Available online at www.sciencedirect.com
ScienceDirect
Economics of Education Review 27 (2008) 211–220
Economics of Education Review
www.elsevier.com/locate/econedurev

Gains in standardized test scores: Evidence of diminishing returns to achievement

Donna Driscoll, Dennis Halcoussis*, Shirley Svorny

Department of Economics, California State University, Northridge, 18111 Nordhoff St. Northridge, CA 91330, USA

Received 25 June 2005; accepted 27 October 2006

Abstract

In many states, standardized tests are used to hold schools accountable for student academic achievement. To motivate improvement in test scores, financial awards are given to teachers and administrators in schools that show the greatest gains. However, failure to adjust for initial conditions may put awards out of the reach of some schools and fail to produce the desired incentives. In this paper, we examine factors that influence gains in test scores using school-level data from California (1999–2003). We find evidence that validates the existing California award system—test scores improved the most for weak schools.

JEL classification: I22; I28

Keywords: Educational economics; Educational finance; Resource allocation

1. Introduction

In many states and school districts, standardized tests are used to hold schools accountable for student academic achievement. To motivate improvement in test scores, financial awards are sometimes given to high achieving schools, teachers and administrators. In setting awards, program administrators must decide whether goals and rewards should be the same across all schools. Optimally, awards should be structured to distinguish between test results that reflect the actions of school administrators and teachers rather than the characteristics of the student population.[1]

If adjustments are not made for characteristics of the student population in setting the award structure, the difficulty of meeting goals in some settings may discourage teachers and district officials from efforts to improve academic performance. As a result, efforts to improve academic performance through standardized tests may fail.

Although there have been studies that examine the characteristics of schools that lead to higher or lower test scores, little is known about the factors that contribute to gains in test scores over time.

*Corresponding author. Tel.: +1 818 677 4566; fax: +1 818 677 6079.
E-mail address: dennis.halcoussis@csun.edu (D. Halcoussis).

[1]Most studies find socioeconomic characteristics and parents' education to be the principle determinants of education outcomes (Hanushek, 1997).

0272-7757/$ - see front matter
doi:10.1016/j.econedurev.2006.10.002

An award system in California was predicated on the premise that there are diminishing returns to achievement. Diminishing returns imply that high initial scores would be generally associated with smaller measured gains over time. Under the California award system, schools with high initial scores did not need to see as great an improvement as other schools to qualify for awards.

In hope of providing evidence as to what is an effective award system, we examine the various factors that influence gains in test scores over time. Using school-level data from California over the period between academic years 1999 and 2003, we test the premise of the California award system—that high initial scores are associated with smaller gains. We also test to see if there is a relationship between other characteristics of schools and measured gains in test scores over time. These characteristics include family income, parent's education, class, school and district size, competition from private schools, and use of emergency credentialed teachers.

We find evidence that validates the California award system—test scores improve the most for weak schools. Two other characteristics of schools show a statistically significant relationship to test score gains over time—competition from private schools (negative) and family income (positive).[2] Although the exact relationship will vary across school districts and over time, states designing or revising a school-based reward system might want to take these results into consideration.

2. California goals and rewards

In California, the state's academic performance index (API) is used to measure student achievement. The goals set for schools in California assume diminishing returns to academic effort; California uses growth targets that are lower for schools with higher initial test scores. As explained on the California Department of Education's web site, "The annual growth target for a school is five percent of the distance between a school's API Base and the statewide performance target of 800" (with a minimum growth target of one point).[3] When funding is available, schools that meet the growth targets are eligible for monetary awards. The growth targets are also used to select California Distinguished Schools.[4]

In 1999–2000, $227 million was appropriated to the Governor's Performance Award Program. Schools that met the selection criterion—including meeting or exceeding the growth targets—earned $150 per student tested, to be used at the discretion of the "site governance team/school wide council representing major stakeholders" (ratified by the local school board). The same year, $100 million was appropriated to the Certificated Staff Performance Incentive Awards Program. One thousand certificated staff members in schools with the largest API growth earned $25,000 each, with smaller awards going to staff in schools coming in second and third.[5] Also in 1999–2000, $350 million was awarded through the Schoolsite Employee Performance Bonus Program ($800 per staff member, half going to the staff member, the other half for school use).[6] Clearly, when funds are available, much is riding on the growth targets selected.[7]

3. Measuring academic performance over time

Attempts to measure changes in academic performance over time are subject to several complications (Kane and Staiger, 2002a). First, year-to-year comparisons may be distorted by changes in the student population or by outside events. The arrest of a student or an early morning earthquake could shift student focus from the test, affecting a school's overall performance (Kane and Staiger, 2002a).

Also, attempts to rank schools by relative improvement are stymied by the fact that sample size influences measured test score gains over time. The smaller the classroom or the smaller the school, the more variation in test scores from one year to the next. Any ranking of schools on the basis of

[2]Hoxby (1994, 1988) examine competition between private and public schools. Hoxby (2003) is also of interest here.

[3]http://www.cde.ca.gov/ta/ac/ap/apidescription.asp (viewed March 14, 2006).

[4]http://www.cde.ca.gov/ta/ac/ap/apidescription.asp (viewed March 14, 2006).

[5]3750 certificated staff members were awarded $10,000 each and another 7500 were awarded $5000 each. This information is taken from the California Department of Education web pages, http://www.cde.ca.gov/ta/sr/ci/progdesc0001.asp and http://www.cde.ca.gov/ta/sr/gp/program00.asp.

[6]California Educator, vol. 5, Issue 5, February 2001, http://www.cta.org/CaliforniaEducator/v5i5/Feature_2.htm.

[7]Funding in 2000–2001 was limited to the Governor's Performance Award Program; $144 million was apportioned among the winning schools. No funds have been allocated to the programs since that time.

year-to-year gains will put the smallest schools disproportionately at the top (and bottom) of the list (Kane and Staiger, 2002a).

Although these concerns cannot be eliminated altogether, researchers suggest that using multiple grades in calculating year-to-year improvements and observing change over a relatively long period of time is the best way to deal with these issues (Kane and Staiger, 2002b). The California API is well suited to addressing these issues. First, it is a weighted index of the scores in several grades in each school, limiting the variability likely to result from small sample sizes. Second, improvements in test scores can be measured over a 4-year period.

The API is a numeric index that ranges from 200 to 1000, reported annually (in January). Initially based solely on the Stanford Achievement Test, Ninth Edition (Stanford 9) test scores, the state switched in 2002 to the California Achievement Test, Sixth Edition (CAT/6). In addition, between 1999 and 2003, the weights on these "norm-referenced tests" have declined as the state has factored in other indicators, such as the California Standards Test (first included in 2001), and the California High School Exit Exam (included in 2002).[8] The state has taken steps to ensure "that the statewide average API does not fluctuate solely as the result of adding new API components," adding "Scale Calibration Factors" to adjust scores by grade span.[9] Only those students who were enrolled in the district the prior year are included when a school's API is calculated.

Over time, the state plans to incorporate indicators of graduation rates and attendance rates into school level API scores. The federal Public Schools Accountability Act of 1999 requires that at least 60% of the API reflect test scores.[10]

4. Empirical tests

To assess the influence of initial scores and other factors on test score gains, we employ ordinary least-squares regression techniques to examine school level data. The sample includes 4174 elementary schools, 768 middle or junior high schools, and 745 high schools. We examine each level separately. The dependent variable is the change in school-level API scores between 1999–2000 and 2003–2004. Independent variables include measures of initial test scores, parents' wealth and education, efforts toward class size reduction, credential status of teachers, school size, district size, and competition (from private schools).

Diminishing returns to learning would imply that high-scoring schools would show the smallest gains over time. The alternative hypothesis is that schools with relatively high initial scores experience bigger gains over time. This could be the case if observed high initial scores suggest any kind of an advantage in preparing students for the form and content of standardized tests. It has been said that much of the improvement in test scores in the several years following the introduction of a new test reflects increasing teacher/student familiarity with the tests and test content, rather than actual achievement.[11] If this is the case, schools with high initial test scores may be best positioned to take steps to raise test scores.

We include measures of other characteristics of the test-taking population to further explore the circumstances that affect achievement gains over time. First, we include measures of parental income and education; these variables are consistently among those that have been shown to influence test score levels. Here we examine whether they impact gains as well. The percent of the school's students receiving free or reduced price meals measures the extent of poverty in a school. Median household income in dollars (based on school zip codes) adds a measure of the distribution of income. To capture parental education, we use the percent of a school's parents that have graduated from college.[12]

[8]For more details, see the California Department of Education's "Information Guide" for the "2003 Academic Performance Index Base Report" (March 2004) available at http://www.cde.ca.gov/ta/ac/ap/documents/infoguide03b.pdf (viewed April 23, 2006).

[9]The reader is referred to the California Department of Education web pages, http://www.cde.ca.gov/ta/ac/ap/expnotes0Xb.asp where X is 1, 2, or 3 for 2001, 2002, and 2003, respectively (viewed April 23, 2006)

[10]See the California Department of Education's "Information Guide" for the "2003 Academic Performance Index Base Report" (March 2004) available at http://www.cde.ca.gov/ta/ac/ap/documents/infoguide03b.pdf (viewed April 23, 2006).

[11]Koretz (1996).

[12]Household income, population density, and the private school data are from US Census Bureau and are measured for calendar year 2000 (US Census Bureau, 2000). Zip codes and location names were used to match these data with the appropriate schools. The remaining variables are all in the California API database and are for academic year 1999–2000,

As in Driscoll, Halcoussis, and Svorny (2003), we argue that these socioeconomic measures not only reflect the characteristics of the student population but teacher characteristics as well. The premise is that teachers sort themselves so that teacher quality is not independent of the student population of the school.

In addition to measuring relative wealth, median household income serves as a proxy for homeownership. Interest in property values may reinforce parents' propensity to encourage schools to achieve significant gains in test scores over time. Boghossian (2004) reports that Southern California parents are willing to pay a premium to live in neighborhoods where API scores exceed 900 points. In 2000, a home in Los Angeles County (served by the much maligned Los Angeles Unified School District) sold for $100,000 less than a similar home in a similar neighborhood across the Ventura County line (served by the Los Virgenes School District).[13] Where homeownership is high, parents have strong incentives to work toward test score improvements, even if their own children will be out of the specific schools when the achievements are reported.

The regressions also include measures of class size reduction from 1999–2000 to 2003–2004 as an independent variable. Specifically, the measures indicate the decrease in the average number of students in a class, so that a decrease is expressed as a positive value. This allows the coefficient to represent the change in API gains from a decrease in class size, ceteris paribus (most schools experienced decreasing class sizes during this period). In 1996, the California legislature passed Senate Bill 1777 and appropriated funds to encourage class size reduction in California public schools. Districts were given an extra $650 per pupil for students in classrooms with 20 or fewer students. Schools were required to reduce all first grade classes first, then all second grade classes, then either all kindergarten or all third grade classes (Bohrnstedt and Stretcher, 2002). Schools across the state took steps to reduce class size.

Including the decrease in class size in each regression allows a test of the relationship between class-size reduction and test score gains. The cost of the program suggests that legislators must have had high expectations with respect to improvements in education. We are able to test one aspect of this here.

One consequence of class size reduction was a dramatic increase in the demand for teachers in the state. Many schools were forced to rely on emergency credentialed teachers (individuals without teaching credentials, but making progress toward a degree). The distribution of emergency credentialed teachers was not uniform across California schools. The increase in demand for teachers created additional opportunities at the best-reputed schools. Many credentialed teachers transferred to those schools. This left the least-sought-after schools with vacancies and resulted in substantial reliance on un-credentialed teachers in these schools.[14] To control for any possible impact on test score gains, we include an additional explanatory variable in our regressions—the percent of teachers working under emergency credentials in every school.

Shkolnik et al. (2002) examined the impact of the credential status of teachers on test score gains from 1999–2000 to 2000–2001 for six large school districts in California (6% of total public school enrollment in the state). They did not find credential status to be statistically significant when student and classroom characteristics were controlled for.

The school size (the number of students enrolled) is included in the regressions as an independent variable to control for difficulties facing smaller schools in implementing class size reduction. In smaller schools "grade groups are not easily divisible into groups of 20" (CSR Research Consortium July 27, 2002 Press Release) stymieing efforts toward class size reduction. The addition of this variable also allows us to examine whether school size has an independent influence on achievement gains. School size has been shown to influence academic achievement (Cotton, 1996). It may also influence the ability of a school to achieve academic gains over time. Larger schools may permit better matches between students and their learning needs, facilitating improvements in test scores. On the other hand, more individual attention to students in smaller schools might motivate achievement gains.

(*footnote continued*)
except for the decrease in class size and the change in API, which are both measured from 1999–2000 to 2003–04.

[13]Conversations with real estate brokers.

[14]Jepsen and Rivkin (2002) found that, in some cases, gains from smaller class size were offset by a decline in average teacher quality.

Table 1
Means and standard deviations

Variable	Elementary schools $n = 4174$		Middle schools $n = 768$		High schools $n = 745$	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Change in API	102.08	60.75	56.42	44.64	40.87	39.32
1999 API	630.19	136.79	631.46	126.10	621.08	107.59
District size	64,925.05	154,465.22	50,757.73	133,276.67	49,708.61	131,751.98
School size	372.41	157.64	779.87	405.64	1,086.49	607.60
Meal	42.24	32.96	33.85	29.13	21.87	23.64
College grad	21.28	14.93	24.90	11.32	26.62	10.49
Private school	0.11	0.090	0.11	0.088	0.11	0.089
Pop density	7.52	8.31	6.91	7.95	6.28	7.60
Median HH income	49,299.58	19,553.26	50,111.41	19,787.12	49,612.21	19,619.99
Emergency credential	11.79	12.16	12.86	11.28	11.37	8.45
Decrease in class size grades K-3	0.58	1.92				
Decrease in class size grades 4–6	0.18	3.57	0.29	3.59		
Decrease in class size grades 7–12			0.27	3.15	1.23	2.57

Several additional variables attempt to identify circumstances that lead to larger achievement gains over time. The first is school district size, also measured by student enrollment. Driscoll et al. (2003) found that school district size has a negative effect on student performance. However, financial awards for achievement gains might have motivated district administrators to resolve observed weaknesses in programs. If so, we would expect larger academic achievement gains in schools in large districts. Alternatively, the bureaucracy of large districts may stand in the way of efforts to improve teaching or even to familiarize students with the tests and facilitate test taking, resulting in smaller gains over time. To isolate the impact of district size on performance gains, we include population density (per acre) as a control variable in the regressions (see Driscoll et al., 2003).

The portion of students in each school's territory that are attending private school is included as an independent variable. Competition with local private schools might motivate public schools to expend more effort to improve their achievement scores.[15] Driscoll et al. (2003) found that, all else constant, API scores were higher in public elementary and middle schools in communities with relatively high shares of students in private schools. Here, we add the private school variable to observe any association with test score gains over time.

We lack measures of non-labor inputs, including capital equipment, facilities and textbooks. Because we are looking at test gains over time, the lack of controls for such inputs is less of a problem than it might otherwise be. Driscoll et al. (2003) argue that the lack of strong empirical connections between these variables and student performance in previous studies (see Hanushek, 1986) minimizes concerns over the lack of such controls in regressions on student achievement.

[15]See Testa and Sen (1999) and Hoxby (1988, 2000, 2003).

5. Results

The means and standard deviations for all variables are listed in Table 1. Table 2 shows means for the same variables shown in Table 1 but here, for each type of school, separate values are presented for schools that scored in the lower third, middle third, and upper third for their school type (based on the 1999 API score.) Looking at the first row of Table 2, for each type of school, the average increase in the API score from 1999–2000 to 2003–2004 declines for schools with a higher 1999 API score.

The regression results are shown in Table 3. In each case, the dependent variable is the change in the school's API score between 1999–2000 and 2003–2004. The 1999 base period API scores have a negative and significant (at the 1% level) effect on gains over time. Each initial point lowers the expected gain by 0.39 points for elementary schools, 0.28 points for middle schools and 0.23 points for high schools. This suggests diminishing returns in achievement dominate any superior ability of high-scoring schools to manipulate scores by familiarizing

Table 2
Means for schools sorted by 1999 API score

Variable	Elementary schools			Middle schools			High schools		
	Lowest third of API	Middle third of API	Top third of API	Lowest third of API	Middle third of API	Top third of API	Lowest third of API	Middle third of API	Top third of API
Change in API	160.88	96.72	48.64	97.60	49.22	22.43	70.33	36.38	15.92
1999 API	476.00	627.79	786.78	490.24	630.73	773.41	501.61	621.14	740.48
District size	117,271	63,874	31,129	98,908	33,155	20,208	91,725	38,008	19,439
School size	417.83	352.62	346.79	877.75	721.67	740.19	1189.83	1048.87	1020.92
Meal	66.83	43.71	16.18	53.86	35.34	12.35	37.81	20.68	7.12
College grad	10.47	19.85	33.51	15.41	24.50	34.79	16.01	27.55	36.29
Private school	0.09	0.10	0.14	0.09	0.11	0.13	0.10	0.11	0.12
Pop density	10.21	6.86	5.49	9.76	6.01	4.95	8.96	5.38	4.50
Median HH income	37,603	46,418	63,879	38,115	46,753	65,465	38,898	47,831	62,114
Emergency credential	18.80	12.09	4.48	18.49	14.08	6.01	14.60	13.27	6.27
Decrease in class size grades K-3	0.83	0.68	0.23						
Decrease in class size grades 4–6	0.46	0.19	−0.11	0.41	0.46	0.00			
Decrease in class size grades 7–12				0.42	0.16	0.23	1.10	1.37	1.22

students with the test. An elementary school with a base API score 100 points lower than a similar school could expect an API gain 39 points larger. To put this in context, these 39 points are larger than the California gain target for the average elementary school over the period of this study.[16] For middle schools, a school with a base API score 100 points lower than a similar school could expect an API gain that covers 84% of the average middle school's API gain target for the period of this study. For high schools, the figure is 63%.[17]

In addition to the results shown here, the model is also tested by replacing the initial test score with its natural logarithm as an independent variable. This variation of the regression model explicitly takes into account the "diminishing returns" aspect of higher initial scores on the change in API scores. The results do not change substantially; they are not presented here.[18]

Household income is positively associated with improved test scores at each level of education. Schools with similar base scores but wealthier families see larger improvements over time. These results are statistically significant but the impact is smaller than that of the base 1999 API score. An elementary school that has a median household income \$10,000 higher than a similar school could expect an API gain that is 4 points higher. This is approximately 12% of an average elementary school's California gain target for the period of this study.[19] Parents' education has a positive effect on test score gains as well, but only at the elementary level.

[16]Following the state of California's formula for the annual gain target, calculating 5% of the distance between 800 and the mean elementary school 1999 API score (630.19) gives a target of 8.49. There are four 1-year change periods covered in this study, so the cumulative gain target for the average elementary school over the course of the study is 34 points.

[17]Using the same method of calculation as in the previous footnote, for middle schools, the cumulative gain target is 33.7. Taking the 1999 API coefficient of −0.282 from Table 3, multiplying it by −100 (since the school in the example has a base 1999 API 100 points lower than the other) and then dividing this by 33.7 gives 84%. For high schools, the cumulative gain target is 35.7. The 1999 API coefficient (from Table 3) is −0.225, multiplying this by −100 and dividing by 35.7 gives 63%.

[18]These results are available from the authors upon request.

[19]The expected gain is actually 4.165 points. Dividing this by 34, the cumulative gain target for the average elementary school for the period of this study, gives 12%.

Table 3
Ordinary least-squares cross-section results by type of school, dependent variable: Change in API (2003 API-1999 API)

Variable	Regression 1: elementary schools		Regression 2: middle schools		Regression 3: high schools	
	Unstandardized coefficients	Standardized coefficients	Unstandardized coefficients	Standardized coefficients	Unstandardized coefficients	Standardized coefficients
Constant	335.414		236.514		159.724	
	(62.444)**		(22.267)**		(13.430)**	
1999 API	−0.394	−0.891	−0.282	−0.818	−0.225	−0.610
	(−57.805)**	(−57.805)**	(−17.035)**	(−17.035)**	(−10.417)**	(−10.417)**
District size	6.714×10^{-5}	0.175	-1.187×10^{-5}	−0.041	-4.278×10^{-5}	−0.130
	(16.517)**	(16.517)**	(−1.183)	(−1.183)	(−3.567)**	(−3.567)**
School size	−0.003	−0.008	0.001	0.007	0.001	0.008
	(−0.833)	(−0.833)	(0.214)	(0.214)	(0.245)	(0.245)
Meal	−0.112	−.0.061	−0.069	−0.046	0.177	0.106
	(−4.824)**	(−4.824)**	(−1.364)	(−1.364)	(2.747)**	(2.747)**
College grad	0.138	0.034	−0.056	−0.014	0.031	0.008
	(2.718)**	(2.718)**	(−0.364)	(−0.364)	(0.147)	(0.147)
Private school	−17.451	−0.025	−38.383	−.084	−51.437	−0.103
	(−2.363)*	(−2.363)*	(−2.904)**	(−2.904)**	(−3.010)**	(−3.010)**
Pop density	0.307	0.042	−0.036	−.0.007	0.624	0.119
	(4.064)**	(4.064)**	(−0.240)	(−0.240)	(3.409)**	(3.409)**
Median HH income	4.165×10^{-4}	0.131	3.351×10^{-4}	0.154	4.290×10^{-4}	0.216
	(10.476)**	(10.476)**	(4.417)**	(4.417)**	(5.379)**	(5.379)**
Emergency credential	0.029	0.006	0.255	0.062	0.157	0.029
	(0.508)	(0.508)	(2.060)*	(2.060)*	(0.884)	(0.884)
Decrease in class size	−1.022	−.035				
grades K-3	(−3.982)*	(−3.982)*				
Decrease in class size	0.096	.006	−0.682	−0.098		
grades 4–6	(0.665)	(0.665)	(−4.052)**	(−4.052)**		
Decrease in class size			−0.558	−0.043	−0.347	−0.023
grades 7–12			(−1.775)†	(−1.775)†	(−0.787)	(−0.787)
Adjusted−R^2 N	0.69	0.69	0.58	0.58	0.37	0.37
	4174	4174	768	768	745	745

Numbers in parentheses are t-statistics.
†significant at a 10% error level.
*significant at a 5% error level.
**significant at a 1% error level.

The percent of students on free or subsidized meal plans (an indication of poverty) is negatively related to test score gains in elementary schools, but positively related to test score gains among high school students. The high school finding is hard to explain in light of previous studies of the effect of poverty on student performance. Considering the size of the coefficient estimates in the two cases where the results were statistically significant, the practical effect of the meal plan variable is small.

The coefficient for the percent of students in a community who are enrolled in private schools consistently has the same sign across all three levels of education. The more students in private schools, the lower the test score gains. This is not what we expected as we thought that competition from private schools would motivate public schools to improve their scores. Perhaps the lower test score gains in areas with relatively greater enrollment in private schools reflects the fact that, in these neighborhoods, a sizeable portion of the very parents who might pressure local public schools to improve test scores enroll their children in private schools.

District size is positively related to test score gains for students in elementary schools but negatively related to test score gains for high school students. Again, this inconsistency in the results is hard to explain.[20]

[20]Driscoll et al. (2003) use instrumental variables estimation to control for endogeneity between district size and the level of API scores. Here, Hausman tests do not show evidence of an endogeneity problem with the district size and the change in API scores. These results are available from the authors upon request.

Table 4
Predicted versus actual values for change in API, based on results in Table 3, for the nine schools with 1999 API scores closest to the mean for the school's type and third

	Elementary schools			Middle schools			High schools		
	Lowest third of '99 API	Middle third of '99 API	Top third of '99 API	Lowest third of '99 API	Middle third of '99 API	Top third of '99 API	Lowest third of '99 API	Middle third of '99 API	Top third of '99 API
School name	Brooklyn Elementary	Morse (Barbara Comstock)	Township Elementary	Mulholland Middle	Greene Middle	East Avenue Middle	De Anzi Senior High	Modesto High	Mammoth High
District name	San Diego City Unified	Elk Grove Unified	Simi Valley Unified	Los Angeles Unified	Natomas Unified	Livermore Valley Joint Unified	West Contra Costa Unified	Modesto City High	Mammoth Unified
School's 1999 API	476	628	787	502	630	772	502	621	740
School's district size	104,515	33,068	15,420	551,229	4259	10,546	26,629	10,303	955
School's median HH income	26,505	36,001	67,311	37,178	44,674	75,026	61,494	30,567	46,858
Predicted value of change in API	151	96	51	81	61	29	66	36	11
Actual value of change in API	156	145	41	99	56	37	45	20	32
Residual	5	49	−10	18	−5	8	−21	−16	21

The positive coefficient on population density for the elementary and high schools is interesting. If urban areas suffer from relatively lower parent participation (more moms work), standardized test results may motivate program reform that otherwise would not occur.

Class-size reduction appears to have had a negative effect on test score gains in elementary and middle schools. Perhaps the disruption caused by the shift to smaller classes had an effect. Otherwise, we can only conclude that class size must proxy other characteristics of the school that we are not aware of or do not control for. Clearly, there is no basis for giving class size reduction any credit for the increase in test scores over the 1999–2003 period. When the result of a change in the independent variable is compared with the average API gain target (as done above for median household income and the base 1999 API score) the practical impact of the district size, population density, and class size reduction variables is small. These variables still serve a purpose as control variables.

As for the use of emergency credentialed teachers, there appears to be no negative effect. Is this evidence that teacher-training programs offer little value added in terms of improving class outcomes? Or, are there other characteristics of the emergency credentialed teachers that are producing the counterintuitive result? It is possible that many individuals, out of work in California during the recession of the early 2000s, went into teaching because the educational prerequisites necessary for a credential had been set aside. For these individuals, the required additional coursework might have been too costly (including the opportunity cost of their time) to justify a career shift into teaching. But, as the required coursework did not preclude employment, individuals were attracted to teaching who otherwise would have found work elsewhere. If this were the case, lowering credential requirements might increase the supply of talented individuals from which the system could select.

The results from Table 3 can be used to predict changes in API scores for specific schools in different situations. These predictions can then be compared to the actual results. Table 4 shows predicted changes in API scores for nine schools. For each type of school, three schools were chosen: one to represent those that had a 1999 API score in the lowest third of schools of that type, one for the middle third, and one for the upper third. In each case, the school with a 1999 API score closest to the mean for that third was selected[21] (these means are listed on Table 2). Table 4 also includes values for some of the independent variables for the schools selected. For example, the first school shown is Brooklyn Elementary, part of San Diego Unified. Brooklyn Elementary scored in the lower third of all California elementary schools on the 1999 API, with a score of 476. The regression model for elementary schools in Table 3 predicts that Brooklyn Elementary's API should increase 151 from 1999 to 2003. The actual gain was 156. In other cases, the model was not as accurate. Morse Elementary, with a 1999 API score of 628, represents elementary schools in the middle third of 1999 API scores. The model predicts Morse will have a gain of 96 points, but in reality, it gained 145 points. For the top third of elementary schools, the model predicts Township Elementary, part of Simi Valley United, would enjoy an increase of 51 points, but the actual increase was 41 points. The model predicts smaller gains for schools with higher base scores. Looking over Table 4, the cases shown are generally consistent with the conclusions made from the regression results, although the degree of accuracy in predicting the API gain can vary.

6. Conclusion

The evidence suggests that, where standardized tests are used, variations in achievement gains over time reflect characteristics of the schools and the student population. Our findings say that observers can expect test scores to increase the most where initial test scores are low, where private school enrollment is low, and where median household income is relatively high. If rewards are not adjusted for these conditions, the difficulty of meeting goals may discourage teachers and schools from efforts to improve academic performance. For this reason, states revising their reward system or setting one up might want to take base scores and characteristics of the school and its population in consideration in setting school-based rewards.

Acknowledgments

The authors thank William Kaempfer, the participants in the "Determinants of Educational Quality" session at the 2005 Public Choice Meetings, and an anonymous referee for helpful comments.

[21] If more than one school had a 1999 API score closest to its group's mean, one of these schools was chosen randomly.

References

Boghossian, N. (2004). School scores guide home-buying parents. *Daily News of Los Angeles*, *22*, 6.

Bohrnstedt, G. W., & Stretcher, B. M. (Eds.). (2002). *What we have learned about class size reduction*. Sacramento, CA: California Department of Education.

Cotton, K. (1996). School size, school climate, and student performance. *School Improvement Research Series,* (North-West Regional Educational Laboratory) May available at ⟨www.nwrel.org/scpd/sirs/10/c020.html⟩ on December 3, 2006.

Driscoll, D., Halcoussis, D., & Svorny, S. (2003). School district size and student performance. *Economics of Education Review*, *22*(2), 193–201.

Hanushek, E. A. (1986). The economies of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, *24*(3), 1141–1177.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, *19*(2), 141–164.

Hoxby, C.M. (1988). What do America's traditional forms of school choice teach us about school choice reforms? *Federal Reserve Bank of New York, Economic Policy Review*, *March*, 47–59.

Hoxby, C.M. (1994). *Do private schools provide competition for public schools*? National Bureau of Economic Research working paper no. 4978.

Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review*, *90*(5), 1209–1238.

Hoxby, C. M. (2003). School choice and school productivity or could school choice be a tide that lifts all boats? In C. M. Hoxby (Ed.), *The economics of school choice* (pp. 287–341). Chicago: University of Chicago Press.

Jepsen, C., & Rivkin, S. (2002). *What is the tradeoff between smaller classes and teacher quality*? National Bureau of Economic Research working paper no. 9205.

Kane, T.J., & Staiger, D.O. (2002a). Volatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp. 235–283) Washington, DC: Brookings Institution Press.

Kane, T. J., & Staiger, D. O. (2002b). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, *16*(4), 91–114.

Koretz, D. (1996). Using student assessments for educational accountability. In E. A. Hanushek, & D. W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp. 171–195). Washington, DC: National Research Council, National Academy Press.

Shkolnik, J., Hikawa, H., Suttorp, M., Lockwood, J.R., Stecher, B., & Bohrnstedt, G. (2002). The relationship between student characteristics and student achievement in reduced-size classes: A study of 6 California districts. In G. W. Bohrnstedt & B. M. Stretcher, (Eds.), *What we have learned about class size reduction* (Appendix D). Sacramento: California Department of Education.

Testa, W.A., & Sen, S. (1999). School choice and competition [Special issue]. *Federal Reserve Bank of Chicago, Chicago Fed Letter*, (143a).

US Census Bureau. (2000). *Census CD 2000, Long form SI3, Release 2.0*. East Brunswick, NJ: GeoLytics, Inc.