Resources and Student Achievement: An Assessment

Julian R. Betts

Public Policy Institute of California and UC-San Diego

Anne Danenberg

Public Policy Institute of California

DRAFT--- March, 2001---Please do not cite without permission.

Prepared for the Joint Committee to Develop a Master Plan for Education – Kindergarten through University

Summary

Although California has striven to equalize revenue across school districts, disparities in student achievement and school resources – broadly defined to include teacher quality as well as revenue – remain large. This paper asks whether increasing school resources in low-achieving schools can reduce or eliminate these achievement gaps in California.

We begin by reviewing the national and state-level evidence on the relationship between school resources and student achievement. Although the research results are mixed, most studies indicate a weak relationship between resources and achievement, especially compared to the strong correlation between student performance and socioeconomic status (SES). The most recent research using California data shows that teacher education, experience, and full credentialing are associated with modest gains in student performance.

With these California findings in mind, we simulate the benefits and costs of increasing resources in low-performing schools. We focus in particular on the achievement gap between California schools whose fifth-grade students score at the 25th percentile and 50th percentile on reading and math tests. Holding all other factors constant and equalizing teacher characteristics at the two sorts of schools, we predict that 1.3 percent more fifth-grade students at the low-achieving school would score at or above the median on standardized math and reading tests than is currently the case. This change would reduce the achievement gap by about 10 percent. We also estimate the benefits and costs of a more dramatic change in school resources: namely, raising teacher characteristics at low-achieving schools to the 90th percentile level for teacher qualifications statewide. In our simulation, this change reduces the achievement gap by about one third.

The cost of this more dramatic increase in teacher qualifications is approximately $300 per student, or 6 percent of per pupil spending at such schools. However, this estimate must be taken with caution. Unobserved factors may be driving the observed

correlations between school characteristics and student outcomes. If so, additional school resources may not have the predicted effects on student performance. Also, the actual costs of increasing teacher qualifications in this way would probably be much higher, as salaries would have to rise by an unknown amount to attract and retain the requisite number of qualified, experienced teachers. Finally, we cannot predict how much additional compensation would be needed to place these teachers in the neediest schools within districts. Given current salary arrangements between teachers and districts, intra-district salary bonuses might be necessary to attract a large number of certified, highly educated, and experienced teachers to low-performing schools.

In light of these and other uncertainties surrounding the relationship between school resources and student achievement, the report concludes with three recommendations for implementing and assessing large school reforms. First, such reforms should be phased in over five or six years. In the initial years, participating schools would be selected randomly. The state would then be able to evaluate the effectiveness of a reform by comparing student outcomes in participating and non-participating schools. This approach also allows for mid-course corrections before reforms were implemented statewide.

Second, California should develop a statewide student data system that allows policymakers to track improvements in student achievement over time. This “longitudinal” database, which Texas has had for several years, would greatly improve our knowledge of cost-effective school spending. It would also improve the Academic Performance Index, which is used to assess school quality in California, and do much to reconcile the divergent estimates of the dropout rate currently provided by the California Department of Education.

Third, we recommend that the state continue to use the Stanford 9 test as part of its statewide testing system. This continuity makes it possible to assess the long-term effects of recent reforms, such as class size reduction and the new state accountability system. Together, these three measures could shed new light on how to narrow inequality and raise achievement in California schools.

Introduction

Since the early 1970s, California has done much to equalize revenue across school districts (Sonstelie, Brunner, and Ardon, 2000). Yet large disparities in both student achievement and school resources---broadly defined to include teacher characteristics and curriculum as well as school revenue---still remain. Both sorts of disparities are strongly related to socioeconomic status (SES). In general, students from low-SES families perform worse on standardized tests than other students . At the same time, school resources vary positively and systematically with student SES. Compared to other students, those from low-SES families attend schools with less educated and less experienced teachers. Low-SES students also attend high schools that offer fewer advanced courses (Betts, Rueben, and Danenberg 2000).

These findings raise the question: How much of the achievement gap in California can be traced to inequalities in school resources? This study addresses that question in two steps. First, it reviews the evidence on the relationship between school resources and student outcomes. Second, it asks how much California would need to spend to reduce or eliminate this achievement gap. This second question is implicit in adequacy-based reforms both in California and other states.[1] California's Public School Accountability Act of 1999, for example, allows schools with particularly low test scores to receive additional funds through the Immediate Intervention/Underperforming Schools Program (II/USP). Many other education bills and programs also provide additional resources to low-achieving schools or to those with high proportions of economically disadvantaged students with the expectation that these extra resources will boost academic performance.

However, the relationship between school resources and student outcomes is not nearly as clear as the one between SES and student achievement. As a result, the extent to which increased spending can reduce or eliminate California's achievement gap is uncertain. Because some of this uncertainty can be traced to the ways California gathers

[1] For discussion of adequacy-based reform, see Rose (2000).

data and implements educational reforms, we conclude the paper with three recommendations regarding data collection and program implementation.

School Resources and Student Outcomes: The Evidence

Research on the relationship between school resources and student outcomes has been conducted intensively for almost four decades. Although the research results are mixed, they yield two basic lessons. First, school resources as we define and measure them do not account for large, systematic differences in student performance. Second, student SES overshadows all school-related factors in determining student achievement. The following discussion highlights the key national and state-level evidence for these conclusions.

National Evidence

Most of what we know about school resources and student achievement comes from studies using national data-sets. One early and influential study, now known as the Coleman Report, examined variations in test scores across a huge sample of students in the mid-1960s (Coleman et al. 1966). The authors found that differences in class size, teacher education, and teacher experience accounted for very little of the large disparities in test scores across schools. Instead, the key factor appeared to be large variations in student SES. The Coleman Report generated considerable controversy, and researchers since that time have used different data sets to test its results. Although this later literature sometimes finds that school resources do affect student achievement, the report's main finding -- that resources matter less than the socioeconomic status of students -- has weathered these replication attempts well.

In a series of influential summaries of test-score research, Hanushek (1986, 1996) concludes that a small proportion of studies have found that additional school resources lead to significantly higher achievement.[2] Table 1, from Hanushek (1996), summarizes some of his findings. For many measures of school resources, such as class size, most

[2] Although Hanushek's claims have been influential, they are not universally accepted. See, for instance, the exchange between Hedges and Greenwald (1994) and Hanushek (1994).

studies find no significant link to student achievement. Other studies even find a link suggesting that more resources are associated with lower achievement. Of the various school resources examined in these studies, teacher experience is found most regularly to have a significant positive relation with student achievement. Overall spending per pupil and teacher salary are the school resources found to matter the second and third most often. Few studies have found that teacher education affects student achievement.

Other national studies reach different conclusions. A key example is a recent study by Grissmer, Flanagan, Kawata, and Williamson (2000), which models the average test scores in each state that participated in National Assessment of Educational Progress (NAEP) between 1990 and 1996 as a function of class size, teacher education, teacher experience, and several other measures of educational resources. They find that class-size variations explained more of the achievement gap than did variations in other measures of school resources, including teacher education and experience. In addition, the authors find that the test-score gap between minorities and whites was smaller in states with smaller class sizes. In light of these findings, the authors maintain that the most efficient use of education dollars is to reduce pupil-teacher ratios in states with high proportions of minority and disadvantaged students, encourage pre-kindergarten, and provide more adequate teaching resources. They also conclude that substantial productivity gains can be made with the current teaching force if working conditions are improved.

The Grissmer study draws on a large number of student test scores, but it measures resources at the state rather than at the school or district level. Even after combining state-level reading and math scores with the NAEP results, the study relies on only 271 observations. Thus, the results should not be seen as definitive. Klein, Hamilton, McCaffrey, and Stecher (2000) examine NAEP data from a slightly different set of years in the 1990s and find that Texas, which the Grissmer report ranks at the top of participating states, outpaced the national average in only one of four tests they examined.[3] Darling-Hammond (2000) examines NAEP data from 1990 to 1996 and finds that teachers' credentials and experience were the two most important factors explaining inter-state variations in test scores, with class size being far less important. These conflicting conclusions indicate that aggregating data at the state level has its limitations. Small changes in the specifications and time period can lead to very different results. Furthermore, these data do not capture the striking variations across schools and districts, especially in a state as large and diverse as California.

In addition to the large body of research on school resources and test scores, a smaller literature examines the relation between school resources and the earnings of students after they leave school and enter the labor force. It may seem odd to ask whether school resources affect students' wages years later, but a key goal of public schooling is to prepare students for successful work lives. It is also possible that the link between school inputs and test scores is weak because tests do not measure the gains in skills that will prepare students for successful work lives. In the end, earnings may be a more useful measure of student success than test scores.

A number of studies have found a relation between adult males' earnings and school resources in their state of birth, but the literature is by no means unanimous (Betts 1996). Work by Betts (1995), Grogger (1996), and others shows that when school resources are measured at the school actually attended, the relationship between school inputs and earnings is not statistically significant. More to the point, the estimated effect of raising school spending on students' subsequent earnings is extremely small. This is true regardless of whether one measures school resources at the school actually attended, in the district attended, or whether one instead uses the person's state of birth to create a rough proxy for school resources.

It is also useful to examine whether school resources are related to the amount of schooling students ultimately attain. Betts (1996) reviews this relatively small body of research and finds weak evidence that school resources affect educational attainment.

[3] For a critique of these two studies, see Hanushek (2001).

In short, four decades of intensive research at the national level suggests a relatively weak relationship between school resources on the one hand and student achievement, educational attainment, and future earnings on the other.

State-Level Evidence: Tennessee and Class-Size Reduction

Perhaps the most famous state-level experiment of the last two decades is Tennessee's class-size reduction of the 1980s. Students in kindergarten through third grade were randomly assigned to one of three groups. The first group had class sizes as low as 15 students; the second group had class sizes in the low 20s and one teacher's aide per class; and the third group had class sizes in the low 20s. Since then, numerous studies have compared test scores for the three groups.

The results indicate that students placed in the small classes learned more quickly than other students. Most of the gains accrued to students in the first year they were in smaller classes, and low-SES students gained somewhat more than others. However, these gains virtually disappeared after students were returned to regular-sized classes (Krueger and Whitmore 1999). Specifically, students in smaller classes had a 4.5 percentile point advantage over other students at the end of third grade, but this advantage had diminished to 1 percentile point by the end of eighth grade. (For example, students that ranked in the 50th percentile on a national test would have risen to the 49th percentile.) In percentage terms, the deterioration in the test-score advantage was slightly higher for students receiving free lunch and slightly lower for black students.

The Tennessee experiment offers the most persuasive evidence to date for reducing class size. Even so, the results suggest that such reductions produce very modest gains, especially if students are placed in larger classes in later grades.

Evidence from California

A number of recent studies have examined school resources and student achievement in California. For example, Betts, Rueben, and Danenberg (2000) analyze the distribution of resources and test scores at the school level for 1997-98. They found that teachers serving low-SES students were considerably less prepared and experienced than teachers serving other students (Figure 1).[4] They also found that low-SES schools had relatively low test scores, raising the question of whether their low achievement was caused by a lack of resources or by the direct effects of poverty.

Regression analyses suggest that school resources did affect achievement, but only slightly. Figure 2 shows the predicted effects on student performance when, holding all other factors constant, a typical school moves from the 25th to the 50th and then to the 75th percentile in SES, class size, or teacher characteristics. As the first trio of bars indicates, the predicted effects of SES on student achievement are large. Schools at the 75th percentile in SES would have 57.5 percent of its students performing above national norms, compared to just 26.8 percent of students at schools at the 25th-percentile. The remaining bars show the predicted effects of changing measures of school resources. All variables in the figure except for class size have a statistically significant impact on student achievement. But the predicted impacts of changing teacher credentials, experience, education, or class size are minor compared to the impact of student SES.

The CSR Consortium (1999, 2000) has also studied the effect of recent class-size reductions in California. As the Consortium authors note, limitations in the state's student data system along with the wholesale implementation of the reform itself prevent them from drawing firm conclusions. The first two reports by the CSR Consortium provide some evidence that third-grade test scores have risen modestly because of class size reductions. In the first year of the study, the CSR Consortium (1999) compared test scores in the state test, the Stanford 9, between students at elementary schools that had implemented class size ceilings of 20 students with students at schools that had not yet adopted the reform. However, the students at schools that did not implement class-size reduction in the first year came from lower- SES families, making any simple comparison problematic. The authors attempt to adjust statistically

---

[4] In this study, SES is measured by the percentage of students receiving full or partial lunch assistance.

for this problem but express reservations about the reliability of their results. The second CSR Consortium report (2000) uses a more complex comparison technique to estimate the effects of class-size reduction. Again, the authors find statistically significant but modest effects of class-size reduction and indicate that the lack of a true comparison group prevents them from generalizing their results.

In short, research in California and the nation as a whole has failed to overturn the main finding of the Coleman Report (1966). Compared to SES, school resources appear to play a modest role in determining variations in student achievement. Many observers regard the class-size reductions in California as improvements in school quality, but the effects appear to be smaller than in the Tennessee experiment.

Estimated Costs of Narrowing the Student Achievement Gap

With these results in mind, we turn to the question of how much California would need to increase school resources to reduce or eliminate achievement gaps. Using data from previous PPIC work,[5] we simulate the allocation of additional resources to low-scoring schools and gauge the effects of these changes on test scores. The three central questions for the simulation are:

- How much would we need to increase resources at schools at the 25th percentile of student achievement to match test scores at schools at the 50th percentile?[6]
- Among the measures of teacher quality, which appear to be the most cost-effective ways of increasing student achievement at low-performing schools?
- How much would such increases in school resources cost the state?

We omit class size from the analysis because the results of Betts, Rueben, and Danenberg (2000) indicate that it had no significant effect on student achievement.

[5] Specifically, we use school resource estimates from Betts, Rueben, and Danenberg (2000), teacher salaries from Rueben and Herr (2000), and overall school costs from Sonstelie (2000).

[6] The simulation examines the performance of fifth graders on reading and math tests in spring 1998. We compare the average characteristics of schools that rank between the 45th and 55th percentile of test scores with those of schools that rank between the 20th and 30th percentile of test scores. We can think of these schools as representing the ‘middle’ or ‘median’ schools in the former case and ‘bottom-quarter’ schools in terms of student achievement.

It should be noted that the simulation is meant to be illustrative rather than prescriptive. Credentialed, experienced, and highly educated teachers cannot be produced by fiat or compelled to teach at particular schools. Instead, supply and demand, collective bargaining agreements, and other labor market and policy considerations govern these arrangements. This exercise simulates student outcomes if teacher characteristics could be distributed without regard for these considerations.

Figure 3 shows the achievement gap in the percentage of students at or above the national median scores in reading and math under various scenarios. The dark bars show the actual gap in this achievement measure between the schools at the 25th and 50th percentiles of student achievement. These gaps are roughly 15 percent, which is to say that schools at the 25th percentile of school achievement have 15 percent fewer students scoring at or above the national median than do schools at the 50th percentile. The cross-hatched bars show a very slight reduction in the predicted gap if teacher characteristics were equalized across the two types of schools. The gap in math scores is predicted to drop from 15.6 percentage points to 14.3 percentage points, while the gap in reading achievement is predicted to drop from 15.1 percentage points to 13.8 percentage points—gains of only 1.3 percent more students scoring at or above the national median for each test.

The lightest bar shows the predicted gap if policymakers were able to make much more significant changes in teacher quality at low-performing schools. Specifically, it indicates predicted outcomes if the state could raise teacher quality at these schools to match teacher quality at schools that rank in the 90th percentile for teacher characteristics. Even when we increase resources to these levels, the predicted increase in test scores at the low-performing schools is rather modest---on the order of 5 percent to 5.5 percent more students would score at or above the median. According to these calculations, even very large increases in teacher quality would not eliminate the achievement gap; rather, they would decrease the gap by about a third.

Table 2 presents the estimated benefit, cost, and benefit-cost ratios of improving teacher qualifications. Each of the three teacher characteristics – experience, educational attainment, and full credentialing – is considered separately. Benefits are defined as the predicted gain in the percentage of students scoring above national norms. Additional costs are derived from salary schedules in Rueben and Herr (2001), and the benefit-cost ratio is the ratio of these two terms.[8] Our sample of low-performing schools includes about 10 percent of all elementary schools in the state.

Table 2 suggests that, at least for fifth-grade math and reading achievement, the benefit-cost ratios for teacher characteristics vary substantially. Having a fully credentialed teacher in every classroom has the greatest benefit relative to its cost, followed by increasing the percentage of teachers with a master's degree. For reading achievement, reducing the number of teachers with at most a bachelor's degree and increasing teacher experience are the third and fourth most cost-effective reforms. For math achievement, these last two measures are reversed; teacher experience ranks third and increasing the percentage of teachers with at most a bachelor's degree ranks fourth.

As the table indicates, in order to improve math scores, the total cost of improving teacher characteristics in this way comes to almost $82.3 million, or just over $306 for each student at these low-performing schools. For reading, the costs are virtually identical. Given that average spending per pupil in 1997-98 for the typical elementary school was $4,881 (Sonstelie 2000), this spending increase seems modest. However, it must be emphasized that these changes would only narrow, not eliminate, the achievement gap. Furthermore, schools that perform below the 25th percentile would require larger increases for the same proportion of students to reach national norms. Finally, our figures represent the cost to a district of a more credentialed, educated, and experienced staff given its salary schedule. To attract such a staff, however, a district would probably have to raise its salary schedule. We have not included this extra cost in our calculations. Detailed longitudinal studies of teachers' careers in California over

[8] The details of the approach appear in Appendix A.

many years could shed light on how much the supply of teachers might respond to such changes in salary schedules. Unfortunately, California at present lacks a data system that tracks teachers over time in this way.

Data Collection and Program Implementation

The simulation above shows how policymakers can use existing research to predict the likely effects of changing school resources on student achievement. The utility of this research depends on the thoroughness and accuracy of the data used to perform the simulation. If the analysis omits important determinants of student achievement, the results may be unreliable. Yet the way in which California currently collects data and implements major education reforms makes it difficult to identify these important determinants. As a result, we learn surprisingly little about the effectiveness of these reforms. For example, a lack of student-level data on gains in performance over time creates large uncertainties. In our simulation, too, the observed variation in school resources in any given year may pick up unobserved variations in other characteristics of students, parents, teachers or school administrators. Although the simulation should give pause to those who believe that equalizing school spending can by itself eliminate the achievement gap, the lack of adequate data for the analysis leaves many questions unanswered.

As a consequence of these methodological difficulties, California policymakers are often forced to rely on national research that may not be wholly relevant to California. For example, the class-size reduction (CSR) initiative appears to have been based on a demonstrably uncertain body of literature that is mostly national in nature. Although the Tennessee class size experiment has drawn national attention, Tennessee's student population differs in important ways from California's. These differences raise the possibility that class-size reduction in California might have quite different effects than those observed in Tennessee. Moreover, the Tennessee experiment reduced class size from about 23 to 15, while the California reform reduced class sizes from the upper 20s to 20. If the effects of class size on achievement are non-linear, the Tennessee experiment might not provide an accurate guide to outcomes in California.

California policymakers deserve considerable credit for commissioning a formal evaluation of the CSR initiative. Because the reduction in class sizes was not phased in over time, however, it will be extremely difficult to evaluate its effects. The central problem is the lack of a control group against which to compare the gains in achievement of students placed in smaller classes. As mentioned earlier, the first CSR evaluation could only compare test scores of students in small classes to those of students who did not get smaller class sizes. However, the latter group of students does not represent a valid comparison group. As the CSR Consortium authors are careful to indicate, the students in larger classes were a highly non-random group: Schools with more disadvantaged students were markedly less likely to reduce class size in the first year of the program. So the finding that students in larger classes have lower test scores may in part arise simply because of these students' relative disadvantage.

In later years, state evaluators can compare achievement of students who received up to four years in small classes compared to just one, two, or three years for older cohorts. But a problem with comparing test scores of older and younger students is that these students will vary in their familiarity with test-taking, which affects test scores over time. Koretz (1996) recounts evidence that rising test scores in one school district reflected growing student (and teacher) familiarity with the test form. Because California has used the same test form since spring 1998, we cannot simply compare different student cohorts that have had differing degrees of exposure to the Stanford 9 test. This problem threatens to invalidate the evaluation of the relationship between test scores and class-size reduction in California.

A related and severe problem affecting the CSR evaluation is that the state's student test score databank does not follow individual students over time. This forces analysts to compare different cohorts of students in two different years. This is a potentially dangerous approach because different cohorts could vary in achievement for reasons quite unrelated to schools and teachers. The lack of a statewide database of this sort also creates problems for the Academic Performance Index, which the state uses to rank public schools. If test scores fall in second grade at Lincoln Elementary between 1999 and 2000, are teachers doing a less effective job as time passes, or does the decline represent some unobservable change in the students and their capabilities? We can never know with certainty. The only solution is to examine gains in test scores for individual students over time.

A Proposal

We propose several straightforward reforms that would vastly improve the ability of California to analyze the effectiveness of school resources and evaluate major educational reforms. If even a part of these proposals were implemented in the new Master Plan, it could revolutionize the quality of education research in California. It would also lessen the state's current reliance on out-of-state research, which might not apply to California's student population.

Our proposal has three parts:

1. Any major educational reform should be phased in over five or six years. If more schools apply for the new program in initial years than the state can accommodate, the state should select schools randomly through a lottery.

This reform will achieve several goals. First, the phase-in can save money by lowering up-front costs and allowing for cost-saving mid-course corrections based on early evaluations. Second, schools that do not win the lottery create a group against which the schools undergoing reform can be compared. This change would allow for the first truly valid evaluations of education reforms. Also, a lottery will be perceived as fairer to schools compared to an opaque selection process. Of course, if policymakers wanted to direct the initial stages of a program to a particular group of schools, say, those with low test scores, they could still do so by restricting the program to those schools or, less drastically, by having a series of lotteries with different odds of "winning" for schools in different categories. Perhaps most usefully, the state could select a “stratified” random sample of schools in its lottery. It would sample schools across the socioeconomic spectrum; rural schools, suburban and urban schools; small schools and large schools. In this way the state could scientifically determine whether a specific reform worked better in some types of schools than others.

This approach might become all the more important given recent discussions in Sacramento about the possibility of moving away from categorical, top-down reforms to a more decentralized block-grant approach. Statewide evaluations of the sort we propose could do much to prove or disprove the notion that “one size fits all” in school reform. If the evaluations suggested that, in fact, one size did not fit all, then the evaluations would at the same time provide strong clues to each district about what might work best in its schools. Notably, few districts could afford to conduct similar evaluations on their own, and would probably lack a sufficiently large number of schools to learn anything with the same degree of precision.

2. The state must maintain one or more consistent measures of achievement statewide over many years. In particular, it should continue to use the SAT9 test even as it expands other components of the school accountability system

California has a history of introducing and then abandoning state test instruments. It is easy to find fault with any of the existing or proposed forms of state tests. But without continuity, policymakers are doomed to learn little about trends in achievement or the effectiveness of reforms such as CSR or recently implemented expansions in teacher training.

Our recommendation applies to current measures of student performance as well to new ones being phased in, such as the high school exit exam. But it is especially important to maintain the current SAT9 test, even if it is not geared to the recently adopted state content standards. The SAT9 is the only way that we can track student performance from the late 1990s forward. It is also the only component of the future proposed version of the API that provides comparison to a nationally representative comparison group.

3. California should create a database that tracks the achievement and transcripts of individual students over time.

Although such proposals may raise concerns in some quarters about confidentiality and fairness, a longitudinal system similar to the one we propose has been in use in Texas for some years and has survived legal challenge. There are four important reasons why California must move to such a system soon. First, evaluations of education reforms such as CSR would be greatly improved if they were based on analysis of gains in achievement on a student-by-basis, instead of relying on comparisons of successive cohorts.

Second, non-experimental analyses of the impact of school resources on student outcomes, such as that in Betts, Rueben and Danenberg (2000), would be far less prone to uncertainty if they were based on gains in the scores of individual students.

Third, the state accountability system places great reliance on the Academic Performance Index. Yet under the current system, a school's API ranking might fall from one year to the next simply because of unobserved student mobility between one school and another. Because inner city schools tend to have relatively high student mobility, the API is likely to provide a relatively less accurate measure of changes in school quality over time for such schools. If the API were based on gains in achievement among individual students who had spent two years in the school, it would eliminate such distortions.

Fourth, the inability of California to follow individual students over time has seriously affected the quality of even the most basic education data available to legislators and other policymakers. To give one example, California's database on dropout rates is surprisingly weak. High schools have difficulty verifying whether students who leave have dropped out, moved to other districts, or left the country.

According to the California Department of Education, between 1995-96 and 1998-99 the one-year dropout rate averaged about 3.5 percent. This figure implies that by the time a ninth-grade cohort reaches the end of their senior year, just over 13 percent of students will have dropped out. But if we compare enrollment in ninth grade with the actual number of high school graduates four years later, we obtain a dropout rate of 30 percent. (This estimate applies to any cohort graduating in the late 1990s.) This second dropout estimate is more than double the rate from the first method! If California moved quickly to create a longitudinal database with a unique student identification code, we could do much to reconcile this huge discrepancy.

A statewide longitudinal data system would do much to solve such critical problems. California School Information Services (CSIS), an experimental longitudinal data system in which a number of districts currently participate, might form the kernel of a more ambitious statewide system. However, at present, CSIS covers only a minority of students in California; one California Department of Education official warns that it might take another ten years before California has implemented a statewide electronic student data system (Asimov 1999).

In sum, three simple reforms could transform California from a net importer of education policy research to a leader in the field. Given the amount of money California spends on its educational system, more accurate, disaggregated analyses upon which to base estimates of benefits and costs for resource allocation in California are crucial. The legislature could consider adopting an oversight law that was triggered by any educational reform that costs more than, perhaps, $100 million a year when fully implemented. The legislation would require that such a reform must be phased in slowly over five or six years, with a lottery mechanism for selecting early participants. The resulting evaluation would compare outcomes in schools that won and lost the lottery in order to provide a reliable assessment of the impact of the given educational reform. In addition, the legislation might automatically set aside financial resources for a state-sponsored evaluation of new reforms.

Conclusion

Thirty-five years of educational research has consistently shown that school-level variations in student disadvantage explain more of the achievement gap than do variations in measures of school resources such as class size, teacher education, and teacher experience. Some research has found that specific measures of school inputs "matter," either for test scores, graduation rates, or future earnings, but the effects, when present, are small. Even in the more sophisticated recent research, poverty still drives most of the large variations in student achievement across schools.

This fundamental fact has given rise to the notion of "educational adequacy." Exact definitions of this concept vary, but in its purest form, educational adequacy means that schools with high proportions of disadvantaged students require greater than average resources to provide an adequate education. In other words, three decades of court-induced revenue equalization has not done enough. The decision to spend more than average on schools in disadvantaged areas is a political one. But the facts are clear: Funding equalization will not equalize test scores across schools.

We therefore examine how increasing resources at schools near the 25th percentile of California test scores might move achievement at these schools toward achievement of schools near the 50th percentile of California test scores. First, we confirm that equalization of resources between these two types of schools would barely put a dent in the test-score gap. If resources were equalized, for example, we predict that the existing gap of 15.1 percent in the percentage of fifth-grade students scoring at or above national norms in math would shrink by only 1.25 percent -- to 13.8 percent.

Second, we examine what an "adequate" level of resources might look like by simulating the extent to which one could reduce this test score gap by increasing teacher qualifications at low-achieving schools. (Class size does not have a statistically significant effect on test scores in fifth grade, so we focus instead on teacher preparation.) If we raise teacher preparation at these low-achieving schools to the level observed in schools with the most experienced, highly educated teaching staffs, we find that the test-score gap between low-performing and median schools drops by one third.

These gains in student achievement appear to come at a rather modest cost. In the low scoring schools, spending per pupil would have to rise by just over $300 per pupil. However, this estimate provides a lower bound on the true costs, which could easily be twice or even five times as great as we have estimated. We have assumed that low-achieving schools could simply hire more educated, experienced, and fully credentialed teachers. Our cost calculations therefore assume that the only cost to the school would be the higher salaries such teachers command. But as Betts, Rueben, and Danenberg (2000) report, schools with low test scores, often in the inner city or rural areas, sometimes suffer severe shortages of qualified teachers, even if salaries do not lag far behind those paid in high-achieving schools. In general, fully credentialed, highly experienced, and highly educated teachers in California prefer to work in schools with high test scores and low levels of student disadvantage. As a result, it may take higher than average salaries to attract such teachers to the schools in greatest need.

A recurring theme throughout the analysis has been uncertainty about the effect of school resources on student achievement in the nation and in California. We discuss some of the roadblocks preventing social scientists and policymakers from obtaining more precise and accurate measures of the relative effectiveness of various types of school resources. There have been three perennial problems in California. First, California lacks a systematic way of evaluating important reforms such as class size reduction or teacher training programs. Suppose that a claim is made that a specific educational reform has made California schools better. The skeptical listener should ask, “Better than what”? For instance, test scores in California’s elementary schools have risen considerably since the spring of 1998, when the new state test was implemented. Could the class-size reduction program that began several years ago have generated these gains? Perhaps, but it is extremely difficult to know for sure. How much would test scores have risen if class size had not been reduced? It is almost certain that test scores would have risen anyway, given research findings that when a new test is introduced, scores increase over the first few years because teachers and students become more familiar with the test format. In the case of the SAT9, the problem is compounded by the fact that the same test questions have been used in a given grade each year.

In order to allow for more accurate evaluations of important educational reforms like CSR, we recommend that all major new education initiatives be phased in over five or six years. In the early years, schools should be chosen to participate using a lottery. This system provides policymakers with a group of schools against which to compare gains in student outcomes at the schools that undertake the reform. This comparison would provide a compelling test of whether taxpayers' dollars were being spent productively. This change would also save taxpayers money in the early years, and research results from the first five years of the evaluation could allow for improvements to the program before it was implemented statewide.

California also lacks a statewide database, similar to that used by Texas, which follows individual student progress over time. State-mandated evaluations currently rely on average achievement at each school, and so are subject to error. Similarly, the Academic Performance Index (API) measures do not take account of students who switch schools within a district. Because disadvantaged students change schools relatively frequently, the API may be biased against schools that serve large populations of such students. A statewide student-level database could solve this problem permanently. Similarly, such a dataset could help to resolve longstanding controversies about the high school dropout rate in California.

Finally, California should continue to use the SAT9 test for some time to come. Without this continuity, it will prove impossible to evaluate important reforms such as the class size reduction initiative or the school accountability program implemented in 1999.

References

Asimov, Nanette, "Dropout Rate Drops – But Not Far Enough. State Among Worst for Earning a Diploma," *San Francisco Chronicle*, June 8, 1999.

Betts, Julian R., "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *Review of Economics and Statistics* Vol. 77, 1995, pp. 231-50.

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, Washington, D.C., 1996

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

Bohrnstedt, George W., and Brian M. Stecher, eds., *Class Size Reduction in California: Early Evaluation Findings, 1996-1998* (CSR Research Consortium, Year 1 Evaluation Report), American Institutes for Research, Palo Alto, California, 1999.

California Department of Education, Education Demographics Unit, October 1997 Professional Assignment Information Form (PAIF) electronic database, Sacramento, California, 1998.

California Department of Education, Education Demographics Unit , October 1997 School Information Form (SIF) electronic database, Sacramento, California, 1998.

California Department of Education, Education Demographics Unit, October 1997 Aid to Families with Dependent Children (AFDC) electronic database, Sacramento, California, 1998.

California Department of Education, Education Demographics Unit, April 1998 Stanford 9 (STAR) electronic database, Sacramento, California, 1999.

Coleman, James et al., *Equality of Educational Opportunity*, Government Printing Office, 1966.

CSR Research Consortium, *Class Size Reduction in California 1996-98: Early Findings Signal Promise and Concerns*, American Institutes for Research, Palo Alto, California, 1999.

CSR Research Consortium, *Class Size Reduction in California: The 1998-99 Evaluation Findings*, American Institutes for Research, Palo Alto, California, 2000.

Darling-Hammond, Linda, "Teacher Quality and Student Achievement: A Review of State Policy Evidence," *Education Policy Analysis Archives* Vol. 8:1, January 2000.

Grissmer, David W., Ann Flanagan, Jennifer Kawata, and Stephanie Williamson, "Improving Student Achievement: What NAEP State Test Scores Tell Us," RAND, Santa Monica, California, 2000.

Grogger, Jeff, "Does School Quality Explain the Recent Black / White Wage Trend?" *Journal of Labor Economics* Vol. 14, 1996, pp. 231-53.

Hanushek, Eric A., "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* Vol. 24, 1986, pp. 1141-77.

Hanushek, Eric A., "Money Might Matter Somewhere: A Response to Hedges, Laine and Greenwald," *Educational Researcher* Vol. 23, 1994, pp. 5-8.

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, Washington, D.C., 1996, pp. 43-73.

Hanushek, Eric A., "Deconstructing RAND,“ *Education Matters* Vol. 1, No.1, January 2001.

Hanushek, Eric A., "RAND vs. RAND,“ *Education Matters* Vol. 1, No.1, January 2001.

Heckman, James, Anne Layne-Farrar, and Petra Todd, ”Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship," in Gary Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success* Brookings Institution, Washington, D.C., 1996, pp. 192-289.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald, “Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes,” *Educational Researcher* Vol. 23, 1994, pp. 5-14.

Hedges, Larry V., and Rob Greenwald, "Have Times Changed? The Relation between 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, Washington, D.C., 1996, pp. 74-92.

Klein, Stephen P., Laura S. Hamilton, Daniel F. McCaffrey, and Brian Stecher, “What Do Test Scores in Texas Tell Us?” RAND Issue Paper #202, 2000 downloaded 2/2001 from http://www.rand.org/publications/IP/IP202/.

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,* National Academy Press, Washington, D.C., 1996, pp. 171-195.

Krueger, Alan B., and Diane M. Whitmore, “The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR,” Princeton University Industrial Relations Section, Working Paper #427, 1999.

Rose, Heather, "The Concept of Adequacy and School Finance,” Draft Report, Public Policy Institute of California, October 2000.

Rueben, Kim S., and Jane Leber Herr, "Teacher Salaries in California," Draft Report, Public Policy Institute of California, November 2000.

Sonstelie, Jon, "Towards Cost and Quality Models for California's Public Schools,” Draft Report, Public Policy Institute of California, December 2000.

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

Appendix A

1. **Differences Between Low-performing and Medium-performing Schools**

We begin the simulation by examining the differences between the two groups of schools. We first rank all California elementary schools that offer grade 5 by their grade 5 math and reading scores. Table A.1 shows the average test scores, socioeconomic status (measured by the percentage of students receiving free or reduced-price lunch), class size, teacher experience, teacher education, and teacher credentials for 20th to 30th percentile and 45th to 55th percentile schools. Because much of the variation in test scores among schools in California reflects differences in the percentage of students who are English Learners (EL), we focus on the gap in test scores of non-EL students. Meaningful variations in test scores, socioeconomic status (SES), and school resources emerge from this comparison. Clearly the two biggest differences between these two sets of schools are the 15 percentage point gap in the percentage of students scoring at or above the national median in reading and math, and the roughly 20 percentage point gap in the share of students who receive free or reduced-price lunch. School resources, especially related to teacher background, also differ, but to a lesser extent.

2. Expected benefits from change in teacher characteristics.

We calculated the increase in the percentage of students that would be expected to score at or above the national median if we were to increase the average resources at the low-performing school to the average level observed at the medium-performing school, which we define as schools scoring between the 45th and 55th percentile of students at or above the national median. We start by taking an enrollment-weighted average of selected characteristics in two groups of elementary schools that include grade 5 and have test scores in the two ranges that represent a low-performing school (20th to 30th percentile) and a median-performing school (45th to 55th percentile). This selection yields

12,498.9 FTE teachers in 379 low-performing schools that had 5th grade reading tests, and 12,463.32 FTE teachers in 398 low-performing schools that had 5th grade math tests. We then calculate the difference between the average resource levels for the low-performing group of schools and the medium-performing schools, and multiply the difference by the expected gain (loss) per unit for each characteristic to obtain an expected gain or loss in the percentage of students who would score at or above the national median.

Table A.2 shows what an equalization of resources would produce. We use the results from Betts, Rueben and Danenberg (2000) to predict how much student achievement would rise at schools with low test scores if they received the level of school resources observed at schools with typical test scores in California. The top part of the table shows results for math achievement, and the bottom part shows the results for reading. Betts, Rueben and Danenberg (2000) found that larger class size has an unexpected positive relationship to test scores, and was not statistically significant. But we include changes in class size simply to illustrate the relative predicted effects of changing class size and the various measures of teacher preparation.

The first two columns of numbers in Table A.2 simply repeat the resource level shown in Table A.1 for the low-scoring and median-scoring schools. The third column shows the increase in the resources that would be needed at the low-scoring schools so that they would have the same resources as the median-scoring schools. Column 4 shows the predicted change in the percentage of students scoring at or above the national median from a one-unit change in the stated school resource. These estimates are based on the regression analysis in Betts, Rueben and Danenberg (2000). For an example of how to interpret these numbers, the number in the first row tells us that if average teacher experience rises by one year, the share of students predicted to score at or above the national norm in math would rise modestly, by 0.235 percent. To calculate the predicted effect from improving each measure of teacher preparation, we multiply the change in resources from column 3 by the predicted effects of a one-unit change in the resource, in column 4, to give the predicted change in column 5. The table shows that most of the predicted gains from equalizing resources come from narrowing the gap in the percentage of teachers who are not fully credentialed. At the bottom of each section of the table the sum of the predicted changes in the share of students at or above national norms is calculated. For both math and reading, resource equalization is predicted to increase the percentage of students at or above the national median by just over one percentage point. It became immediately apparent that the expected gain of 1.25 to 1.3 percent more students scoring at or above the national median would be so small that we would need to increase teacher resource-levels beyond the equalization point for these two groups of schools.

Given the small benefits expected from equalizing resources between low and medium scoring schools, we estimated the expected benefit of increasing the average level of teacher resources at the low-performing schools to the 90th percentile of school resources observed statewide. Table A.3 shows the changes in resources in the simulation, along with the predicted changes in student achievement and the gap in student achievement between the low-performing schools. The top half of the table performs the simulation for equalizing math achievement; the bottom half repeats the analysis for reading achievement.

We also repeated this exercise calculating the difference between resources at the same low-performing schools and high-performing schools, defined as the group of schools between the 85th and 95th percentile of *test scores* in California.

In Table A.4, we show the expected benefit from increasing the average resource level of a low-performing school to that of a high-performing school. Again, the expected gains are small—on the order of 2.5 to 3 percent more students are expected to score at or above the national median. We did not discuss this simulation in the main text because of space constraints.

3. Expected Costs.

New Salary Schedule. Appendix Table A.5 shows the average salary schedule we used to assign costs to teachers. Building upon the average salaries calculated by

Rueben and Herr (2000), we collapsed education and experience categories to match data from the Professional Assignment Information Form (PAIF) filled out by or on behalf of teachers. The PAIF has only six educational categories, whereas the J-90 form used by Rueben and Herr has numerous combinations of education and experience, which they collapsed into 30 distinct cells. We further collapsed this salary schedule into the 15 cells shown in Appendix Table A.6 by taking a weighted average across the salaries in Rueben and Herr categories that corresponded to education levels that would have salaries equivalent to a Master's or more in the PAIF data.

Baseline cost estimation. Each FTE above is assigned to a cell in Table A.5 according to their combination of experience and education to compute a baseline salary cost in reading and math for the low-performing groups of schools. Table A.6 shows the baseline salary matrix for the teachers in each education and experience combination for schools that have grade 5 reading and math test scores. Using Rueben and Herr's estimate of average health benefits and Sonstelie's estimate of retirement and workers' compensation benefits, we further estimated a baseline cost for the combination of average experience and education observed at the low-performing group of schools. Following these authors, we estimate that some of these benefits costs are proportional to wages, equaling 12.19 percent of wages, while other costs are fixed.

Simulated change in teacher characteristics. Next we simulated how teachers would be expected to shift cells if we increase from the average level of characteristics seen at the low-performing school to the 90th percentile of characteristics seen in the state. To do this, we first moved individual teachers across experience categories. In the first stage, we moved teachers from the experience category "0-2 years" as required to reduce the percentage of teachers who lacked full credentials. The J-90 form, which collects salary data from districts, does not indicate how the possession or acquisition of credentials is reflected in the salary schedule, yet lacking a full credential has a strong negative relationship to test scores. After analyzing the distribution of teachers lacking full credentials, we find that low experience and lack of credentials are positively and fairly highly correlated (0.46), and that the majority of teachers without a full credential also have low experience. We thus use a proxy for lack of credential, measured as the percentage of teachers with fewer than 3 years, to estimate the numbers of teachers who would move from being uncredentialed to credentialed in order to incorporate them in the salary matrix. We do not treat teacher experience and the percentage of teachers with full credentials as completely separate policy tools in the sense that improving one component of teacher experience cannot in practice be done without improving the other component. Then we calculated the remaining increase in average years of experience that was required to raise average experience to the needed level, and re-allocated teachers accordingly. Next, we changed the percentage of teachers in education categories to match the 90th percentile statewide. We reallocated teachers from one education level to the next higher level in proportions so that the overall mix of experience ranges for at most a Bachelor’s and at least a Master’s education level were maintained after the shift. For instance, if 4 percent of teachers with a minimum of a Master’s had 0-2 years experience before the shift, then the same percentage would have this experience level after the shift.

We multiplied the elements of the new matrix of the numbers of teachers in each cell by the corresponding elements in the baseline salary matrix and added the estimated benefits package to calculate an increased cost for the simulated set of teacher characteristics. We then subtracted the baseline cost from the new cost. The cost differences were divided by the number of pupils in the low-performing schools for math (268,690 students) and reading (271,425 students) to estimate per-pupil incremental costs.

Costs for simulated change in teacher characteristics. Column 2 in Table 2 in the main text shows the incremental costs for changing each characteristic. We calculated each of these separately by estimating the total cost of funding a school at the given level and then subtracting the baseline cost from Table A.6. We do not show these tables with total cost estimates because of space constraints. However, these tables are available upon request.

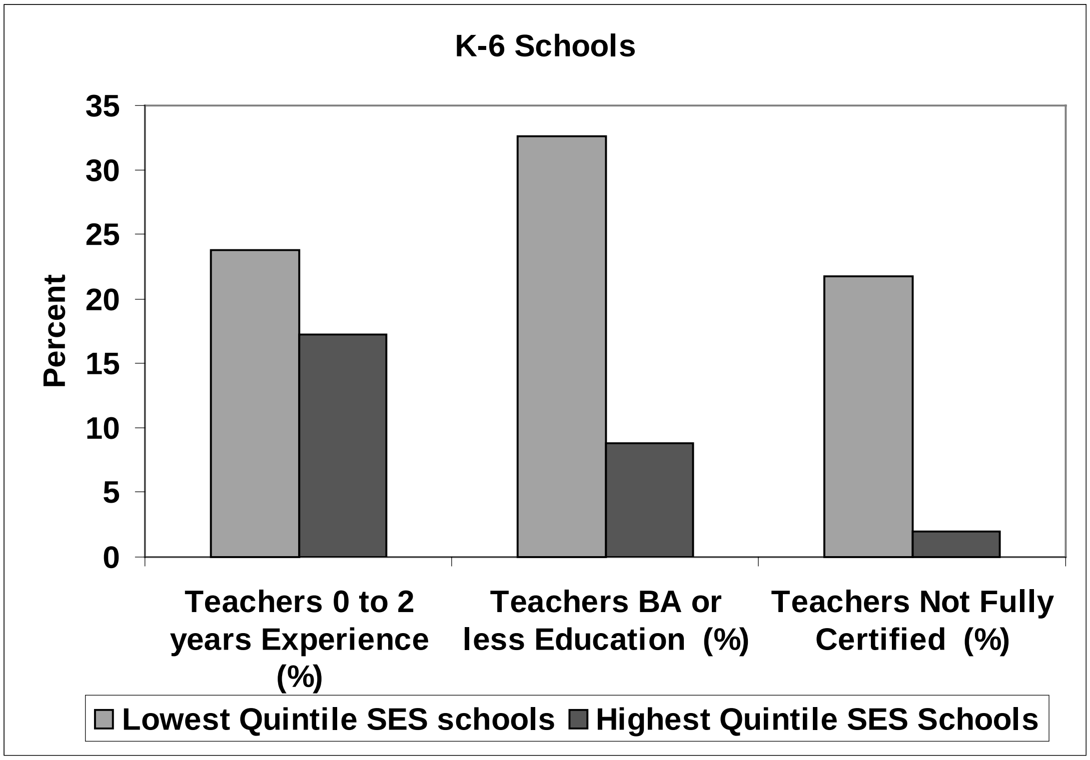
4. Benefit to cost ratios.

After estimating the expected benefits and costs, we calculated a benefit to cost ratio by dividing the expected benefit by the estimated cost per pupil. In addition to calculating a change in all characteristics, we calculated the expected benefit and cost of changing each characteristic independently of the others one-by-one. For example, all other things being equal, we wanted to see how much it would cost if only the experience of teachers was increased. This allowed us to calculate the benefit to cost ratios for each of the changes to estimate which characteristic would be most cost-effective to change. In practice we found very little difference in the total cost per student when we changed all teacher characteristics separately rather than independently.

Larger class size has an unexpected positive relationship to test scores (see Tables A.2 and A.3). Furthermore, class size was not significant in the regression results from Betts, Rueben, and Danenberg (2000). Therefore in the simulation we increase only the measures of teacher preparation listed above, while leaving class size unchanged.

TABLES AND FIGURES

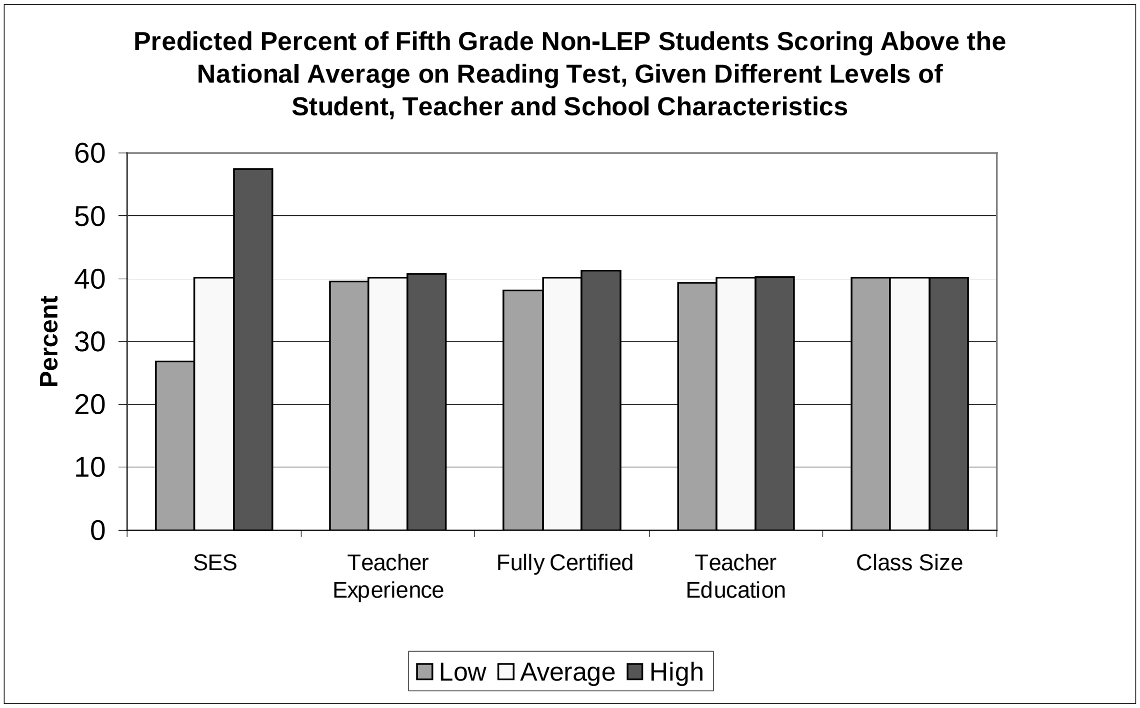
Figure 1. Teacher Characteristics and Student Socioeconomic Status, 1997-1998



Source: Betts, Rueben and Danenberg (2000).

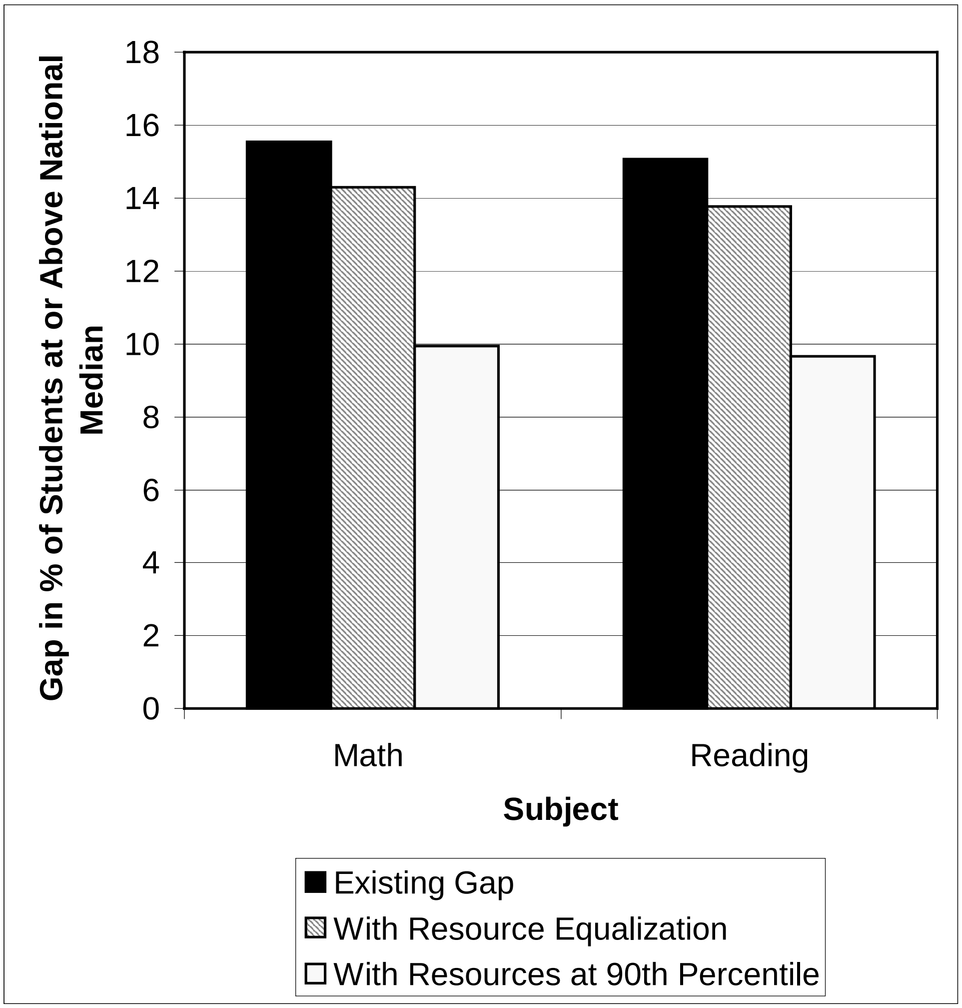
Figure 2

Predicted Percentage of Students Scoring Above the National Average by SES and School Resources



Note: Source: Betts, Rueben and Danenberg (2000). Low, average, and high levels of resources refer to the resource level at the schools ranked at the 25th, 50th and 75th percentile of the resource in the state.

Figure 3. The Inter-School Gap in the Percentage of Students Scoring at or Above National Norms when Resources Are Equalized or Increased to the 90th Percentile



Note: The inter-school gap refers to the gap between schools at the 50th and 25th percentile of test scores in California.

Table 1. Percentage Distribution of Estimated Effects of School Resources on Student Performance as Calculated by Hanushek (1996)

Resource	Number of Estimates	% Positive and Statistically Significant	% Negative and Statistically Significant	% Statistically Insignificant
Teacher-Pupil Ratio	277	15	13	72
Teacher Education	171	9	5	86
Teacher Experience	207	29	5	66
Teacher Salary	119	20	7	73
Expenditure per Pupil	163	27	7	66
Administrative Inputs	75	12	5	83
Facilities	91	9	5	86

Source: Hanushek (1996), Table 3.

Table 2
Benefits and costs of changing each teacher measure independently

	Benefit (additional % students scoring in top half)	Incremental cost across low-performing schools ($)	Per pupil incremental cost ($)	Benefit / cost ratio
	MATH			
Measure				
Increase Years Experience	1.18	46691059.21	173.77	0.007
Increase % Master's	1.82	20,472,485.57	76.19	0.024
Reduce % Bachelor's	0.31	7,145,442.63	26.59	0.012
Reduce % non-Credentialed	2.30	7,960,261.49	29.63	0.078
TOTAL	5.61	82,269,248.89	306.19	0.018
	READING			
Increase Years Experience	1.58	46461178.58	171.18	0.009
Increase % Master's	0.99	20,084,623.64	74.00	0.013
Reduce % Bachelor's	0.11	8,335,270.54	30.71	0.003
Reduce % non-Credentialed	2.74	7,732,498.03	28.49	0.096
TOTAL	5.42	82,613,570.78	304.37	0.018

Note: The incremental costs in the second column refer to the total predicted costs of improving the various measures of teacher preparation to the 90th percentile observed among all California elementary schools. We include all schools ranked between the 20th and 30th percentile of state test scores, or about 11% of all elementary schools, in these cost calculations.

Table A.1
Characteristics of low-performing schools and middle-performing schools, 1997-1998

	Reading			Math		
	Low school	Median school	Difference	Low school	Median school	Difference
Percentage of non-ELL students at or above national median	26.38	41.93	15.55	26.26	41.32	15.06
Percentage of students receiving free/reduced-price lunch	80.41	60.25	-20.16	77.16	59.48	-17.68
Average teacher experience	11.61	12.47	0.86	11.62	12.43	0.81
Percentage of teachers with at least Master's	24.55	26.12	1.57	24.00	26.46	2.47
Percentage of teachers with at most Bachelor's	28.70	19.88	-8.82	25.93	19.56	-6.37
Percentage of teachers not fully credentialed	17.24	11.42	-5.82	16.09	10.78	-5.31
Average class size	25.39	24.95	-0.44	24.81	25.02	0.22

Table A.2
Changes in resources from low-performing school level to resource levels at median-performing school and expected gain in percentage of students scoring at or above the national median, 1997-1998
Math

Characteristic	Low school	Median-performing school	Difference	Predicted Gain or Loss from 1-Unit Change*	Benefit (difference *gain)
Average teacher experience	11.62	12.43	0.81	0.235	0.19
Percentage of teachers with at least Master's	24.00	26.46	2.47	0.086	0.21
Percentage of teachers with at most Bachelor's	25.93	19.56	-6.37	-0.013	0.08
Percentage of teachers not fully credentialed	16.09	10.78	-5.31	-0.143	0.76
Average class size	24.81	25.02	0.22	0.035	0.01
Total predicted percentage of additional students scoring in top half					**1.25**

Reading

Characteristic	Low school	Median-performing school	Difference	Predicted Gain or Loss from 1-Unit Change*	Benefit (difference *gain)
Average teacher experience	11.61	12.47	0.86	0.315	0.27
Percentage of teachers with at least Master's	24.55	26.12	1.57	0.048	0.08
Percentage of teachers with at most Bachelor's	28.70	19.88	-8.82	-0.004	0.04
Percentage of teachers not fully credentialed	17.24	11.42	-5.82	-0.159	0.93
Average class size	25.39	24.95	-0.44	0.042	-0.02
Total predicted percentage of additional students scoring in top half					**1.29**

* from Betts, Rueben, and Danenberg (2000) Table 8.2
** from Betts, Rueben, and Danenberg (2000) Table 8.1

Table A.3
Changes in resources from low-performing school level to statewide 90th percentile resource Level and expected gain in percentage of students scoring at or above the national median, 1997-1998.

Math

Characteristic	**Low school**	**90th Percentile Statewide**	**Difference**	**Predicted Gain or Loss from 1-Unit Change***	**Benefit (difference*gain)**
Average teacher experience	11.62	16.64	5.02	0.235	1.18
Percentage of teachers with at least Master's	24.00	45.16	21.16	0.086	1.82
Percentage of teachers with at most Bachelor's	25.93	1.87	-24.06	-0.013	0.31
Percentage of teachers not fully credentialed	16.09	0.00	-16.09	-0.143	2.30
Average class size	24.81	20.00	-4.81	0.035	-0.17
Total predicted percentage of additional students scoring in top half					**5.45**

Reading

Characteristic	**Low school**	**90th Percentile Statewide**	**Difference**	**Predicted Gain or Loss from 1-Unit Change***	**Benefit (difference*gain)**
Average teacher experience	11.61	16.64	5.03	0.315	1.58
Percentage of teachers with at least Master's	24.55	45.16	20.61	0.048	0.99
Percentage of teachers with at most Bachelor's	28.70	1.87	-26.83	-0.004	0.11
Percentage of teachers not fully credentialed	17.24	0.00	-17.24	-0.159	2.74
Average class size	25.39	20.00	-5.39	0.042	-0.23
Total predicted percentage of additional students scoring in top half					**5.19**

Table A.4

Changes in average resources from low-performing school level to a high-performing school and expected gain in percentage of students scoring at or above the national median, 1997-1998

Math

Characteristic	**Low performing school**	**High performing school**	**Difference**	**Predicted Gain or Loss from 1-Unit Change***	**Benefit (difference*gain)**
Average teacher experience	11.62	13.44	1.82	0.235	0.43
Percentage of teachers with at least Master's	24.00	29.52	5.52	0.086	0.47
Percentage of teachers with at most Bachelor's	25.93	13.11	-12.82	-0.013	0.17
Percentage of teachers not fully credentialed	16.09	5.82	-10.27	-0.143	1.47
Average class size	24.81	25.49	0.69	0.035	0.02
Total expected percentage of additional students scoring in top half					**2.56**

Reading

Characteristic	Low performing school	High performing school	Difference	Predicted Gain or Loss from 1-Unit Change*	Benefit (difference*gain)
Average teacher experience	11.61	13.58	1.97	0.315	0.62
Percentage of teachers with at least Master's	24.55	29.48	4.93	0.048	0.24
Percentage of teachers with at most Bachelor's	28.70	12.11	-16.59	-0.004	0.07
Percentage of teachers not fully credentialed	17.24	4.28	-12.95	-0.159	2.06
Average class size	25.39	25.74	0.36	0.042	0.01
Total expected percentage of additional students scoring in top half					**3.00**

* from Betts, Rueben, and Danenberg (2000) Table 8.2

** from Betts, Rueben, and Danenberg (2000) Table 8.1

(A low-performing school is defined as being between the 20th and 30th percentile of scores. A high-performing school is defined as being between the 85th and 95th percentile of scores.

Table A.5

Average Annual Salary, by Experience and Education Categories, 1997-1998

Experience Levels	Average Annual Salary $		
	MaxBA	BA + 30 Units	MinMA
0-2 Years	29,873	30,893	32,839
3-5 Years	31,892	33,860	36,610
6-10 Years	36,163	39,262	43,316
11-19 Years	38,004	42,536	50,004
20 or More Years	38,645	43,426	53,238

Source: Rueben & Herr (2000)

Table A.6

Baseline Costs for Teachers in Low-performing Schools for Math and Reading, 1997-1998

Math

Experience Levels	FTE Teachers MaxBA	FTE Teachers BA + 30 Units	FTE Teachers MinMA	Average Annual Salary $ Max BA	Average Annual Salary $ BA + 30 Units	Average Annual Salary $ Min MA	Cost ($) MaxBA	Cost ($) BA + 30 Units	Cost ($) MinMA	
0-2 Years	1581.23	1034.54	221.03	29,873	30,893	32,839	47,236,084	31,960,044	7,258,501	
3-5 Years	540.9	1030.29	277.36	31,892	33,860	36,610	17,250,383	34,885,619	10,154,086	
6-10 Years	421.86	1404.32	519.01	36,163	39,262	43,316	15,255,723	55,136,412	22,481,318	
11-19 Years	261.08	1506.19	822.96	38,004	42,536	50,004	9,922,084	64,067,298	41,151,695	
20 or More Years	359	1317.31	1166.24	38,645	43,426	53,238	13,873,555	57,205,504	62,088,332	
TOTAL	**3164.07**	**6292.65**	**3006.6**				**103,537,829**	**243,254,877**	**143,133,932**	**489,926,638**
Retirement and workers' compensation	0.1219						12,621,261	29,652,770	17,448,026	59,722,057
Insurance Benefits package: average	4,455	12463.32								55,524,091
TOTAL COST										**605,172,786**

Reading

Experience Levels	FTE Teachers MaxBA	FTE Teachers BA + 30 Units	FTE Teachers MinMA	Average Annual Salary $ MaxBA	Average Annual Salary $ BA + 30 Units	Average Annual Salary $ MinMA	Cost ($) MaxBA	Cost ($) BA + 30 Units	Cost ($) MinMA	
0-2 Years	1677.28	917.3	208.01	29,873	30,893	32,839	50,105,385	28,338,149	6,830,932	
3-5 Years	641.5	1022.97	279.43	31,892	33,860	36,610	20,458,718	34,637,764	10,229,868	
6-10 Years	433.21	1305.6	542.06	36,163	39,262	43,316	15,666,173	51,260,467	23,479,746	
11-19 Years	314.7	1443.36	885.79	38,004	42,536	50,004	11,959,859	61,394,761	44,293,477	
20 or More Years	465	1206.18	1156.51	38,645	43,426	53,238	17,969,925	52,379,573	61,570,326	
TOTAL	**3531.69**	**5895.41**	**3071.8**				**116,160,060**	**228,010,714**	**146,404,349**	**490,575,123**
Retirement and workers' compensation	0.1219						14,159,911	27,794,506	17,846,690	59,801,108
Insurance Benefits package: average	4,455	12498.9								55,682,600
TOTAL COST										**606,058,831**