(Note: This paper was published as Julian R. Betts and Anne Danenberg, "The Effects of Accountability in California," in Paul E. Peterson and Martin R. West, eds. **No Child Left Behind? The Politics and Practice of Accountability**, Washington, D.C.: Brookings Institution, pp. 197-212, (2003).

The Effects of Accountability in California

Julian R. Betts
Department of Economics
University of California, San Diego
and Public Policy Institute of California

and

Anne Danenberg
Public Policy Institute of California

This paper was prepared for the conference "Taking Account of Accountability: Assessing Politics and Policy," held at Harvard University in June 2002. This research also received financial support from the Public Policy Institute of California (PPIC). We thank Macke Raymond, William Howell, Paul Peterson and Martin West for helpful comments.

1. *Introduction*

California has been a relative latecomer to the national trend towards school accountability. In 1998, the state reintroduced a standardized statewide test and began to develop content standards. The third part of the reform was the introduction of an accountability system outlined in the Public Schools Accountability Act of 1999 (PSAA). The PSAA sets out a list of financial awards for schools, staff and students and sanctions for schools and administrators based on a school's ranking on the Academic Performance Index (API). The API, which is a weighted average of test scores of students at each school, has become the cornerstone of the legislation. Another key element of this system is the Immediate Intervention / Underperforming Schools Program (II/USP), which provides financial assistance to schools that roughly fall in the bottom half of the API distribution. [1]

Some important policy questions surround these reforms. First, how have school resources, such as class size and the qualifications of teachers evolved in response to the accountability reforms? Second, what is happening to achievement in California since these reforms? Third, given the large variations in student achievement among schools, how successful is the part of the PSAA that targets assistance towards low-performing schools, that is, the Immediate Intervention / Underperforming Schools Program?

In a sense II/USP is the most important aspect of the PSAA because previous work by Julian R. Betts and Anne Danenberg; Julian R. Betts, Kim S. Rueben and Anne Danenberg; and others has shown a shockingly high gap in student achievement among schools, with the socioeconomic status (SES) of each school's students being the single best predictor of test scores.[2] The II/USP program tackles this issue head on by targeting

additional financial resources at the lowest performing schools in the state. However, questions naturally arise concerning the effectiveness of these additional expenditures. In addition, the fact that participation in II/USP is voluntary raises important questions about whether the schools that most need help elect to participate. Because participating schools receive additional funding and face additional sanctions should they fail to improve, the question of ***which*** schools choose to participate in II/USP takes on new importance.

2. Has School Accountability Influenced the Resource Gap between Low- and High-Score Schools?

Trends among Schools at Different Achievement Levels

The goal of accountability is to have students at all schools performing at high levels. A necessary step then, is for low-performing schools to improve more quickly than high-performing schools. However, one potential side effect of school accountability is that the threat of sanctions may induce talented teachers to shy away from low-performing schools. Our ultimate goal in this section is to answer the first question above by testing whether the accountability legislation of 1999 has induced more highly experienced and educated teachers to move away from low-scoring schools. For this reason, we rank K-6 schools according to their mean scaled math scores for grade 5 and divide these schools into quintiles.[3] We compare selected school and teacher characteristics in the bottom and top fifth of schools to see whether there is any difference in these characteristics. More to the point, we examine whether gaps in

resources between high-scoring and low-scoring schools have widened or narrowed over the last few years. We also present the percentage of students receiving free or reduced price lunch for comparison purposes.

Table 1 shows that in low-score schools there are between approximately four and six times as many students on free or reduced-price lunch as there are in high-score schools. Clearly, the students in low-score schools are less economically advantaged than those in high-score schools. Moreover, the gaps between low- and high-score schools in this example have increased from 1995-96 to 2000-01.

Turning to other school and teacher characteristics, we find that in K-6 schools low-score schools have higher shares of novice teachers, teachers with at most a bachelor's degree, and who lack full credentials, while these schools have lower shares of teachers who have at least a master's degree and higher levels of experience. The gaps between low- and high-score schools have also widened for the share of novice teachers, teachers with at most a bachelor's degree, and teachers who lack full certification. The latter of these three widening gaps grew by over 13 percent, suggesting that despite efforts to equalize per pupil resources in California, uncertified teachers are concentrated disproportionately at elementary schools with the lowest test scores.

Gaps between average class sizes in low and high test-score schools have remained relatively static in K-6 schools, and have actually shifted in an unexpected manner. In 1995-96 low-score schools had higher average class size than high-score schools, yet in 2000-01 high-score schools had slightly higher average class size than low-score schools. This finding is somewhat surprising given the financial incentives to reduce class size in grades K-3 that were implemented by the state legislature in 1996,

along with the assumption that teachers would rather teach in smaller classes than in larger ones.[4] However, in a school with only one Grade 1 class with 25 students, it would be practically impossible to implement smaller classes—even under the best of circumstances.[5]

The discussion above makes abundantly clear that the gap in teacher qualifications between low-score and high-score schools has widened in all three grade spans between 1995-96 and 2000-01. Ideally, we would like to test in a statistical sense whether this widening inequality occurred mostly before the 1999-2000 school year, when school accountability really began in earnest in California, or whether the accountability system created a "shock" that widened the resource gap more quickly after it was introduced.

Using regression analysis techniques (not shown), we tested whether there was anything special about the post-PSAA years (1999-2000 and 2000-01) in terms of widening of the gap in teacher qualifications among these high-achieving and low-achieving schools. We find that in most cases there has been a fairly consistent trend over the entire period of widening resource gaps between lower-scoring and the highest-scoring schools. [6]

What about the question of whether these trends may have been aggravated by the passage of the PSAA in 1999 (or by some unobserved event occurring at the same time)? In almost all cases, we find that the legislation did not widen the resource gaps, and that in some cases, pre-existing trends of widening in the resource gap between the highest API quintile of schools and other schools actually began to *reverse* for at least one lower API quintile of schools.

Overall, then, the conclusions are quite clear-cut: there exists strong evidence of steadily widening gaps between the fifth of schools with the highest achievement and lower quintiles of schools between 1995-96 and 2000-01 in teacher education, experience and credentials. The trend toward smaller class size that began in 1996 may be partly responsible for these trends. However, there is no evidence that passage of the PSAA in 1999 exacerbated the trends.

3. The Immediate Intervention/Underperforming Schools Program

Overview

Figure 1 shows the cumulative distribution of grade 5 math test scores across California's schools, for non-Limited English Proficient students. The vertical axis shows the percentage of students attending schools at which average test scores are at or below the scores listed on the horizontal axis. The graph shows that the distribution of test scores has shifted upwards over time. [7] However, perhaps a more striking pattern is how test scores tail off among the lowest performing schools. This is true in all years, and is meaningful because the scores in the underlying test have been scaled psychometrically so that a 5-point difference is intended to capture the same absolute gap in performance anywhere along the test score distribution.

This "S-shaped" distribution of schools by test scores suggests that the bottom-performing schools lag considerably behind other schools in the state, and so deserve special attention. Therefore, the financial aid and potential penalties directed to the bottom-performing schools under the school accountability system arguably constitute

the most important part of the PSAA. We now devote considerable attention to this system, known as the Immediate Intervention/Underperforming Schools Program (II/USP). Below we refer to this mostly as "the Immediate Intervention program."

This program provides financial assistance to public schools that, roughly speaking, rank in the bottom half of API rankings each year. The first cohort of schools was selected in 1999.[8] Participating schools draft an Action Plan for reform in year one and implement the plan in years two and three, with an initial planning grant of $50,000 and grants of $200 per student during the implementation years. The first cohort of II/USP schools began their planning phases during the 1999-2000 school year, in conjunction with an external evaluator pre-screened by the state. In addition, California began implementing similar Comprehensive School Reform Demonstration Program (CSRD) grants in the same year with federal money. (For a list of differences between the regular II/USP program and CSRD see the conference version of this paper.) Below we will refer to both types of schools as participating in Immediate Intervention.

Both flavors of the Immediate Intervention program include not only financial carrots but also accountability "sticks." Within 12 months, the district must hold a public hearing to discuss initial progress during the implementation phase. If after 24 months the school does not meet its API growth targets, the state may provide one additional year of grants if the school shows some evidence of improvement. But if the school's API score has not improved sufficiently, the school will be subject to a sliding scale of state sanctions, culminating in possible state takeover of the school.

In a state with persistent resource inequalities across schools, this intervention program is notable for devoting additional financial resources to schools that lag behind

in the test-score rankings. These schools almost always enroll above-average numbers of disadvantaged students. The II/USP program reflects the recognition on the part of the California legislature that additional resources are needed for the very worst performing schools.

Nonetheless, we have many important questions about the efficacy of both the planning and implementation phases of the program. Most importantly, the question arises of whether the II/USP has started to improve student performance in participating schools. Additional questions relate to concerns that the funds are being distributed too widely, the potentially non-random nature in which schools have applied for the program, and the ease with which schools have planned for II/USP funding.

Preliminary Evidence on the Relative Performance Gains of II/USP and other Low-Performing Schools

Below we will attempt to assess the API score growth for schools participating in II/USP between spring 1999 and spring 2001. But it seems almost unfair to expect that II/USP reforms could bear fruit in two years, given that participating schools spend their first year planning the reforms that take root only in the second year. For CSRD schools, there is slightly more hope that we can evaluate the success of the reforms, as the first cohort of these schools began implementing reforms in 1999-2000, rather than 2000-01.

We ran simple models to explain the gains in API scores between 1999 and 2001 for schools that entered the II/USP in 1999-2000 and a sample of comparison schools. Each model included an indicator variable for schools that began to participate in II/USP

(including CSRD schools) in 1999-2000 as well as dummy variables for the school's grade span.

We tested whether, among the *sorts of schools that are apt to apply for reform programs such as II/USP,* II/USP participation causes the school's achievement to rise. Table 2 shows the coefficient on the II/USP indicator variable in the regression. The first column lists the treatment group, that is, the sample or subsample of II/USP participants included in the regression. The second column lists the control group, that is, the comparison group that provides the benchmark against which the II/USP schools are compared. Notably, our comparison group is schools that not only were eligible for II/USP but in addition applied and were ***not*** selected. This is a valid comparison group because these schools exhibited similar motivation in that they applied, but did not "win" the randomized selection process.

We find II/USP participation to be a positive and highly significant predictor of gains in school API scores over time. The predicted effect, about 17.5 points, represents roughly one third of the mean gain in API scores for schools in the sample over the period. This is a large effect. [9] [10]

Gaming the System? Self-Selection of Schools into the Immediate Intervention Program

A common criticism of school accountability systems is that schools will "game the system." In particular, schools that are likely to gain from participating in voluntary programs with the least effort are the most likely to apply, and these are not necessarily the schools that accountability systems primarily seek to help. To assess this issue, we now examine whether the voluntary nature of participation has caused schools to self-

select into the Immediate Intervention program in unintended ways that may undermine the effectiveness of the program.

These potential problems related to self-selection were not inevitable. It is easy to imagine how the Department of Education could have dictated which schools were to participate in the program, by starting at the bottom of the API distribution and working upward until the quota of 430 new schools per year was filled. Instead, a wide swath of schools was deemed eligible. Schools that were in the bottom 5 deciles of state rankings based on both spring 1998 and 1999 scores were deemed eligible for the first round of the program. Beginning in the 1999-2000 school year and in later years a school became eligible if it ranked in the bottom five deciles for that year, and in addition failed to meet the state-imposed target for growth in the API.

This voluntary nature of participation in the Immediate Intervention program potentially creates some severe incentive problems related to which schools volunteer to participate, and whether some schools actually have an incentive to let test scores slide. Some of the chief possibilities are as follows:

• Schools' API scores can rise or fall between years due to random variations in test scores that are beyond teachers' control. Are schools that had an "unlucky" year, with atypically low test-scores, more likely to apply for the program in the knowledge that they are more likely than average to improve in subsequent years?

• A special case of the above argument relates to the number of students in the school. Thomas J. Kane and Douglas O. Staiger argue that smaller schools are more likely than large schools to have unusually high or low test scores in any given year, because of greater random noise in their average test score.[11] Because II/USP schools

can dodge the worst penalties by showing improvement in one or two of their first three years, random chance works in the favor of smaller schools. In addition, there is a much more direct reason why smaller schools may be more likely to participate: principals at such schools may believe that reform is easier to implement when the number of teachers and students is relatively small.

There is another incentive mechanism, however, which works in the opposite direction. Schools that apply for Immediate Intervention must invest a large fixed cost of time to complete the application and in year one to plan how to spend the dollars that begin to flow in earnest in the second year of program participation. A large school can spread this fixed cost over more students (and teachers) than can a smaller school. Thus, simple financial incentives suggest the opposite to what we hypothesized above: larger schools may be more likely to apply for Immediate Intervention than small schools.

• The API growth targets for each school below the official target of a score of 800 (out of 1000) is 5 percent of the gap between 800 and its previous API score. This means that schools with particularly low performance must increase by more *in an absolute sense* than do schools that initially had higher scores. II/USP schools must meet their API growth targets within 24 months to avoid state sanctions. Does this combination of policies reduce the probability that the worst-performing schools will volunteer for the program? This seems to represent a potentially severe problem given the evidence that the very lowest performing schools fare disproportionately poorly on the state test.

To answer these questions, we ran ordinary least squares models of the probability that a school that was eligible for II/USP in fact applied for the program. We

performed these analyses separately for elementary, middle and high schools. We modeled this probability as a function of whether the API at the school had declined the previous year, total enrollment at the school, and the size of the API growth target set for the school by the state. The first of these controls is a dummy variable set to one if the API score declined in the most recent year, which we take as a proxy for what could well be "random noise" that affected a school's test scores in that year only. Our earlier hypothesis was that such schools may be particularly apt to apply for the program because the school's administrators rightly believe that test scores will naturally rebound or "regress to the mean."[12]

Table 3 shows the results when we model whether a school that was eligible to enter the II/USP in the 1999-2000, 2000-01 and 2001-02 school years applied. Most of the pessimistic hypotheses that we outlined above do not gain support from our regressions. Most importantly, schools at the bottom end of the API range, which therefore have the largest API growth targets assigned to them, in fact are *more*, not less, likely to apply to participate in II/USP. To any policymaker concerned about narrowing the achievement gap across schools, this appears to be very good news.

What about the size of the school as measured by enrollment? There was only one case in which we found support for the idea that smaller schools are more apt to apply, and this was in the model for elementary schools. However, the effect is very small (an increase in enrollment of 100 students is associated with a 0.75 percent drop in the probability of applying). Moreover, we find the opposite result for middle schools, where larger schools seem more likely to apply; in high schools enrollment also enters significantly, but is not significant statistically. Finally, we note that when we repeated

these models by individual years the negative enrollment effect among elementary schools was never statistically significant. We conclude that overall there is little support for the idea that smaller schools are more apt to apply because they realize that their scores are particularly likely to rebound.

The model also tests whether schools that recently experienced a decline in their API score would be more likely to apply for the program if administrators believed that the recent decline was a statistical fluke. [13] As the table shows, no significant link emerges between a recent decline in a school's API and its probability of applying for the Immediate Intervention program.

Overall, we have found little evidence that perverse incentives related to II/USP has created adverse patterns of self-selection among schools eligible for the program. By far the strongest and most consistent pattern was that schools with the lowest API's, which faced the largest state-mandated targets for growth in the API, were the most likely to apply for the program. This is a welcome sign that schools that need the most help are indeed the most likely to take advantage of program funding. [14]

Is the II/USP Program Sufficiently Focused on the Bottom-Performing Schools?

Our earlier discussion of the distribution of test scores makes patently clear that schools at the very bottom of the test score distribution are disproportionately in need of help. The Immediate Intervention program targets limited resources to a very broad range of schools in the bottom half of the API rankings. It would seem to make more sense to concentrate the funding for this program on the schools most truly in need, perhaps those in the bottom one or two deciles of performance. Not only would this

improved targeting do more to equalize student achievement, but it would also allow the state to invest proportionately more per student in each of the participating schools.

In this regard, it is useful to compare the state's decile rankings of schools that were eligible, that applied, and that were chosen for the program. Figure 2 illustrates for the 1999-2000 II/USP round. The eligibility criteria led to a relatively even distribution across the bottom four deciles and a drop-off in the fifth decile. In contrast, applicants from the eligibility pool did come disproportionately from the lower deciles. The lottery system through which schools were selected effectively evened out the distribution across deciles as shown. Thus, both the rules for eligibility and for selection into the II/USP program tended to draw fairly equal numbers of schools from all of the bottom deciles of school performance.

The disproportionate number of applicants from the bottom API decile was more exaggerated in the 2000-01 application process than in 1999-2000. However, in both that year and 2001-02, the selection procedure was more representative of the applicant pool than occurred in 1999-2000.[15] This most likely reflects the much smaller eligible pools in the two later years, where in *addition* to falling in the bottom half of the API rankings a school also had to have failed to meet its API growth target. This reduction in the number of applicants clearly allowed the Department of Education to select schools in a way that more closely represented the applicant mix.

Overall, both the eligibility rules and the selection rules, but especially the former, have allowed schools from across the achievement distribution to participate in the program. A more effective way of targeting funds towards the lowest performing schools would have been to select 430 schools each year from the bottom of the API rankings.

Ironically, a simple lottery from among eligible applicants without regard to their initial API ranking would also have focused the II/USP funds more narrowly on bottom-performing schools.

Perhaps in recognition of the fact that II/USP may have been distributing financial aid too widely instead of focusing more intensively on the bottom-performing schools, California legislators in 2001 passed Assembly Bill 961, which established the High Priority Schools Grant Program. This program will funnel twice the per capita funding stipulated under Immediate Intervention to schools most in need. The California Department of Education predicts that schools in the bottom decile will be the sole recipients of this additional funding. This innovation represents a partial move to a more dramatic focusing of new dollars on the schools in greatest need.

The Success of II/USP Schools' Planning Period and CSRD Schools' First Year of Implementation

II/USP schools undergo a one-year planning cycle. Given the multidimensional problems facing many failing schools, how successful have they been in developing coherent Action Plans within the space of 12 months? In the case of CSRD schools, they lack any planning period at all, but in return receive an additional year of implementation grants. This could work reasonably well, because the CSRD schools must select from a menu of existing reform approaches, rather than develop a plan from scratch. On the other hand, a custom-made plan for an II/USP school might work better than a one-size-fits-all type of reform.

The Department of Education conducted a review of the first year experiences of II/USP and CSRD schools, based in part on site visits and principal surveys to a subsample of schools. In this paragraph and the next, we summarize some of the key findings from Anne E. Just, Larry E. Boese, Rachel Burkhardt, Linda J. Carstens, Marsha Devine and Tim Gaffney.[16] Some principals complained that evaluators were aloof, unprepared, inexperienced, or unaware of the details of the II/USP program. But principals for the most part strongly endorsed the role played by the external evaluators chosen to help each school. For instance, 90% of principals responding stated that the external evaluator collaborated closely with the school and 84% would recommend their evaluator to other schools.

School site visits and the evaluation of the Action Plans by the Department of Education suggested less optimistic conclusions. Of the initial Action Plans submitted, the Department rejected 38% because they did not meet the stipulations of the II/USP program. Moreover, site visits revealed that both participant schools and their external evaluators complained about the timeline for planning, which in reality was compressed to about four months. In well over half of cases, the site visit teams inferred that schools were relying on the external evaluator to generate a reform plan instead of cooperating actively. [17]

Additional information on the first year of the II/USP plan comes from a one-day conference held by the California Education Policy Seminar and the California State University Institute for Education Reform in November 2000. Presentations by a panel of external evaluators and by panel respondents are reported by Kathleen Beasley and are summarized below. [18] The external evaluators' comments in general support the findings

of the Department of Education evaluation. Evaluators felt that many of these schools genuinely wanted to improve. They generally agreed that progress during the planning phase could have been greater with an extended period of planning time, especially for teachers. Several evaluators also stated that their schools had often made little progress at implementing reforms in curriculum at the classroom level. In his response to the panel of evaluators, John Mockler, at the time Interim Secretary for Education in California, reported on his own reviews of schools' Action Plans. One of his main concerns was that "There was no discussion of curriculum. And there was very little mention of standards, so it was difficult to tell what the plans were really talking about."

Overall, the qualitative reports discussed above suggest some teething pains for the implementation of II/USP, and in some schools a lack of focus on the central issues, but that on the whole participating schools are taking the Immediate Intervention reforms quite seriously.

4. *Conclusion and Policy Discussion*

This paper extends the initial evaluation of California's school accountability program in Betts and Danenberg along two lines of inquiry. Our earlier paper suggested a widening gap in school resources between low-achieving and high-achieving schools. The current paper extends this analysis over a six-year period in order to determine whether the growing inequality in school resources, especially related to teachers, has continued well into the "post-accountability" period beginning with the 1999-2000 school year. Second, the paper studies the Immediate Intervention program of financial assistance and potential sanctions that the state has targeted at the lower-performing

schools. Given the gulf in achievement between these schools and top-achieving schools, this aspect of the Public School Accountability Act is to many observers the most important component of the accountability reforms.

First, we find that there has been an overall trend toward fewer highly educated and experienced teachers and more novice and uncertified teachers. We also find that the gaps between low- and high-score schools have widened for most measures we examine, with the exception of average class size in K-6 schools, which has equalized almost universally—probably due in large part to legislation to reduce class size implemented in 1996. Regression analyses support these findings after controlling for student characteristics that might affect a school's selected resource levels. However, we find that there is no evidence that passage of the PSAA in 1999 exacerbated these trends.

We briefly examine trends in test scores in the state, and note a steep and persistent drop-off in student achievement in the worst-performing schools. The plight of low-performing schools provides powerful motivation for our analysis of the Immediate Intervention (II/USP) program, which is targeted at low-performing schools. We examined whether the first cohort of II/USP schools outperformed their peers in terms of student achievement gains between spring 1999, before entry to the program, and spring 2001. When comparing II/USP schools with schools that applied for II/USP but which were not randomly selected to participate, we found that the II/USP schools' achievement rose substantially more than did achievement at schools that were not randomly selected. Among the sorts of schools that applied for the program, we find surprisingly strong evidence suggesting that Immediate Intervention works. These results, although

promising, are preliminary; additional work will need to be done in future years to confirm that II/USP has indeed produced gains among participating schools.

The second issue we examined was whether the voluntary nature of II/USP has led to schools "gaming the system," in the sense that the schools that are likely to escape the program's sanctions while exerting the least effort are the most likely to apply for II/USP. We tested several variants of this idea, and found little evidence of gaming the system. The strongest pattern we identified was that schools with the lowest test scores were the most likely to apply for II/USP, which is heartening given that these are clearly the schools in most need.

Third, we asked whether the state's guidelines for determining II/USP eligibility and for selecting schools from among the applicant pool both worked to spread program funding too widely among the bottom half of schools. We did find evidence that both the eligibility and selection criteria worked to distribute funds across all five of the bottom deciles, and concluded that the state might consider reforms to target its resources more narrowly toward the lowest-performing schools. However, the legislature did pass a bill in 2001 that will provide additional funding for schools at the very bottom of the achievement distribution.

Finally, our review of two qualitative studies suggested that the implementation of II/USP in its first year (1999-2000) ran into several problems. School administrators complained that they had not had enough time to plan how to spend their II/USP grants, and several outside observers stated that the implementation plans devised by schools and their contractors very often sidestepped the key issue of curriculum and how the plan would link to California's subject content frameworks.

Thus, the II/USP program despite some start-up problems and a lack of focus on the schools at the very bottom of the achievement distribution, is off to a reasonably strong start, and we have uncovered preliminary evidence that the program may be helping participating schools quite substantially.

Overall, the Public School Accountability program is now well underway, and trends in California's schools suggest both a rising tide of achievement and a narrowing in the achievement gap among schools, at least in lower grades and especially in reading.[19] Areas of concern for policymakers should include slow progress in student achievement in the higher grades, the relentless trends that have widened the gap in teacher qualifications between low- and high-achieving schools, and a need to find better targeting of financial resources to the schools that lag furthest behind.

Of course, there are other technical, financial, and political challenges that our paper has not had space to address. On the political side, there is a small but rapidly growing opposition to the state test and the accountability system more broadly, with opponents including the powerful California Teachers Association. Furthermore, in 2001 legislators passed AB 961, which provides for alternative penalties—often called ***soft sanctions***—to those in the PSAA. In addition, new legislation proposed in 2002 in SB1310 provides more flexibility in imposing sanctions on failing schools in California.[20] However, there are actions that can be taken at district level, such as reassigning principals or converting the school to a charter school (ostensibly under new administration and a largely changed staff), before the state becomes involved. It remains to be seen what these factors might portend for the future path of school accountability in California.

Table 1: Comparison of average school and teacher characteristics in cross sections of high test-score schools and low test-score schools for grade 5 math, K-6 Schools, from 1995-96 to 2000-01.

	1995-1996		2000-2001		
School Characteristic	1997-98 Low-score schools	1997-98 High-score schools	2000-01 Low-score schools	2000-01 High-score schools	Gap change 1995-96 to 2000-01
% Free/reduced lunch	81.7	19.7	82.8	16.1	4.7
Avg. Class Size	30.4	30.1	22.0	22.4	-0.7
Avg. Teacher Experience.	12.3	15.1	10.6	13.0	0.3
% at most BA	21.7	10.7	32.2	14.5	6.7
% at least MA	27.3	34.8	22.6	32.3	-2.2
% low exp. (0-2 yrs)	14.2	8.6	19.4	12.5	1.2
% lacking credential	3.0	0.7	20.5	4.9	13.4
Number of observations	738	747	862	866	—

SOURCE: Authors' calculations from California Department of Education datasets.

Table 2 Estimated Impact of Participation in II/USP Beginning in 1999-2000 on API Gains between 1999 and 2001, from Regression Models

Treatment Group	Control Group	Coefficient and (Standard Error)
Tests that from among the types of schools that are likely to apply to such programs, II/USP participation boosts student achievement		
All schools that entered II/USP in 1999-2000	All schools that were eligible for II/USP in 1999-2000 and applied but were not selected to participate	17.5**
		(2.4)

Note: All regression models included controls for grade-span of school in addition to dummy variable for II/USP participation beginning in 1999-2000. **indicates significant at 1%.

Table 3 Ordinary Least Square Models of Probability that Eligible Schools Apply for II/USP

Independent Variables	Grade-Span K-6	6-8	9-12
API Fell	0.053	-0.006	0.057
	(0.031)	(0.12)	(0.049)
Enrollment	-0.075*	0.075*	0.024
('000s)	(0.031)	(0.032)	(0.017)
Target API	0.015**	0.026**	0.030**
Growth	(0.002)	(0.005)	0.006
Adjusted R-squared	0.108	0.101	0.067
Sample Size	3196	899	802

Note: Numbers in parentheses are standard errors. *indicates significant at 5%, **indicates significant at 1%. Years included: 1999-2000 through 2001-2002. Other regressors not shown include a constant and year dummies. API refers to the Academic Performance Index, which is the state's overall measure of student achievement in each school. Target API Growth is the number representing the target set by the State for each school's growth in API. Lower scoring schools are given a larger growth target.

Figure 1

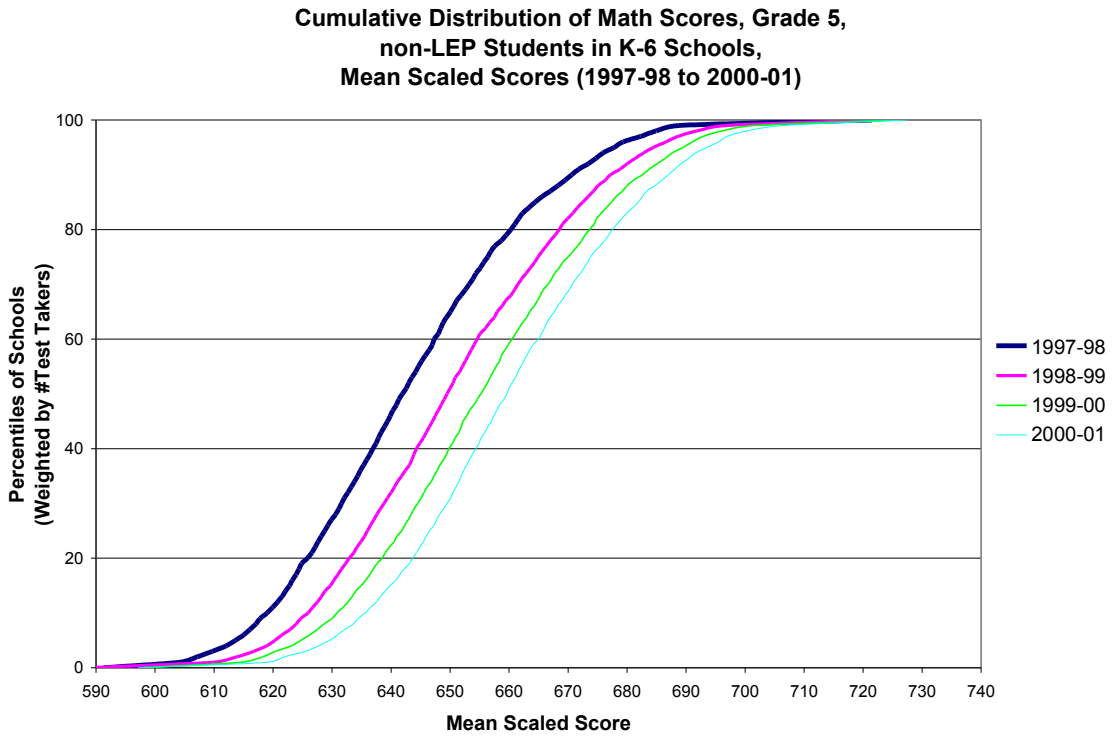
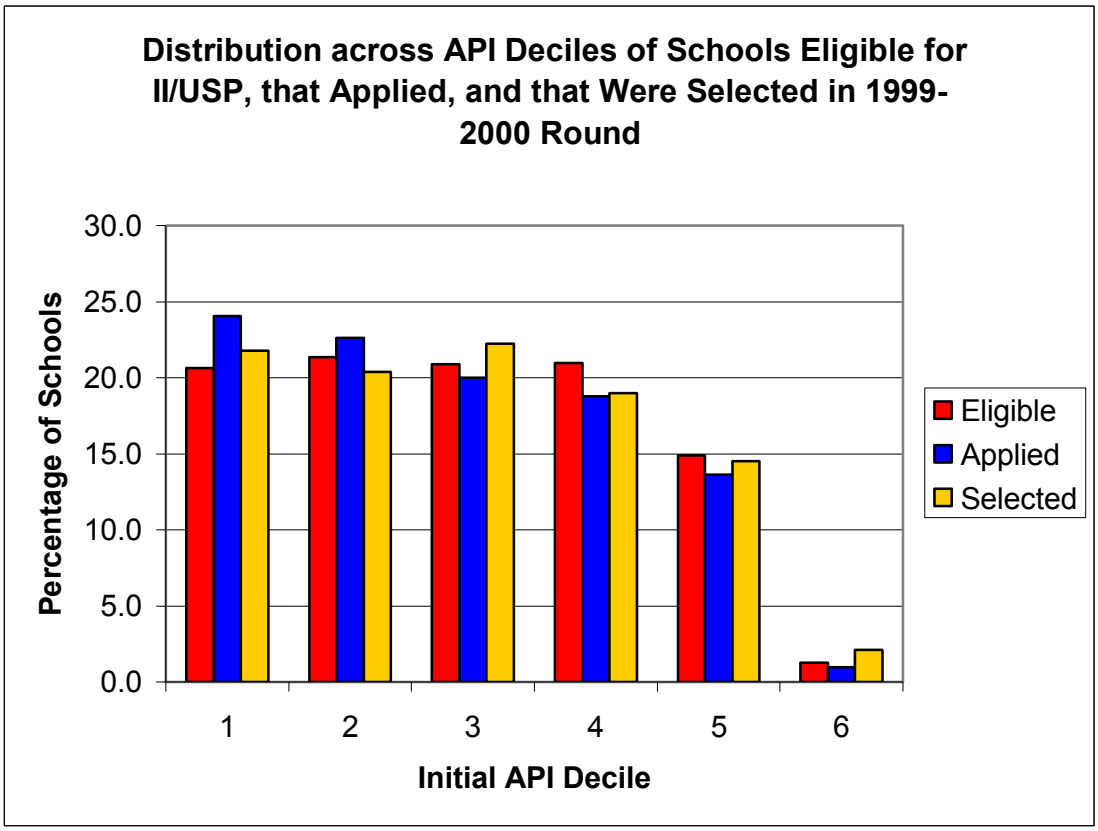

Cumulative Distribution of Math Scores, Grade 5,
non-LEP Students in K-6 Schools,
Mean Scaled Scores (1997-98 to 2000-01)
Percentiles of Schools
(Weighted by #Test Takers)
100
80
60
40
20
0
590
600
610
620
630
640
650
660
670
680
690
700
710
720
730
740
Mean Scaled Score
1997-98
1998-99
1999-00
2000-01

Figure 2


Distribution across API Deciles of Schools Eligible for II/USP, that Applied, and that Were Selected in 1999-2000 Round
Percentage of Schools
30.0
25.0
20.0
15.0
10.0
5.0
0.0
1
2
3
4
5
6
Initial API Decile
Eligible
Applied
Selected

[1] For a more detailed description of the PSAA, see the conference version of this paper at www.ksg.harvard.edu/pepg/index.htm and Julian R. Betts and Anne Danenberg, "School Accountability in California: an Early Evaluation," in Diane Ravitch (Ed.), *Brookings Papers on Education Policy - 2002*, (Washington, DC: Brookings Institution Press, 2002).

[2] Julian R. Betts and Anne Danenberg, "School Accountability in California: an Early Evaluation," in Diane Ravitch (Ed.), *Brookings Papers on Education Policy - 2002*, (Washington, DC: Brookings Institution Press, 2002) and Julian R. Betts, 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)

[3] See the conference version of this paper for overall means in each grade span, and for middle and high school high- and low-scoring schools and change in the resource gaps between such schools. For the 1995-96 and 1996-97 years, when there was no statewide test, we rank schools using the spring 1998 test scores, while in later years we rank schools based on their test scores in that year.

[4] The state legislature passed the Class Size Reduction (CSR) initiative in 1996, which gave schools a strong incentive to reduce class size to 20 first in grades 1 and 2 and starting in 1997 in kindergarten and grade 3 as well. It is likely that the CSR initiative affected the labor market for teachers. As more experienced teachers switched to teaching K-3 with class sizes of 20 or less in 1996-97 and later years, school districts had to hunt far afield for new teachers to replace them. Furthermore, minority and low-income students received a disproportionately higher share of the less experienced teachers who entered as replacements. This pattern will of course tend to widen the resource gap between high- and low-scoring schools, quite independently of any separate effect of accountability. Christopher Jepsen and Steven Rivkin, *Class Size Reduction, Teacher Quality, and Academic Achievement in California Public Elementary Schools,* (San Francisco: Public Policy Institute of California, 2002).

[5] See Julian R. Betts, 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) for a discussion of the contributions of CSR and collective bargaining contracts to this pattern.

[6] Note that Table 1 for the most part identified a different set of low-score and high-score schools each year. Here, we instead keep the makeup of these groups fixed, based on the 1999 API. In addition, we use the overall API given our evidence that using reading or math scores, two of the most important subcomponents of the API, yield similar trends in resource gaps. We ran regression models for each of the school resource measures listed in Table 1 as a function of the school's test score quintile, time trends specific to each test-score quintile of school, a dummy variable for the period beginning in 1999-2000 (our post-PSAA period), interactions of this dummy variable with the API quintiles, and characteristics that might independently influence the resources the school receives through categorical programs—the percentage of students receiving free/reduced-price lunch, the percentage of Limited English Proficient (LEP) students, and the enrollment at the school. See the conference version of this paper at www.ksg.harvard.edu/pepg/index.htm for the coefficient values and the full regression results.

[7] For a detailed analysis of trends in California test scores please see the conference version of this paper.

[8] In fall 1999, any schools performing below the 50th percentile on the STAR test in *both* spring 1998 and spring 1999 were eligible to participate in the II/USP, and the state's Academic Performance Index (API) determined subsequent cohorts' eligibility. In September 1999, a first cohort of 430 schools of over 3,000 thousand eligible schools was chosen for external evaluation. Second and third cohorts of 430 schools were chosen based on the 2000 API and the 2001 API, respectively. The PSAA text describes a random selection process inside each of fifteen categories (or cells) of schools—deciles one through five for each of elementary, middle, and high schools. However, in reality the selection process depends partially on which schools apply to the program, and which schools are already in the program. In theory, all II/USP-eligible schools would eventually enter the program as the eligible pool of schools shrinks each year after a new cohort is chosen.

[9] In regression results not shown, we could find no evidence of any difference in improvement rates between CSRD and regular II/USP schools.

[10] We also estimated a model in which the comparison group was *all* schools that were eligible to apply, even if they chose not to apply. In this case we found no significant effect of participation in the program.

It appears that there was considerable selectivity bias in this model, because eligible schools that did not apply appear to be quite different from those that did apply. See the conference version of the paper for more detail.

[11] Thomas J. Kane and Douglas O. Staiger "Volatility in School Test Scores: Implications for Test-Based Accountability Systems", in Diane Ravitch (Ed.), *Brookings Papers on Education Policy - 2002*, (Washington, DC: Brookings Institution Press, 2002) and Thomas J. Kane and Douglas O. Staiger "Racial Subgroup Rules in School Accountability Systems," (in this volume), a version of which can be found at www.ksg.harvard.edu/pepg/index.htm.

[12] We also ran probit models to take account of the binary nature of the dependent variable, and in addition ran simple correlations between whether schools applied to II/USP and each of the explanatory variables to guard against collinearity. Both of these methods suggested similar conclusions to those we report in the main text.

[13] For the first year, 1999-2000, we set the dummy for whether the API had declined equal to zero, because schools had been issued an API only once by that point. However we also include year dummies to account for variations in this variable definition across years. Regressions that include only the later two years of the three years included in Table 3 produced similar results, as did separate models estimated for each year.

[14] We also examined a fourth possible perverse incentive, namely, the incentive that some schools that ranked slightly above the cutoff for II/USP eligibility would have to let their scores slip in the subsequent year in order to qualify for funding. We found no evidence that this occurred.

[15] See Table 4.2 in the conference version of the paper for the 2000-01 and 2001-02 distributions of these schools.

[16] Anne E. Just, Larry E. Boese, Rachel Burkhardt, Linda J. Carstens, Marsha Devine and Tim Gaffney, "Immediate Intervention/Underperforming Schools Program (II/USP): How Low Performing Schools in California are Facing the Challenge of Improving Student Achievement," California Department of Education, Division of Policy and Evaluation, Research Summary, May 2001.

[17] See the conference version of this paper at www.ksg.harvard.edu/pepg/index.htm for a summary of the investigators' analysis of CSRD schools, which was overall quite similar.

[18] Kathleen Beasley, "Immediate Intervention/Underperforming Schools Program: First Year Overview," The California Education Policy Seminar and the California State University Institute for Education Reform, November 2000.

[19] See the conference version of the paper for detailed documentation on these trends in achievement.

[20] Jim Sanders noted in a recent *Sacramento Bee* article ("School Takeover Threat Relaxed" May 1, 2002), that it seems unlikely that the state will take schools over.