

Similar Students, Different Results: Why Do Some Schools Do Better?

A large-scale survey of California elementary schools serving low-income students



AMERICAN INSTITUTES FOR RESEARCH

Initial Report of Findings

October 26, 2005

***Similar Students, Different Results:
Why Do Some Schools Do Better?***

A large-scale survey of California elementary schools serving low-income students

Research Team:

EdSource

Trish Williams, executive director, study project director

Mary Perry, deputy director

Carol Studier, senior research associate

Noli Brazil, research assistant

Stanford University

Michael Kirst, Ph.D., principal investigator

Edward Haertel, Ph.D., senior technical consultant

Sean Reardon, Ed.D., technical advisor

University of California, Berkeley

Elisabeth Woody, Ph.D., principal research scientist

Melissa Henne, assistant research scientist

American Institutes for Research (AIR)

Jesse Levin, Ph.D., senior research scientist

Roger Levine, Ph.D., director of Cognitive Survey Laboratory

Suggested citation:

Williams, T., Kirst, M., Haertel, E., et al. (2005). *Similar Students, Different Results: Why Do Some Schools Do Better? A large-scale survey of California elementary schools serving low-income students*. Mountain View, CA: EdSource.

For more information, contact:

Trish Williams or Joan McRobbie at 650/917-9481

To request copies of this report, contact:

Maggie Rancatore,

520 San Antonio Road, Suite 200,

Mountain View, CA 94040-1217;

650/917-9481; edsource@edsource.org

This report is also available at: www.edsource.org

Copyright 2005 EdSource, Inc. All rights reserved. Call 650/917-9481 for permission to reprint.

Similar Students, Different Results: Why Do Some Schools Do Better?

EXECUTIVE SUMMARY

Why do some California elementary schools serving largely low-income students score as much as 250 points higher on the state's academic performance index (API) than other schools with very similar students? This study sought answers to that question by surveying principals and teachers in 257 California elementary schools serving similar student populations and analyzing the results to determine which current K-5 practices and policies are most strongly associated with higher levels of student achievement.

Our study differs from previous effective-schools studies primarily in its scale, standards-based content, and targeted yet comprehensive approach. Strong participation rates within schools provided extensive data from approximately 5,500 teachers and 257 principals across the state. We examined statewide implementation of California's standards-based reforms, yet focused on schools serving large numbers of low-income students. Using the API as our measure, we included high-, medium-, and low-scoring schools, which gave us a basis for comparing practices.

The policy context for the study is California's standards-based accountability system. Many experts consider this state's K-12 academic content standards, adopted in the late 1990s, to be among the nation's most challenging. School APIs are based on student test scores on the California Standards Tests, which measure how well students at each school are mastering grade-level academic standards. Given this context, we used each school's most current (2005) API score as the primary performance outcome.

The sample of schools was drawn from the 25th to 35th percentile band of the state's 2003–04 School Characteristics Index where student demographic challenge factors are substantial, but not the most severe.

After reviewing the effective schools literature, we developed and field tested the principal and teacher surveys, which were designed to explore school qualities, policies, and practices related to school success. Specific domains explored were: implementing a coherent, standards-based instructional program; involving and supporting parents; using assessment data to improve student achievement and instruction; encouraging teacher collaboration and professional development; ensuring instructional resources; enforcing high expectations for student behavior; and prioritizing student achievement.

Extensive analysis of the survey findings used regression analysis to determine which activities more common at high-performing than at low-performing schools were correlated with higher API scores. The practices found to be associated with high performance were:

- **Prioritizing Student Achievement.** Where teacher and principal answers to multiple survey questions indicated higher expectations for students, their schools had, on average, higher API scores than schools whose staffs indicated lower expectations. In more-successful schools, both teachers and principals reported that their school has well defined plans for instructional improvement and that they put priority on meeting the state's API goals and the federal adequate yearly progress goals. Teachers and principals also report that their schools set measurable goals for exceeding the mandated API student subgroup growth targets for improved achievement.
- **Implementing a Coherent, Standards-based Curriculum and Instructional Program.** Teachers who report the following were more likely to be in higher performing schools:

schoolwide instructional consistency within grades; curricular alignment from grade-to-grade; classroom instruction guided by state academic standards; curriculum materials in math and language arts aligned with the state's standards; in a district that addresses the instructional needs of English learners at their school. Principals were more likely to be in higher performing schools if they reported that: the district has clear expectations for student performance aligned with the district's adopted curriculum, and the district evaluates the principal based on the extent to which instruction in the school aligns with the curriculum.

- **Using Assessment Data to Improve Student Achievement and Instruction.** Strongly correlated with a higher API was the extensive use of student assessment data by the district and the principal in an effort to improve instruction and student learning. For example, principals more often reported that they and the district use assessment data from multiple sources (curriculum program and other commercial assessments; district-developed assessments; the California Standards Tests and the CAT/6) to evaluate teachers' practices and to identify teachers who need instructional improvement. Principals report using this data to develop strategies to follow up on the progress of selected students and help them reach goals. According to these principals, the district expects all of its schools to improve achievement, evaluates principals based on student achievement, and provides support for site-level planning related to improving achievement.
- **Ensuring Availability of Instructional Resources.** Where more teachers reported having regular or standard certification for teaching in California, schools had, on average, higher API scores. The same was true of schools where principals more often reported that their districts provide sufficient and up-to-date instructional materials as well as support for supplementary instruction for struggling students and for facilities management. Teachers with at least five years of full time teaching experience were more likely, on average, to be from schools with higher APIs. Principal experience was also correlated with higher school achievement.

Besides signaling critical, interrelated practices of more-effective schools, these findings indicate that the principal and the district play key roles in school success. Specifically, it appears that:

- **Principal leadership in the context of accountability-driven reform is being redefined to focus on effective management of the school improvement process.** In general, API scores were higher in schools with principals whose responses indicate that they act as managers of school improvement, driving the reform process, cultivating the school vision, and extensively using student assessment data for a wide variety of school improvement areas of focus, including evaluation of teacher practice and assistance to struggling students.
- **District leadership, accountability, and support appear to influence student achievement.** Principals' responses indicate that district practices may contribute to a higher API in a variety of ways. These include setting clear expectations that schools meet API and AYP growth targets, including for subgroups, as well as providing schools with achievement data and evaluating principal performance and teacher practices based on that data. They also include ensuring: that math and language arts curricula are aligned with state standards; that instruction is focused on achievement; that schools have adequate facilities and textbooks as well as resources for struggling students.

Across California, schools serving similar types of student populations can vary widely in how well they score on the API. The 257 elementary schools studied were drawn from a fairly narrow student demographic band. Yet their 2005 Growth API scores varied by about 250 points. This range of scores suggests that while student socioeconomic background is one predictor of academic achievement, it is not the sole predictor. What schools do—and what resources they have to do it with—can make a difference. With that in mind, the interrelated practices identified in this study may help schools in their efforts to improve student achievement.

Similar Students, Different Results: Why Do Some Schools Do Better?

A large-scale survey of California elementary schools serving low-income students

In California, the correlation or relationship between school-level student demographics (such as percentages of students from families with low incomes and education levels) and school-level academic achievement (as measured by the state's Academic Performance Index, or API) is quite high. However, it is also true that two schools serving similarly challenging student populations can have very different levels of performance – a difference of as much as 250 points on the API (on a scale of 200-1000). Why? This was the central research question we set out to answer.

By conducting a large scale survey of elementary schools across California serving similarly challenged low-income student populations, a collaborative research team (EdSource, Stanford University, U.C. Berkeley, and American Institutes for Research) sought to find which current K–5 practices and policies are most strongly correlated with high achievement. The study's surveys focused on concrete and actionable practices and policies at the school level, but also gathered teacher and principal reports about district and classroom practices.

Specifically, this study surveyed teachers and principals at California elementary schools in the 25th to 35th percentile band of the School Characteristics Index (serving high proportions of low-income students), then analyzed their survey results against the school's Academic Performance Index score for 2005.

The collaborative research team is choosing to release its initial findings in October of 2005 because as elementary schools receive their API growth scores, these findings may be particularly helpful in guiding their deliberations around school improvement. Early in 2006 the research team will issue a follow-up report with implications from the findings for practice and policy, as well as with the results of additional analyses of the survey data and of interviews with 21 district superintendents with schools in our sample.

How this study is different

Over the years, many research studies have examined which practices and policies make schools most effective. This study differs in a variety of ways—particularly in terms of its scale, standards-based context, and targeted yet comprehensive approach.

While many studies have examined a group of districts or schools within a region, few have examined *such a large number of schools located across a broad geography*. In total, 257 schools from 145 districts throughout California participated.

Strong participation rates within each school provided extensive data from *approximately 5,500 teachers and 257 principals*. All schools in our sample returned the principal survey and the bulk of them returned surveys from at least 80% of their K-5 classroom teachers.

Further, many studies have focused on high-performing schools only, using a variety of measures to identify the schools and then examining their practices. By surveying teachers and principals *from the full range of school API performance*—high, middle, and low—this study sheds light on what high-performing schools may be doing that low-performing schools are *not*.

In addition, the *survey questions help identify “intensity” of practice or policy implementation* with response scales ranging from weak to strong agreement, or from infrequent to very frequent as to how often a reported practice occurs.

Although the study has a large sample and has focused on both high- and low-performing schools, *it is also targeted—examining a specific organizational and policy context as well as a particular student population*. In other words—the study looks only at elementary schools serving a specific student population (those in the 25th to 35th percentile of the School Characteristics Index or SCI) that are operating in California’s current standards-based school policy context. The 2004 research review, *How Leadership Influences Student Learning* by Leithwood, Louis, Anderson, and Wahlstrom, recommends targeting studies in this way to gain broader knowledge about effective leadership behavior in certain environments.

In addition, few studies have examined *the implementation of standards-based reforms by schools across California*. While this study was designed mainly to inform schools and districts about useful practices and policies related to improving student achievement, its results also inform the state of its progress in implementing standard-based reforms.

Finally, to ensure that the study findings would be most useful to schools and districts, survey questions focused primarily on actionable items that can be implemented by other schools, rather than on general theories that are less clear in their implications for practice.

Education Policy Context and Background for California

In California, as in many other states, standards-based reform is currently the driving force behind public education policy. While elements of a standards-based system were in place as early as the mid-1980s in California, it was during the mid-1990s that an aligned standards-based education system began to develop more significantly. During this period, the state also focused most on elementary schools, especially their reading programs.

The general principle of K-12 standards-based reform is that all elements — curriculum, assessments, professional development, financial resources, and accountability systems — are aligned to widely agreed-upon, explicit academic content standards set at the state level. The standards specifically describe what students should know and be able to do at each grade level. Schools’ ability to help students learn the content standards depends in large measure on how well the state aligns all of the key elements of the system.

In 1995 the state initiated a process for the development of California’s academic standards and assessments. As of 1999, the State Board of Education (SBE) had adopted statewide academic content standards in the four core subject areas—English language arts, mathematics, history/social science, and science—with math and English language arts completed first.

According to the Fordham Foundation’s [The State of State Standards 2000](#), California adopted the most rigorous academic standards in the nation. Fordham gave California an overall grade of A-, making it at the time the only state to reach the “A” level for the rigor of its standards. In [Making Standards Matter 2001](#), the American Federation of Teachers gave California their highest ranking for the state’s academic standards in the four core subjects, reporting that the state’s standards were “clear and specific.”

The assessment system was established by state law in 1997 and implemented in 1998, but has evolved significantly since that time to align with the standards and curriculum. The accountability system aligned with the assessment-based system was initially established in 1998-99. The Public Schools Accountability Act (PSAA) was passed in the spring of 1999, and the first school APIs were issued in November 1999, based on tests from the prior spring.

Curriculum Frameworks and Textbooks

In California, the school board in each local district has historically been responsible for determining the subjects to teach and how to teach them, within the broad parameters set by the

state. As a result of the standards movement, state requirements and recommendations have had increasing influence on local choices.

Today the state specifies several subjects that all California public schools must teach. In addition, the State Board of Education approves a curriculum framework for each subject. In each of the four core subject areas (math, English language arts, history/social science, science) the framework is based on the state-adopted academic standards. The framework document itself provides an outline of what should be included in a given course of study and is meant to guide school districts and textbook developers.¹

Finally, the SBE recommends curriculum materials and instructional approaches. For grades K–8, the SBE adopts textbooks and other instructional materials for each subject area and each grade level. The state gives school districts funds to purchase materials, and a district must choose a percentage of its textbooks from the approved list in order to receive those funds. (However, districts can request a waiver if they find other materials more appropriate for their schools.) While the local school board ultimately decides on its own schools' textbooks and curriculum, the state's funding of particular textbooks influences those decisions.

For grades K–6, the SBE selected two curriculum programs for English language arts: *Houghton Mifflin Reading: A Legacy of Literature* and *SRA/Open Court Reading*. Grades 4-6 can also use several other texts. The selections for mathematics were more extensive, with districts having a choice of seven programs adopted for the grades K-5 or K-6.

State Testing System

California's assessment system is the Standardized Testing and Reporting or STAR program, which the state established in a 1997 law and began implementing in 1998. California public schools are required to test all pupils in grades 2-11 unless a parent requests in writing that the pupil be exempted. Each summer, the state releases results for testing completed the previous spring. The STAR program currently consists primarily of:

California Standards Tests (CSTs), which are based on the state's academic content standards. The CSTs are primarily multiple choice, but for fourth and seventh graders they also include a writing test.

California Achievement Tests, Sixth Edition Survey (CAT/6) which are taken at the same time as the CSTs and measure basic skills. Scores indicate the performance of each student relative to a national sample of students. Starting in 2005, only third and seventh graders took the CAT/6.

Currently, the CST serves as the key indicator of school performance in the state's accountability system and the basis for the Academic Performance Index (API) for schools, which was the dependent variable or student achievement outcome measure for this study. The State Board of Education developed five performance levels for reporting student results on the CSTs: far below basic, below basic, basic, proficient, and advanced, with the goal of all students scoring at proficient or above. These levels were first applied to the 2001 CSTs in English language arts and were integrated into the state's accountability system beginning in 2001.

Since 1999, a multiple-choice, norm-referenced test has also been a part of the state's accountability system, but to varying degrees over time. From 1999 through 2002, The Stanford Achievement Tests, Ninth Edition (SAT-9) was used, and for the first two years it was the sole accountability measure. Beginning in 2003, the SAT-9 was replaced by the California Achievement Tests, Sixth Edition (CAT/6) survey form. In 2005, the CAT/6 administration was scaled back from grades 2-11 to just grades 3 and 7. Significantly disabled students, who cannot take the CAT/6 or CSTs, take the California Alternate Performance Assessment (CAPA).

¹To see California's curriculum frameworks and other standards-based instructional materials adopted by the SBE, go to www.cde.ca.gov/be/st/fr

The state conducts two other assessments, the results of which are not included in the API calculation. The Spanish Assessment of Basic Education, Second Edition (SABE/2), is a Spanish-language test administered to Spanish-speaking students in grades 2–11 who have been enrolled in California schools for less than one year. It covers mathematics and reading and writing in Spanish. All students whose primary language is not English take the California English Language Development Test (CELDT) when they first enroll in school and each year after until school officials determine that they have become English proficient. In contrast to the SABE/2, the CELDT evaluates a student’s ability to listen, speak, read, and write in English.

California’s Accountability System

Many state policymakers saw an accountability plan as the final ingredient in a standards-based education system. Further, in 1994 Title I of the federal Elementary and Secondary Education Act (ESEA) called for such a plan. In 1997, the state began to examine how to align accountability with the standards, curriculum, and assessments and in 1999 it enacted the Public School Accountability Act (PSAA). Its cornerstone was the Academic Performance Index (API).²

An API score is a one-number summary of each spring’s test scores, with different tests receiving various weights in the index. With rare exceptions, every school in California is assigned an API score. Each school receives an API score between 200 and 1000, which is calculated from student scores on the CSTs and CAT/6 (and CAPA if appropriate). In addition to a score for the school as a whole, schools receive a score for each “numerically significant” subgroup of pupils categorized by ethnicity and poverty³. Under the current formula, to be “numerically significant,” a subgroup must have 100 students or have 50 or more students that constitute at least 15% of the school’s student body. Scores are included for all students who have attended school in the district since the prior October, including EL students.

A school’s API score is used in three different ways. First, schools of the same type (elementary, middle, and high schools) are ranked into deciles that each represent 10% of schools. Schools in Decile 1 have the lowest scores, and schools in decile 10 have the highest. Since schools’ decile rankings indicate how they compare to other schools, there will always be 10% of schools in the bottom decile—no matter how much the overall scores improve.

Second, schools are also compared to the 100 schools most like them in terms of student background and some other relevant factors. Based on their API score, schools receive a “similar school ranking” between one and 10. The similar school ranking indicates how well a school does compared to other schools that face a similar level of challenge. The School Characteristics Index (SCI), yet another composite index, is used to calculate the similar schools rank. The SCI is made up of many student demographic factors and a few school characteristics. All are weighted differently, with parent education level receiving the most weight. Multiple linear regressions are carried out each year to determine how the index will be calculated for every school in the state.⁴

Finally, the state gives schools a target for improving their API score and tracks the change from one year to the next. California’s goal is for every school to have an API score of at least 800. A school’s growth target is 5% of the difference between its current score and the state’s goal. Schools with API scores of 800 and above need only maintain their score at that level or above

² For a comprehensive overview of how the API is calculated and used, see “Overview of the Academic Performance Index School Base Reports 2004” at <http://www.cde.ca.gov/ta/ac/ap/documents/overview04b.pdf>.

³ To date, API subgroups have included: White, Hispanic, African-American, Asian, Filipino, Pacific Islander, American Indian/Alaskan, and economically disadvantaged. APIs will be calculated for both Special Education students and English learners beginning with the 2005-06 cycle.

⁴ For an in-depth description of how the SCI is constructed, see “Construction of California’s 1999 School Characteristics Index and Similar Schools Ranks” at <http://www.cde.ca.gov/ta/ac/ap/documents/tdgreport0400.pdf>.

800. Schools with the lowest API scores have higher growth targets, but they also have more room to grow.

The API system is organized into two-year cycles, with a “Base” score for the first year and a “Growth” score for the second year. (The Base and Growth scores can be thought of as “before” and “after” snapshots.) Early in the calendar year, each school receives a Base score based on its students’ performance on tests from the prior spring. The school is also given growth targets for the school as a whole and its subgroups. The Growth score—based on test scores from the following spring—is released in the fall.

Computing the API score for a school (and its subgroups) involves sorting students’ test scores into five performance levels. An API score is basically a summary of the distribution of scores among the five performance levels, with various subjects and tests receiving various weights in that calculation.

To achieve growth in its API score, a school (or subgroup) needs to have a greater percentage of its pupils score in higher performance bands. The API formula rewards growth from the bottom of the performance distribution upward more heavily than growth from the middle upward. This creates an incentive for a school to work with its lowest-performing students.

The NCLB accountability system, as implemented in California

Federal accountability requirements under No Child Left Behind (NCLB) became law in 2002, well after California had developed and implemented its accountability system. The state has satisfied the federal requirements for accountability under NCLB in part by using elements of its own state program. However, NCLB intensifies the focus on the achievement of every subgroup of students in a school. Further, the primary measure of success under NCLB is that a specific—and gradually increasing—percentage of all subgroups of students scores “proficient” or “advanced” on the California Standards Tests in English and math. Schools are also held accountable for testing 95% of students in each subgroup and the school as a whole. Failure to do so results in the same sanctions that occur if student performance is below expectations. A school’s API score must also meet a baseline expectation or improve by one point. Both individual schools and school districts as a whole that meet these expectations are said to have made AYP—Adequate Yearly Progress.

Schools and districts that fail to make AYP over two or more consecutive years (and that receive funding from the federal Title I program that supports low-income students) face an escalating set of consequences—from allowing students to transfer to other schools (with transportation provided) to shutting the school down—within a process called Program Improvement. While both federal and state accountability programs exact consequences for schools not meeting set targets, more California elementary schools are in jeopardy of missing the AYP targets, so this federal indicator is now driving many school improvement plans.

This study’s focus, however, was on how well elementary schools were performing on California’s API. We chose to not use a school’s AYP status per se as a dependent variable for this study for several reasons. Under AYP, the single “yes” or “no” status question regarding a school’s success is an accumulation of multiple data points related to both student test scores and participation rates. Failure of any student subgroup to meet the official benchmark on either dimension triggers the negative designation. Thus, schools that do not make AYP can be in that situation for a host of reasons. Conversely, the threshold for performance is rather low at this point, rendering a “yes” on AYP less meaningful as an indicator of a high performing school. The next level of analysis, beyond the simple yes or no, would have to look at outcomes for each subgroup. That includes far too many variables—and too much variation among schools—to be helpful for this study.

However, our survey asked questions about the extent to which school staffs prioritize meeting their AYP targets as well as their API targets because schools are expected to meet both.

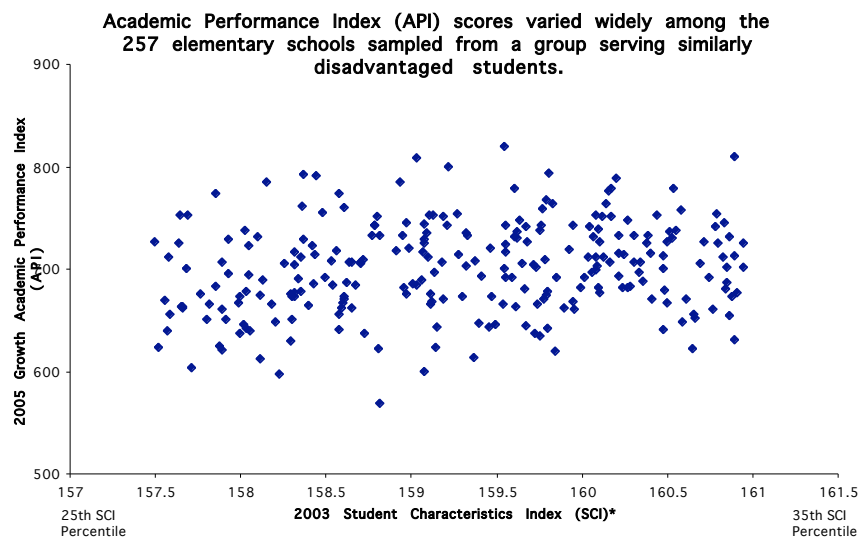
Schools are addressing the demands of both the state and federal accountability systems by focusing on improving student achievement on the California Standards Tests, as these tests form the key measure of performance (API and AYP) for both systems. As a result, API scores have a strong relationship to the AYP system.

Research Process and Methods

The research question—“why do schools serving similarly challenged student populations vary in their performance on the API by as many as 250 points?”—drove the research decisions and plans.

Early in the development of this study the research team agreed on several key decisions:

- The primary performance outcome, or dependent variable, would be the school’s most current (2005) API score.
- The sample of schools would be drawn from a band within the 2003–04 School Characteristics Index (25th to 35th percentile) where the student demographic challenge factors are substantial, but not the most severe. Limiting the study sample to a narrow SCI band helped control for student demographics.



*25th-35th SCI percentile based on most recent data available in winter 2004, when schools were sampled.

The research team began work early in 2004 by reviewing the research literature on effective schools, as well as on high-performing, high-poverty schools. Our initial survey questions for principals and for teachers were grounded in seminal research, covering such broad areas as School Context, District Role, Core Curriculum, Instructional Strategies, English Learners, Kindergarten, Assessment and Data, Professional Development, and Principal Leadership.

In the spring and summer of 2004 these surveys underwent review by academics who commented on their relevance to previous research; by state policymakers to ensure we had captured the state’s policies accurately; and by K-12 educators to get feedback on the surveys’ focus and wording. The research team’s goal was for the questions to be impartial in their wording and

focused on a wide range of potentially meaningful school practices and policies that are actionable and likely to be replicable in other schools.

Field Testing the Surveys

In the summer of 2004 we piloted the surveys in eleven different elementary schools across the state. We asked principals to take the survey before we arrived and then interviewed them about their impressions of the instrument, what questions did not make sense, and how on target the survey was in capturing what was important to them. We followed that with a general conversation with the principal as to what was working or not at the school. These conversations led us to some additional survey questions that were not already evident from themes in the existing literature. The majority of each school's K-5 classroom teachers also participated in taking the surveys. After they finished, we conducted cognitive interviews to ensure that their understanding of the questions matched what we intended to ask.

As the field-testing progressed, what we heard from school principals and teachers made us realize that the research prior to standards adoption in California did not capture all of what schools were now doing to raise student achievement in this context. We refined our surveys to make sure they reflected the general domains of effective schools practices found in previous research but also to capture the specific kinds of classroom, school, and district practices and policies we heard were more relevant to the current standards-based education policy environment in California. We then shortened the surveys in the hopes of increasing participation.

The final products were two 45-minute surveys: one for K-5 classroom teachers (46 multi-part questions totalling 371 items) and one for elementary principals (36 multi-part questions totaling 442 items). **[To view copies of the surveys go to www.edsource.org]**

The Sampling Process: Selecting and Recruiting the Schools

[See Appendix A on the study findings page at www.edsource.org for more details]

The 25th–35th percentile SCI band has approximately 550 elementary schools, most with high levels of low-income and English Learner students, many with high percentages of Hispanic and/or African-American populations. The research team employed a purposive, stratified random sample to choose the schools within this SCI band to approach for participation in the study. Our initial goal was to get a sample of 300 schools to participate, including schools from eighteen targeted districts across the state as well as all of the API Decile 7 or above (highest performing) schools within the band.

EdSource took on the task of recruiting schools for the study because of its 28 year history of strong relationships with districts.

The sampling plan organized schools into groups (strata) by first creating strata representing a small number of specific districts and then classifying the remaining schools into strata defined by API decile. These strata were crossed with a stratification according to EdSource subscriber status. Because it was anticipated (correctly) that it would be easier to enlist the cooperation of schools in EdSource subscriber districts, a higher sampling fraction was specified for these districts. Within each stratum, schools were randomly sampled. District offices for sampled schools were approached first. A refusal at the district level resulted in all schools in that district being dropped. New schools were then randomly sampled and contacted from the districts that agreed to cooperate.

EdSource initially gained the agreement to participate from 269 schools from 154 districts; the final sample included 257 schools from 145 districts.

Research Methodology and Analyses

[See Appendix B on the study findings page at www.edsource.org for more details]

Researchers from each of the four collaborating organizations have met biweekly since March 2005 to discuss the data analysis, while the technical team has met more frequently to specify details of the analyses. Data file construction was carried out primarily by AIR under the direction of Dr. Edward Haertel, the study's senior technical consultant.

A very high level of cooperation among the schools in the selected SCI band was obtained, and comparisons of participating versus nonparticipating schools suggest that any biases due to less than full participation may be safely ignored. To the extent there were differences, the research team adjusted for them using sampling weights so that results reported are representative of the 547 schools in the 25th to 35th percentiles of the SCI.

Teacher surveys were aggregated to the school level. School-level teacher survey data were then merged with principal survey responses as well as demographic and achievement data from EdSource and the California Department of Education to create the final school-level analysis file. Information from several databases was incorporated in the merged file.

The next step in the analysis, using the school-level file, was to combine the reliable teacher and principal survey items to create composite variables (scales) representing school qualities, policies, or practices potentially related to academic success. Within the project—and in this document—these were referred to as “sub-domains.” An in-depth analysis of the sub-domains was performed to ensure that the groups of items had internal consistency and that the included items were positively correlated with achievement.

The primary achievement outcome of interest was the Academic Performance Index (API). Initially, the 2004 "Growth" API was the primary focus of analyses. When the 2005 "Growth" API was released in August, analyses were rerun using the 2005 test data and 2005 demographic control variables. Results were essentially unchanged; the 2005 results are the primary reporting focus.

As mentioned previously, "Growth" API is a cross-sectional measure based on data from a single year. The term "Growth" indicates that it is calculated in the same manner as the preceding year's "Base" API. Thus, the arithmetic difference between current-year "Growth" and prior-year "Base" APIs represents the change in a school's performance from one year to the next. Only cross-sectional measures, not change measures, are used in the analyses for the present study. In addition, an outcome variable based on schools' average API over three years—between 2002 and 2004—with greater weight on more recent years, was examined. This "API Composite" gave results similar to, but generally weaker than, those for recent single-year API outcome variables.

Results were also examined for school-level variables created by averaging CST scores in English language arts across all grades within the school, and similarly for mathematics. Analyses of these separate outcome measures showed little difference in variables predictive of performance in either subject area.

The sub-domains were further organized into eight categories ("domains") representing general areas of focus like the school's instructional program, parental involvement, use of achievement data, and student behavior expectations. For the primary regression analyses, we first regressed the outcome (e.g., 2005 Growth API) on all sub-domains within a given domain, together with a set of demographic variables chosen to control for residual effects of socioeconomic status and student characteristics that were still present within the narrow SCI band specified in framing the study population. The primary statistic of interest in examining these regressions was the difference in percent of variance accounted for (adjusted R^2) for that entire domain and the percent of variance accounted for in a regression including demographic variables only.

For each domain, we next fit a parsimonious regression model that included school demographic characteristics and a subset of the sub-domains that collectively accounted for most of the variation in API scores captured using the entire domain. We then used the estimates from this regression model to predict the API for each school, but assuming each school had demographic characteristics equal to the average of the schools in the sample. The extent to which these predicted API scores vary across schools gives us a measure of the extent to which the composite variables from each domain account for variation in API scores. A final analysis employed the same procedures, but included the final subsets of sub-domains from all eight domains in a single model.

We want to point out that the teacher survey contained a considerable number of questions related to school and classroom instructional practices for English learners. This population is particularly large in California and research that can shed light on the most effective ways to raise their academic achievement is important. However, the research team's technical group found that determining the appropriate dependent variable for studying the responses to this subset of survey questions was too complex and methodologically challenging to accomplish within the timeframe of this study. We will report descriptive findings related to these survey questions in early 2006 (see Further Analyses at the end of this report).

Descriptive Statistics of the Schools in Our Sample and Their Student Populations

[See Appendix C on the study findings page at www.edsource.org for more details]

Overall Demographics and Performance of Schools Participating in the Study⁵

In 2003-04, when the study's sample was derived, 547 non-charter elementary schools were in the 25th-35th percentile of the school characteristics index (SCI) band. Out of the 269 schools identified from this group for the study's sample, 257 agreed to participate in the study and returned the principal's survey. The overwhelming majority of these schools also returned at least 80% of the teacher surveys.

Schools in 145 school districts participated in the study. Fifty-six of these districts had more than one school in the study. According to 2004-05 data, the average school enrollment was 588 students. Of the 257 participating or sample schools:

- 33% were situated in elementary school districts and the remaining 67% in unified school districts.
- 15% of the schools operated on a year-round calendar.
- 98% received Title I funding and 34% participated in Program Improvement—meaning that these Title I schools had not met the federal Adequate Yearly Progress benchmark for at least two consecutive years.

⁵ Data used in this section is from 2004-05 unless otherwise noted.

Student Demographics

The student population attending the participating schools was diverse and faced particular challenges. At the median sample school⁶, 40% did not speak English as their primary language and 78% participated in the free and reduced-price meals program for low-income families. Among all sample schools, the percent of English learners in a school ranged from 1% to 80% and the percent of students participating in the free and reduced price-meals program from 17% to 100%.

The composition of students by ethnicity at the median school was as follows: 68% were Hispanic, 14% white, 4% African American, and 3% Asian. The median value for the combination of American Indian, Filipino, Pacific Islander, and multi-ethnic students, along with those who did not indicate an ethnicity, was 4%. In most but not all schools the majority population was Hispanic. The highest percentage of students of a given ethnicity found in any one school was: 99% Hispanic, 94% African American, 54% white, 42% Asian, 38% American Indian, 31% Filipino, and 10% Pacific Islander. Conversely, the lowest percentage for most ethnic groups was 0%, but all schools had at least 1% Hispanic students.

In the median sample school, 32% of students had parents who were not high school graduates. The median values for other parental education levels were 33% high school graduates, 21% some college, and 11% completed college or graduate school.

School Performance

The 2005 Growth API for the average school participating in the study was 702. The lowest-performing school scored 569 and the highest performing school 821.

The state also assigns a rank from one to 10 to each school in the state at the start of each API cycle, with a 10 identifying the 10% of schools that are the highest performing. For the 2004 Base API, none of the schools in the study were ranked a ten. Only 4% ranked at seven or higher, meaning within the top 40% of elementary schools. Conversely, 7% were ranked at one, the bottom decile. The remaining 89% of schools ranked from two to six, with the majority a three or four.

Description of Study Surveys and Content Domains for Teachers and Principals

As already described, the survey items were organized into sub-domains that represented general categories of school qualities, policies, or practices related to academic success. Those sub-domains were then combined into eight more general domains for analysis purposes, as described below. Our development of these domains rested on an extensive review of existing research related to effective schools, district effectiveness, and standards-based reform.

Prioritizing Student Achievement

This domain examined the importance both the school and district placed on setting clear, high, and measurable expectations for student achievement. Both teachers and principals were asked about the extent to which their school and district communicated high expectations and took responsibility for student achievement. Further, they reported the degree of priority given by teachers, the principal, and the district to meeting API and AYP targets for subgroups of students (such as by race/ethnicity and income level).

⁶ The median is found by arranging schools in order of their values on a particular variable and finding the value in the middle of the group. As a result, there is no one school that represents the median on all of these factors.

Implementing a Coherent, Standards-Based Instructional Program

This domain contained those sub-domains that most clearly indicated that a school's curriculum and instruction are coherent and aligned with state standards. Those included:

- The amount of time spent on mathematics and language arts instruction, and the extent to which they are protected from interruption and integrated with other subjects.
- The proportion of teachers in a school who regularly use the same curriculum packages, and which ones they used.
- The extent to which teachers reported alignment and consistency in curriculum and instruction, planning and materials.
- Teacher, principal, and district use of state standards to guide curriculum and instruction.
- The use of a standards-based report card.
- The extent to which the district had addressed the needs of English learners (EL).

Principals were also asked about the extent to which clear district expectations guided curriculum coherence; whether they saw themselves as knowledgeable about standards and curriculum; and whether the school had implemented a new program for EL students in the last four years.

Using Assessment Data to Improve Student Achievement and Instruction

This was perhaps the most intensively examined domain in our study in terms of numbers of questions asked of both teachers and principals. Under the general topic of data and assessment, questions addressed the types of assessment data teachers and principals received, as well as how they used this data. We categorized the types of data as follows:

- CST (California Standards Tests) and CAT/6 assessment data, the state's standardized tests administered each spring.
- CELDT (California English Language Development Test), an annual assessment of English learners.
- Curriculum program assessments.
- District-developed assessments.
- Other commercial assessments.
- Assessments created by individuals in a school.

Based primarily on item content—but also on the results of our factor analyses—the sub-domains were organized differently for teachers than for principals. Teachers' responses were organized around the frequency with which they reviewed assessment data generally, and the extent to which they used the specific data types to monitor student performance and inform their instruction.

The analysis of principals' responses reflected different questions, including their use of specific types of assessments and the extent to which they used each type to monitor achievement, address student progress, inform school-wide instructional strategies, and monitor and evaluate the practices and performance of teachers. Principals were also asked about the influence of district expectations for improving student achievement, and about incentives and activities specifically targeted at raising CST and CAT/6 scores.

Both principals and teachers also responded to a set of questions regarding the extent to which they addressed student achievement by subgroup.

Ensuring Availability of Instructional Resources

Given the absence of school-level financial data in California, the study was limited in its examination of resources. Data regarding the credentials and experience of educators was combined with teachers' reporting regarding adequate classroom materials, and principals' perceptions of a number of different types of resources, including:

- The skills, knowledge, and attitudes of the teaching staff at the school.
- The school's access to qualified support personnel, supplemental financial resources, and supplemental instructional time for students.
- The extent to which the district provides support for facilities and instructional materials.
- The amount of regular instructional time, including full-day kindergarten and extended school day or year.

Enforcing High Expectations for Student Behavior

The examination of this domain was limited to questions regarding the school's establishment and enforcement of policies related to student behavior. Both principals and teachers reported on the extent to which the school created an orderly and positive environment for student learning, including such areas as attendance policies, enforcement of rules, and respect for cultural differences.

Encouraging Teacher Collaboration and Providing Professional Development

This domain examined three different areas related to the professional environment in the school, and looked at a wide range of activities by teachers, principals, and districts. The first area was teacher collaboration and professionalism. The sub-domains looked at:

- The extent to which teachers felt they had influence over schoolwide decisions.
- The extent to which teachers and principals reported opportunities for teacher collaboration around curriculum and instruction, including for EL students.
- The extent to which teachers and principals perceived that teacher professionalism was supported and encouraged within the school and by the district, and the extent to which they experienced a continuous learning environment.

A second strand in this domain related to the development of educator capacity through professional development, with sub-domains focused on:

- The adequacy, influence, and value of a large number of different teacher professional development opportunities, including training linked to standards generally, specific curriculum programs, instructional strategies, the use of data to inform instruction, and non-instructional issues.
- The frequency of teacher participation in coaching and modeling activities.
- The extent to which principals gave their district credit for providing teacher professional development opportunities.
- Principals also reported on the extent to which their district provided them with professional development, and the value they ascribed to the experiences they had.

Finally, several sub-domains explored the principal's perspective on the hiring, evaluation and firing of the teaching staff at his or her school, including:

- The principal's perceptions about the district's success in building and maintaining a strong teaching staff.
- The principal's capacity to evaluate teacher performance.
- The principal's perceptions regarding his or her ability to hire and remove teachers, including district and school factors that influence that ability.

Involving and Supporting Parents

This domain included sub-domains that looked both at the school's active engagement with parents and its support of parents and families. Teachers were asked about their practices involving parents in students' education; the district's success in building the community's confidence in the school; and the principal's relationship with parents. Principal questions were more comprehensive, including:

- The extent to which the school involves parents in students' education through mechanisms such as parent-teacher conferences, school-wide events, and translators for non-English speaking parents.
- The extent to which the school worked to engage parents in schoolwide decisions and activities.
- The extent to which the school provides support services to parents and families, including such things as ESL classes, health services, and assistance programs.

Initial Study Findings

Within our sample of elementary schools, the "effective schools" domains that proved to be the most significant in distinguishing the responses of teachers and principals from the highest performing schools from those in the lowest performing were:

- Prioritizing Student Achievement;
- Implementing a Coherent, Standards-based Instructional Program;
- Using Assessment Data to Improve Student Achievement and Instruction; and
- Ensuring Availability of Instructional Resources.

Much of what we found in this study was consistent with recent research that has sought to understand how standards-based reforms link to improved instruction and thus to positive student outcomes. *Perhaps the central message is that no single action, or even category of actions, can alone provide a clear advantage related to student performance.*

Education researchers such as Smith and O'Day (1991) have examined the impact of reforms on student achievement, arguing that instructional improvement is unlikely to result from a single policy or practice. As researchers from the Consortium for Policy Research in Education (CPRE)⁷ note, "schooling consists of complex processes, and it would be remarkable if there was only one best way to improve it."

This study's findings are consistent with that theory. Our findings appear to indicate that the cumulative effect of the state's policies on school practices differs among schools that serve low-income children. We have also identified at least some of the attitudes and activities that set the

⁷ Cohen, D.K. and Hill, H.C. *Learning Policy*. New Haven, CT: Yale University. 2001

higher performing schools apart, such as how well or frequently activities are implemented, including high expectations for student performance, curriculum coherence, and the use of assessment data for school improvement management.

Additionally, we found that schools that tended to demonstrate a strong culture of high expectations—as evidenced by their attention to meeting and exceeding state and federal accountability targets and setting high standards for student achievement—on average had higher API scores. Abelmann and Elmore (1999), among others, describe a similar alignment between schools’ internal accountability mechanisms (such as prioritizing student achievement) and external accountability requirements (such as meeting API/AYP targets).

Interpreting the Findings

This study’s findings represent the results of a regression analysis—a tool that uses statistical techniques to identify correlations among variables in large databases. The regression analysis allowed the researchers to isolate separate distinct correlations between schools’ API scores and several related independent variables thought to explain variation in API scores. This study does not prove that the four domains that are *correlated* with higher API scores have actually *caused* the higher API scores. Rather, it indicates that schools that report more strongly that they have implemented more of the practices included in each of the four domains have, on average, higher API scores than schools that report fewer of the practices.

To quantify the magnitude of API effects that might be attributable to the various domains we studied, we estimated the expected differences in API scores associated with changes in teacher, principal, and district engagement across the kinds of activities described. Sets of practices and policies were grouped together for this analysis and it is the predicted effect of the combined practices—not any single practice—that is reflected. The table below presents these domain-specific API effects. Because the practices associated with different domains tend to occur together (i.e., schools high on one domain tend to be high on others), these effects are not additive. Therefore, the API score difference associated with two or more domains of practices will be substantially less than the sum of the separate effects for the same domains.

Magnitude of API Differences

(holds demographics constant at sample mean)

Domain	Standard Deviation of Predicted API Distribution
Implementing a Coherent, Standards-Based Instructional Program	17.6
Ensuring Availability of Instructional Resources	16.9
Using Assessment Data to Improve Student Achievement and Instruction	16.7
Prioritizing Student Achievement	16.3
Enforcing High Expectations for Student Behavior	12.3
Encouraging Teacher Collaboration and Professional Development	11.0
Involving and Supporting Parents	9.9

As these data indicate, our findings did not show that practices to strengthen teacher collaboration and professional development, enforce high expectations for student behavior, or increase involvement and support of parents were unimportant in terms of contributing to a school’s API. They were not, however, nearly as strong as the other four domains in terms of differentiating the lowest performing schools from the highest in our sample of California elementary schools with high percentages of low income students.

The Findings

Prioritizing Student Achievement

Elementary schools where teachers and principals answered multiple survey questions in ways that indicated their higher expectations for students had, on average, higher API scores than similar schools with lower reported expectations. For example, when teachers reported that their school has a vision focused on student learning outcomes, their schools were more likely to be high scoring. The same was true of principal responses on similar questions. And when teachers and principals report that the principal communicates a clear vision for the school, sets high standards for student learning, and makes expectations clear to teachers for meeting academic achievement goals, the school is more likely to be high achieving. Equally important, better school performance seems to be associated with both teachers' and principals' reports that teachers at the school take responsibility for and are committed to improving student achievement.

A shared culture within the school regarding the value of improving student achievement and a sense of shared responsibility for it seems to separate higher performing schools in our sample from those with lower APIs. But beyond “values,” both teachers and principals reported that their school has well defined plans for instructional improvement, and that they make meeting the state’s API goals and the NCLB adequate yearly progress goals a priority. Both teachers and principals report that their schools set measurable goals for exceeding the mandated API student subgroup growth targets for improved achievement. Principals at higher performing schools also report on average that their school’s statewide rank and similar schools ranking on the API influence school wide instructional priorities, and that they are clear about their district’s expectations for meeting API and AYP growth and subgroup targets.

Implementing a Coherent, Standards-based Curriculum and Instructional Program

Our findings are consistent with previous research on the value of a coherent curricular program. Teachers who report schoolwide instructional consistency within grades — and curricular alignment from grade-to-grade — work in schools that performed better on average. Examples of practices teachers report using to accomplish this coherency include examining the scope or sequence of curriculum topics and reviewing a grade-level pacing calendar.

Those teachers who reported that their school has identified essential standards and that their classroom instruction is guided by state academic standards were also more likely to be in high performing schools. They report that the school’s curriculum materials in math and language arts are aligned with the state’s standards and that they frequently map state curriculum standards onto their classroom lesson plans. Teachers at higher performing schools also more often report that their district addresses the instructional needs of English language learners at their school.

Principals who report a strong district role in this domain are also from higher performing elementary schools. These principals say the district has a coherent grade-by-grade curriculum that it uses for all schools and that the district expects the principal to ensure implementation of the curriculum. These principals report that the district has clear expectations for student performance aligned with the district’s adopted curriculum and that it evaluates the principal based on the extent to which instruction in the school aligns with the curriculum.

Using Assessment Data to Improve Student Achievement and Instruction

Another practice strongly correlated with a higher API among our sample of elementary schools was the extensive use of student assessment data by the district and the principal in an effort to improve instruction and student learning. As an example, principals from better performing

schools more often reported that they and the district use assessment data from multiple sources — curriculum program and other commercial assessments, district-developed assessments, and the California Standards Tests and the CAT/6 — to evaluate teachers' practices and to identify teachers who need instructional improvement. Principals who reported frequently and personally using assessment data to address the academic needs of students in their schools led, on average, higher performing schools. They report using this data to develop strategies to help selected students reach goals and to follow up on the progress of selected students. In addition, they review this data both independently and with individual teachers.

These same principals report a clear understanding of their district's expectations for improving student achievement, which may help motivate and support them. The principals report that their districts expect that all schools in the district will improve student achievement and evaluate principals based upon student achievement. The principals report that the district also provides support for site-level planning related to improving achievement.

In schools where assessment data from the California Standards Tests and the CAT/6 influence schoolwide attention to improving student achievement, the API also tends to be higher. Teachers report receiving CST/CAT-6 test data in a variety of formats: for all students in their grade level; disaggregated by specific skills for all students in their classrooms; and disaggregated by student subgroup for students in their classrooms. Principals report using the CST and CAT/6 data to examine school-wide instructional issues, to develop strategies for moving students from below basic and basic to proficient, to compare grades within the school, to identify struggling students and evaluate their progress, and to inform and communicate with parents.

Ensuring Availability of Instructional Resources

API performance was higher in schools where principals reported that a higher proportion of their teaching staff had the following qualities (listed roughly in descending order of importance):

- demonstrated ability to raise student achievement
- strong content knowledge
- good fit with the school culture
- training in curriculum programs
- ability to map curriculum standards to instruction
- supportive of colleague's learning and improvement
- able to use data from student assessments
- familiar with the school community
- excited about teaching
- familiar with state standards

The schools where more teachers reported having regular or standard certificates for California also had, on average, higher API scores.

The principals who more often reported certain district practices were also more likely to be from high performing schools. They said their districts ensure the school has: up-to-date instructional materials, support to provide supplementary instruction for struggling students, enough instructional materials for all students, and support for facilities management. The same was true for teachers who tended to confirm these perceptions, reporting that every student in their classroom has a copy of the current version of the textbook in language arts and math and that the principal ensures the teachers have adequate classroom materials. The principals also reported a strong understanding of what their district expected from them in terms of facilities upkeep and management.

Years of educator experience also mattered, but less so than these other items. Teachers who were more likely to have at least five years of full time teaching experience were, on average, from

schools with higher APIs. Principal years of experience was also correlated with higher school achievement.

Another View: The Role of the Principal and the District

Principal leadership in the context of accountability-driven reform is being redefined to focus on effective management of the school improvement process.

In general, API scores were higher in schools with principals whose responses indicate that they act as managers of school improvement, driving the reform process and cultivating the school vision. In particular, they were more likely to embrace the state's academic standards and to ensure classroom instruction was based on them. They prioritized meeting and even exceeding state and federal accountability targets for school performance. In addition, they reported personally and extensively using student data for instructional management purposes — not only to evaluate the progress of students, but also to examine school-wide and teacher practices, develop strategies to help selected students reach goals, and identify teachers who need instructional improvement. Finally, they ensured that teachers and students had adequate texts and classroom materials.

District leadership, accountability, and support appear to influence student achievement.

Another theme interwoven in the study's data was how the district may influence school performance. While the study was not designed to examine the influence of district policies per se, principals at high-performing schools tended to perceive many aspects of the district role differently from principals at low-performing schools. Based upon principals' survey responses, it appears that districts may contribute to a higher API at these elementary schools in a variety of ways. Specifically, principals at high-performing schools said their districts were clear in their expectations that schools meet both the API and AYP growth targets for the school as a whole and for subgroups. They ensured that the school curricula in math and language arts were aligned with state standards and that instruction was focused on student achievement. In addition, these districts ensured that schools had adequate facilities, textbooks, and resources for struggling students. They also provided schools with student achievement data and evaluated principal performance and teacher practices based on that data.

Further Analyses

The collaborative research team is choosing to release its initial findings in October of 2005 because as schools in the 25th to 35th SCI band receive their API growth scores, these findings may be particularly helpful in guiding their deliberations around school improvement. Additional analyses will be reported in early 2006.

Our survey findings point to a strong district role in providing leadership, direction, and support to improving schools. As part of this study the researchers interviewed 21 superintendents or associate/sub superintendents in 17 different districts that had high-performing schools in our sample. We asked them to talk about the strategies they had used to foster improvements in school performance and student achievement. We plan to report on their observations in early 2006.

In addition, the principal survey in this study included an open-ended question, asking principals to identify the three most effective things their schools had done to improve student achievement. We hope to analyze these responses by the school's 2005 API, to see how those strategies varied and report our findings in early 2006.

High API scores were also correlated with teachers reporting frequent use of *SRA/Open Court Reading* for reading instruction and *Scott Foresman CA Mathematics* for math instruction. These teachers were also more likely to engage in the various effective practices described above. However, this study was not designed in a way that could determine the effects of specific curriculum programs independent of other schooling practices. We plan to run additional analyses to test the robustness of these results and report on them in early 2006.

The teacher survey in our study included a fairly extensive set of questions around instructional practices employed for English learner students. We plan to report the descriptive statistics from this part of the survey in early 2006; if additional funding is secured, we will also analyze these survey findings against appropriate dependent outcome measures to explore which practices seem to be working best.

Other analyses are possible with this rich and unique data set.

In early 2006 we will issue a follow-up report(s) to this study which will include the results of analyses identified above, a connection of the report's major findings to previous education research, and more discussion of the state policy and local school implications of all the findings.

For a copy of this report, more information related to the study, the appendices mentioned in various sections, and a full bibliography including the works cited, go to www.edsource.org and click on the homepage link to the Study.

Conclusion

Across California, schools serving similar types of student populations can vary widely in how well they score on their API. The 257 elementary schools (serving 135,673 K-5 students) that this research team studied were drawn from a fairly narrow band in terms of student demographics (percent low income, English learner, ethnic/racial subgroups). Yet their 2005 Growth API scores varied by as much as 252 points.

A school's API score reflects how well its students are performing on the annual California Standards Tests. This one test is limited: it is not the only way for a school to measure how well their students are mastering the rigorous academic content of the state's grade-by-grade standards; and it also does not measure the other important things that elementary students may be learning at school—about art and science and music, about citizenship and tolerance of differences, about themselves and their sense of place in the world.

But an elementary school's API score provides the state and the public with a consistent and easy way to grasp information about the progress its students are making toward mastering the important math and reading and writing skills that will enable them to succeed academically in later grades.

For this reason, among others related to accountability, a school's API score represents an important measure of student learning.

The range of API scores in our sample suggests that while the socioeconomic backgrounds of students is one predictor of academic achievement, it is not the sole predictor. What schools do—and what resources they have to do it with—can make a difference. With that in mind, the interrelated practices identified in this study may help schools in their efforts to improve student achievement.

Works Cited

Abelmann, C., & Elmore, R. F. (1999). *When accountability knocks, will anyone answer?* Philadelphia: Consortium for Policy Research in Education.

American Federation of Teachers. (2001). *Making standards matter 2001*. Washington, DC: American Federation of Teachers.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy*. New Haven, CT: Yale University Press.

Fordham Foundation. (2000) *The state of state standards 2000*. Washington, DC: The Fordham Foundation.

Leithwood, K., Louis, K.S., et al. (2004) *Learning leadership project: How leadership influences student learning*. New York: The Wallace Foundation.

Smith, M. S., & O'Day, J.A. (1991). Systematic school reform. In S. Fuhrman and B. Malen (Eds.), *The politics of curriculum and testing*, (pp.233-267). New York: The Falmer Press

Acknowledgements

The research team wishes to thank:

- the numerous researchers, K–12 educators, and policymakers who gave candid and invaluable feedback to improve this study throughout the two-year process;
- and the 257 elementary schools – their teachers, principals, and districts – for agreeing to participate in this study and for the hard work they are all doing to improve California’s schools and the achievement and experiences of the students who attend them.

EdSource wishes to further thank:

- the research team for their intelligence, knowledge, skill, dedication, perseverance, and frequent good humor as we worked collaboratively across four institutions over the past two years to complete this project;
- the entire EdSource staff and Board of Directors for their support of and many contributions to this work;
- the cadre of Bay Area school board members who served as research liaisons for this project, and whose effectiveness in communicating with districts and schools contributed to the study’s extremely high school participation and survey return rates;
- and Reed Hastings, CEO of Netflix, for posing an important research question then providing the financial support that allowed us to conduct this independent study to find some answers.