

education sector reports

Measures That Matter: Why California Should Scrap the Academic Performance Index

By Richard Lee Colvin



EDUCATION**SECTOR**

www.educationsector.org

ACKNOWLEDGEMENTS

This report was funded by the Stuart Foundation. Education Sector thanks the foundation for its support. The views expressed in the paper are those of the author alone.

ABOUT THE AUTHOR

RICHARD LEE COLVIN is a visiting fellow at the Woodrow Wilson National Fellowship Foundation and former executive director of Education Sector. He can be reached at rleecolvin@gmail.com.

ABOUT EDUCATION SECTOR

Education Sector is an independent think tank that challenges conventional thinking in education policy. We are a nonprofit, nonpartisan organization committed to achieving measurable impact in education, both by improving existing reform initiatives and by developing new, innovative solutions to our nation's most pressing education problems.

© Copyright 2012 Education Sector

Education Sector encourages the free use, reproduction, and distribution of our ideas, perspectives, and analyses. Our Creative Commons licensing allows for the noncommercial use of all Education Sector authored or commissioned materials. We require attribution for all use. For more information and instructions on the commercial use of our materials, please visit our website, www.educationsector.org.

1201 Connecticut Ave., N.W., Suite 850, Washington, D.C. 20036
202.552.2840 • www.educationsector.org

A few years ago, the symptoms of academic failure at Audubon Middle School southwest of downtown Los Angeles were obvious. Students roamed the trash-strewn campus during class hours, unafraid of consequences. The principal was rarely around, and when he was, he almost never visited classrooms. Observations required for teacher evaluations often were not done, yet teachers still received good ratings. The faculty divided into camps. Some closed their classroom doors and did the best job they could. Others did little more than show videos, knowing it didn't matter. The nearest Subway restaurant did a brisk business delivering sandwiches to classrooms during instructional time. "This was not a functioning school," one teacher said. "It was sink or swim, and we were just barely keeping our heads above water."

A dozen years of warnings and interventions and millions of extra dollars had done nothing to improve the school or its results. At the end of the 2008–09 school year, only 12 percent of seventh-graders and 6 percent of eighth-graders could demonstrate proficiency in math. Only one in five eighth-graders was proficient in English language arts. In the spring of 2010, leaders of the Los Angeles Unified School District had run out of ideas; they were ready to close the school and reopen it under new management. Perhaps a charter operator, some idealistic teachers, or even the teachers union could do better. The school was put on notice that its time was running out.

But though the numbers didn't yet show it, the district's leaders had done something at the start of the 2009–10 school year that within two years would have other schools coming to Audubon to find out its secret. The district had replaced the school's checked-out principal with a serious, methodical veteran named DeWayne Davis. He brought with him a team of experienced administrators who let students know they were expected to be in class, keep the campus clean, and work hard. They began visiting teachers'

classrooms every day. Although weak performers were offered help, it became clear that some were unwilling, or unable, to change. The administrators documented all that they saw to build a foundation of evidence for removing the weakest performers.

Given the school's awful math scores, it was not surprising that as the end of the year neared, Davis and his team informed 18 teachers, including all but one math teacher, that they would receive unsatisfactory ratings. Removing them could have taken years because of the due process protections of the district's labor contract and its policy forbidding the transfer of teachers with negative evaluations. But the district and the teachers union agreed to expedite the teachers' departures.

For all that changed that first year, however, the available data did not show by June that the hard work was paying off. So that fall, then-Superintendent Ray Cortines had to make the final call on the school's fate. Had he relied on proficiency data, the decision would have been straightforward: only 8 percent of eighth-graders were proficient or advanced in math,

and only a third were proficient in English language arts, a poor showing by any standard. Audubon's fate also would have been sealed had Cortines gone by the school's status under the federal No Child Left Behind law. The school that year failed to achieve Adequate Yearly Progress for the 10th straight time—a fact that also would have justified closing it down. The federal law allows schools to satisfy the AYP requirements as long as they reduce the number of underperforming students by 10 percent per year. But Audubon and 1,500 other schools in California hadn't met that goal either.

Cortines had one more measure to consult before he took the drastic step of putting an operating school out of commission: the California Academic Performance Index, a number that is derived from test scores and is the state's main accountability metric. On the one hand, the API showed that Audubon ranked in the bottom 20 percent of the state's middle schools. On the other, the school's index had risen by more than 12 percent in the single year with Davis on duty. That tipped the balance. Cortines removed the school from the takeover list, giving Davis more time to help the school improve.

But although the decision was no doubt ultimately the right one, the year-to-year changes in the API, for all the significance attached to them, actually convey no information about whether *individual students* are doing any better or worse. The reason is that each year's API is based on a distinct group of students. The scores of this year's seventh-graders, for example, are compared to the scores of last year's seventh-graders and, while both are of the same age group, they may differ significantly. In the fall of 2010, for example, Audubon enrolled 147 new students who scored far below the basic level of proficiency.

The year-to-year changes in the API, for all the significance attached to them, actually convey no information about whether *individual students* are doing any better or worse.

Davis feared those students would “handicap the growth that the Audubon staff initiated this year and negatively impact the school's overall scores.”¹

Authorized by the California Legislature in 1999, the API places students in one of five performance bands, from “far below basic” to “advanced,” based on their performance on tests designed to cover the state's academic standards. The tests for math and English language arts contribute the most to the index, especially in elementary and middle schools.² A mathematical formula is applied to translate how many students are in each of those bands to produce a single number on a scale of 200 to 1,000. That formula gives schools an incentive to focus attention on the lowest-performing students, by crediting the school with more points for moving students up from the lowest band than it does for moving proficient students up a notch to advanced.

One of the original purposes of the API was to identify schools for a state intervention program, which doled out substantial sums of money to pay for expert assistance. Another was to identify high-flying schools to receive bonus money that could add up to as much as \$25,000 per teacher, a program that was ended because of its cost. A third was to spur improvement. Schools with APIs below 800 were given annual, though modest, improvement targets and not meeting them was supposed to indicate that something might be amiss.

Today, the API is part of the sales pitches of real estate agents, a discussion topic at PTA meetings, and the focus of school staff meetings. But, over the years, complaints about it have accumulated. The index is based entirely on standardized test scores and is highly correlated with socio-economic factors, which means that it reveals more about how many poor students attend a school than it does about the relative effectiveness of teachers or schools. Because it relies so much on English and math, the API is also widely blamed for curriculums that give social studies and science short shrift.³ The API also is said to give schools reasons to pay the most attention to students just below the cut-off score for the next performance band, to boost them up enough to earn the school points toward its overall score. Student and staff attendance data and graduation rates were supposed to have been included in the API, but they never have been. Today, nearly half of all California schools have

Today, nearly half of all California schools have APIs over the target number of 800, which means they are exempted from state scrutiny, even if they're doing little to help students who are not academically proficient.

APIs over the target number of 800, which means they are exempted from state scrutiny, even if they're doing little to help students who are not academically proficient.

Revamping the API to base it on changes in individual student test scores has been discussed in California policy circles for years, but nothing has been done. This is the case even though the quality of the state's tests and the availability of many years of student-level data mean it would be far easier for California to adopt such a measure than it would be for other states. Meanwhile, such growth measures have been widely adopted across the country.⁴ They are seen to be fairer than measures like the API, because they do a better job of identifying schools most needing help, making it easier to target resources. Data on individual students can also be used to distinguish teachers whose students are making the most progress from those whose students are making less, giving administrators insights into teacher performance. The Obama administration is particularly keen on focusing attention on individual student progress and has given states significant financial incentives to do so.

Despite the lack of movement on the state level, the Los Angeles Unified School District has over the past few years created its own accountability metric, which uses the same tests as the statewide system but is based on individual student achievement. By this measure, which is called Academic Growth Over Time, Audubon has made far more rapid progress than even was implied by the API.

Measuring Student Growth

The idea of measuring academic growth has been around for many years. The 1994 reauthorization of the federal Elementary and Secondary Education Act (then called the Improving America's Schools Act) required states to develop academic standards and to hold schools accountable for student achievement on tests linked to these standards. It called for schools receiving federal funds for disadvantaged students to show that their pupils were making "continuous and substantial, yearly improvement," such that they would become proficient in a reasonable time period.⁵

In 2002, No Child Left Behind replaced the Improving America's Schools Act. NCLB requires states to give a criterion-referenced test to students in grades three through eight and in one grade in high school. The tests are supposed to show whether students have become proficient in the knowledge and skills contained in states' academic standards. Each year, an ever-escalating number of students in each of 11 groups of students defined by race, family income, language, and learning disabilities have to pass the test for the school to make its AYP goal. In California, the AYP is based on the same data as the API but, like it, reports on different groups of students in different years.

The demands of NCLB escalate such that all students are supposed to be proficient by 2014. In the 10 years since NCLB was passed, more and more schools have fallen short of the targets they were supposed to hit. Schools that were making progress, but didn't reduce the percentage of students not proficient by 10 percent in a year, were required to give up some of their federal money to be used for tutoring and students were to be given options to attend other, higher scoring schools. But those same schools might be helping students who came to them two or three grade levels behind make great strides, even if they had not yet reached proficiency. To address concerns about fairness, in 2005 the Bush administration began allowing states to base their accountability systems on students' improvement rather than their absolute performance. Under those rules, 15 states have won federal approval to use student growth as a metric to satisfy NCLB.⁶ California considered modifying the API to make it a measure of student growth and explored a way

of doing it, but ultimately chose not to participate, because of technical and political concerns.

California considered modifying the API to make it a measure of student growth and explored a way of doing it, but ultimately chose not to participate, because of technical and political concerns.

California was bucking what would soon become a national movement. Growth measures now enjoy wide support, based largely on a boost from the Obama administration's \$4 billion Race to the Top competition as well as the rules for receiving part of the 2009 stimulus package. The administration wants states to use student achievement gains as an element of the evaluations of teachers, principals, and schools. More recently, requirements for states seeking waivers from much of NCLB also called on states to create accountability systems based on changes in student achievement. The administration also wants new assessments being developed by consortia of states operating with federal financial support to be "capable of measuring individual student growth" toward career and college readiness.⁷ The budget also calls for replacing the AYP measure, "which is based on a single, static snapshot of student proficiency on academic assessments" with a "broader, more accurate measure of student performance that looks at student achievement, student growth and school progress."⁸

In 2010, a national survey by the Council of Chief State School Officers found that 17 states were using growth models, 13 more were developing them and 11 were considering doing so.⁹ The National Council on Teacher Quality reported in October 2011 that 23 states now require teacher evaluations to reflect changes in student achievement.¹⁰ But there are a number of ways to build growth models, and, according to Douglas N. Harris, an expert on growth measures, some are not much different from

the static measures such as the AYP or API, which simply report on how students performed at a single point in time.

Harris argues that a certain type of growth model known as "value-added" is superior, although not perfect.¹¹ Like Galileo's law of momentum, students' past test scores tend to predict their future test scores. Value-added models are designed to compare students' actual performance to their predicted performance. The prediction is based on characteristics such as family income, language, and ethnicity, as well as their previous performance. The purpose of factoring in these other characteristics is to isolate the effects of influences on learning that a school cannot control from those it can. It doesn't take any fancy mathematics to know that, generally, a student who is a native English speaker and is raised in a well-educated, middle-class family is likely to have higher test scores than a student whose parents are poor, uneducated, and not fluent in English. But some teachers and schools are better than others at bending that learning trajectory upward. Conversely, a value-added accountability model can also help identify which schools and teachers are not serving their students well. The Academic Growth Over Time, or AGT, metric now used in Los Angeles is a value-added model.

Academic Growth Over Time

In the spring of 2009, the Los Angeles Board of Education directed Cortines to develop a more rigorous teacher evaluation system that would establish "measureable learning outcomes."¹² Soon after, the *Los Angeles Times* published a series of articles that detailed the difficulty of firing even the most dysfunctional, ineffective, and dangerous teachers. Those stories, which revealed teachers continuing to be paid even though they had been removed from the classroom, led to the creation of a task force charged with devising a new evaluation system based partly on student achievement.

During this reporting, the *Times* learned that LAUSD had a crucial asset that could help estimate the relative effectiveness of teachers: years of student-level test data that could be matched to demographic data, and back to students' teachers. The data made it possible to follow the academic progress

of individual students from year to year, teacher to teacher, and school to school. Using that data, the *Times*, in August 2010, published stories that ranked teachers from ineffective to most effective, based on an expert's value-added analysis.¹³ The *Times* also published the names of 6,000 teachers for whom it had calculated rankings.¹⁴ The resulting controversy overshadowed the fact that, from a value-added perspective, student achievement at some schools in affluent neighborhoods was lower than would be expected and that some schools serving lower-income students were helping them achieve far more than would have been expected. One outcome of the series put pressure on the district to use the data in its own teacher and school evaluations.

The concept of measuring performance with value-added formulas first gained attention in the early 1990s, when William Sanders, an agricultural statistics professor at the University of Tennessee, used test scores to estimate the relative effectiveness of individual teachers. Since then, researchers have refined and improved such measures. One leading source of that research is the University of Wisconsin, where LAUSD turned to help it come up with a measure accurate enough for evaluating schools, principals, and teachers. The result was the AGT.

Robert H. Meyer, who directs the Wisconsin center and is working with the schools in other states, is leading the work on the AGT. The basis for his work, he said, is “a practical theory of action that says it's good to track reliably whether kids learned what we want them to learn using tests that are sensitive to the curriculum. My whole orientation is thinking about whether I gave a good lesson and did the child learn.”¹⁵ Using data for all of the district's students over at least three years, the calculations underlying the AGT produce an estimate of how each student is likely to score on tests in each subject—math, reading, science, social studies, and history. That prediction is arrived at by factoring in the average effects on performance at all schools across the district of gender, race, family income, English fluency, housing situation, and mobility. The AGT also produces an estimate of predicted gains for each tested subject and school.

The districtwide average gain is expressed as a three on a five-point scale. The year Davis took charge at Audubon, its AGT score for English language arts

was literally off the chart—a six on the five-point scale, which meant that its students' gains were much higher than those of similar students across the district. The gains also were far higher than had been made by Audubon's students for the three years prior to 2010. The schoolwide math gain was smaller, and the school's math AGT was lower—3.7. But that was still slightly above the district average and significantly higher than it had been during the previous three years. This measure, more than any other, was evidence that Davis' efforts were paying off.

The value-added approach also solves another problem—the fact that California tests are grade-specific. Edward Haertel, the Stanford University professor who heads the technical group that designed the API, said the state's tests are aligned with those standards and, thus, quite different from one year to the next; in the parlance of psychometrics, they are not built on a continuous, or vertical, scale, which some argue makes it harder to precisely measure year-to-year growth.¹⁶

But Meyer says that value-added measures such as the AGT do not need tests built on a vertical scale. The AGT can measure the difference between students' and schools' predicted and actual scores using entirely different tests. For example, some students study general mathematics in the seventh grade, algebra in the eighth grade and geometry in the ninth grade—three different bodies of knowledge. What's more important than a vertical scale, he said, is the quality of the tests and the connection between the curriculum taught and what the test covers. He said California's tests are psychometrically solid and tightly related to the state's standards. In that sense, California is better positioned than many other states for adopting a value-added accountability system.

The AGT also provides much more information to parents and the public than does the one-dimensional API. Detailed, 26-page reports show how well a

**California is better positioned
than many other states for
adopting a value-added
accountability system.**

school is serving all types of students in all of the grades and all of the subjects that the state tests.¹⁷ So Davis, for instance, could see that the gains by eighth-grade algebra students at Audubon were relatively smaller than those made by algebra students with similar backgrounds districtwide. Consistent with that, he could see that students who had done well in the past in math also were making relatively smaller gains. That information helped him identify the challenges he had to address in subsequent years.

During Davis' second year, Audubon students continued to make larger gains than their peers across the district. For instance, the second-year rate of growth in math for all grades at Audubon jumped sharply—perhaps reflecting the fact that Davis had been able to replace most of the school's math teachers.¹⁸

Of course, growth is not all that matters. It also matters—a lot—whether students are reaching proficiency. During Davis' tenure the percentage of seventh-graders scoring proficient or advanced in English language arts has increased from 23 percent to 44 percent; the percentage of eighth-graders who are proficient or advanced jumped from 30 percent to 46 percent. In math, the gains were even steeper, rising from 23 percent to 36 percent for sixth-graders and 12 percent to 32 percent for eighth-graders.

When combined, these two measures—the AGT and proficiency rates on the California Standards Tests—provide the potential for interesting and useful analyses. The AGT displays the two types of data on what the district calls a two-by-two matrix. Imagine a square divided into quadrants. Schools that have low proficiency rates and low value-added scores are represented by dots in the lower left quadrant. Schools that are low in proficiency but are making strong gains are seen as dots in the lower right quadrant.¹⁹ The schools gaining the least are represented by red dots. Those gaining slightly more are yellow. Grays represent the district average for growth. Greens are better, and the schools adding the most value are blue.

In mathematics in 2007, Audubon was a yellow dot in the lower left quadrant—about 10 percent of the students were proficient, and the school was adding less value than similar schools across the district. Audubon's dot turns from yellow to gray and rises a

bit the next year. In the 2009–10 school year, the first year Davis was on duty, the dot keeps going up higher and moves into the lower right quadrant, meaning the students were making bigger gains than their peers in other schools. The next year, Audubon's dot becomes blue and proficiency continues to go up. Audubon's progress in English over four years is represented by a different set of dots marking a sharp trend upward. In the latest year, Audubon shows up as a blue dot in the upper right quadrant, meaning its proficiency rates and value-added scores are both well above average.

Davis' focus on rigorously evaluating teachers “changed the culture of the place,” said Joel Parks, a history teacher who chairs the school's teachers union chapter.²⁰ Although Davis offended some teachers, most welcomed the shift and were happy to see the most ineffective teachers gone. “Before, the math teachers gave homework, no one did it, and then they failed the students and explained it away by blaming the students for their failure,” Parks said. Test scores improved, he said, because Davis made the teachers teach.

Although Davis offended some teachers, most welcomed the shift and were happy to see the most ineffective teachers gone.

An English teacher at Audubon said she had been trying for years to find a teaching job nearer to her home. Now, she said, “I don't want to leave.” She attributes her change in attitude to Davis' leadership. She said she had been given many opportunities to improve her practice and be recognized for her efforts. “He set a standard for us and he expects us to meet it and, if not, you're gone,” she said.

To Davis, it's not magic. “Other schools come here and want to know the secret and there isn't anything for me to teach them,” he said, other than monitoring the performance of all students and supporting teachers. “It's just doing the work.”

In contrast to Audubon, Charles Drew Middle School which is in South Los Angeles, had the unwelcome distinction in 2010 of having one of the lowest AGT

Los Angeles is using the AGT to help it sort schools in a more precise and useful way than either the AYP or the API, making it easier to target interventions.

scores in the district. Ninety percent of Drew's 1,200 students are Latino and the rest African-American. More than eight in 10 students are poor.²¹ Proficiency rates are below 20 percent. According to interviews with former teachers, many classes, especially in math, were taught that year by substitutes. The school was on a year-round schedule, and some teachers would earn extra money by substituting during the time they were supposed to be off.

There was a constant churn of students from class to class due to errors in scheduling, the sudden departure of teachers, or attempts by some students to escape from classes with substitutes to get into ones with full-time teachers. Former teachers said that some newer staff members tried to ignore the chaos around them and got good results. But some older teachers would complain if a faculty meeting went five minutes over schedule. "Even if my kids made big gains," one former teacher said, "it was balanced out by other teachers not trying at all." Not surprisingly, Drew's dots were red or yellow showing its students were doing more poorly than had been predicted by the model for four straight years. This year Drew has a new principal and has abandoned the year-round schedule.

The AGT also can reveal weaknesses in schools that, based on average achievement, appear to be stronger academically. Taft High School, located in an affluent area of the San Fernando Valley, has a relatively high API score; nearly two thirds of its students are proficient or advanced in English language arts and nearly 60 percent are on grade level or above in math. But the school's 2011 AGT scores were far below the district average. Delia Estrada, Taft's new principal, said the AGT results "confirmed that the school had a problem with how it is teaching algebra. If you're just pushing out students and saying they can't learn math, it's not going to work."²² One bright spot that

she did not expect was biology, where the school's AGT score was far above the district average.

Estrada thinks the AGT will provide her with more useful data than will the results of the state's standards tests. "Are you doing right by each child, and what's a fair measure of that?" she asked. "The kids who are proficient and advanced, are they staying there and are you serving them well? If that's what it can tell us then it can be a very powerful way of measuring my work."

Los Angeles is using the AGT to help it sort schools in a more precise and useful way than either the AYP or the API, making it easier to target interventions. "We want to help highlight for folks some areas of strength and areas where there is opportunity for improvement," said Noah Bookman, a district official who has worked on the AGT.²³ If judgments about schools were based solely on the API or AYP, he said, "we might have had schools subject to major, costly intervention. But if you look at the same schools over time, they might be making some real, serious progress and should just be left alone."

Using value-added measures as one factor in evaluating schools is not controversial. The more data that is available, the smaller the random error that always accompanies testing. What is controversial is using student achievement data to evaluate teachers. In small part, this is due to the anger stirred up by the *Los Angeles Times* decision to publish teachers' names and ratings. Those articles galvanized statewide opposition to growth measures based on individual achievement, and the controversy spread nationwide as journalists in other cities requested similar information. Nonetheless, LAUSD, led by Superintendent John Deasy, has pushed ahead. The district wants a new evaluation system for the 2012–2013 school year that has, as one element, individual teachers' AGT scores.

In the spring of 2011, the district confidentially shared effectiveness ratings with individual teachers, asking them for feedback. Last fall, it gave the scores to the teachers' principals. Principals are not supposed to use the scores to evaluate teachers, but in exchange for small bonuses, about 1,000 teachers agreed to have their work evaluated this year partly based on the AGT. The arrangement has again raised the hackles of the United Teachers Los Angeles union.

Likewise, the California Teachers Association, which is one of two statewide unions that include UTLA members, opposes using student test scores in teacher evaluations at all. The union's position is that using student test scores in this way "is inherently flawed and meaningless."²⁴ Reliance on test scores, it says, "leads to teaching to the test and a narrowing of the curriculum." It also says that the "misuse of data threatens individual teachers' well-being, creates unhealthy school environments, and undermines otherwise effective evaluation systems." The problem with this argument is that few of those "otherwise effective evaluation systems" exist.

Revisiting the Academic Performance Index

The API was created around the same time that California was putting into effect rigorous new standards in math, science, English language arts, and social studies, and it satisfied several purposes. New tests aligned with the standards in each of those subjects were being developed, and students in grades two through 11 would take them. The tests generated an enormous amount of score data from each school, and the API was a convenient way to combine all of it into a single number that gave the public a crude proxy for student achievement and allowed schools to be ranked.

But it was not designed to give educators much help in analyzing school performance, and it told the public more about who attended each school than how well they were being taught. The shortcomings of the API have been known from the beginning. Over the years, various studies highlighted those problems and lawmakers have proposed legislative fixes. In recent

years, those fixes have been aimed at converting the index into a growth measure based on student achievement. In 2009 a legislative analysis found that, using the API, schools are "often unfairly held accountable for the low performance of the school the pupils previously attended."²⁵

A law passed that year required the state superintendent of public instruction to propose a way of fixing that problem. But no significant changes have been made, largely because of opposition from the powerful California Teachers Association. Not that the CTA is a fan of the API. Patricia Ann Rucker, a CTA lobbyist appointed to the state Board of Education by Gov. Jerry Brown, said she believes the API would be stronger if it were truly a growth measure. She also agreed that information about changes in individual student achievement would be valuable for teachers. But she said student achievement data should not be used by the state for any purpose.²⁶

Another reason the API has not been updated is a concern over the technical challenges involved, with some arguing that the standards and the tests they are based on would also have to also change, which would be costly. Depending on how it were done, changing the API could make historical comparisons impossible.

Last year, Senate President Pro Tem Darrell Steinberg introduced legislation to modify the API so that it measured "group and individual academic performance growth by utilizing individual pupil results from a longitudinally valid achievement assessment system."²⁷ That idea was quickly dropped due to opposition. The bill that eventually passed the Legislature would have created the Education Quality Index, which would have incorporated dropout rates and measures of career and college readiness and pupil engagement and given more weight to subjects other than reading and language arts. The bill would have established growth targets for schools and groups of students, but the gains would not be measured using changes in individual students' scores. The CTA did not oppose the legislation in its final form.

Still Brown vetoed the bill, citing his antipathy toward using test scores to judge schools. "Lost in the bill's turgid mandates," he wrote, "is any recognition that quality is fundamentally different from quantity."²⁸

The shortcomings of the API have been known from the beginning. Over the years, various studies highlighted those problems and lawmakers have proposed legislative fixes.

Arun Ramanathan, executive director of the reform group Education Trust West, said the veto sent a message that Brown “has rejected the use of data in any way, shape or form, and that sort of boggles the mind.”²⁹ Steinberg also is the author of a new bill that once again seeks to supplement the API with other measures, such as the percentage of students advancing to the next grade in middle and high school.³⁰ The bill also calls for school inspections to add human judgment to the process. The API would be retained, but it would only count for 40 percent of a school’s overall rating. It would continue to compare different groups of students and would reveal little about whether the teaching at one school is any better than the teaching at another.

Why California Should Replace the API Now and How It Could Be Done

What Steinberg should propose instead is to scrap the API entirely as the student achievement measure and to replace it with a schoolwide value-added system that incorporates proficiency rates. The API once served a purpose, but today it is simply a placeholder for a real accountability system. A higher index number indicates almost nothing about whether the teaching at one school is any better or worse than the teaching at another. The API is, to a large extent, an indicator of students’ wealth rather than of a school’s educational quality. It places overwhelming emphasis on math and reading, which results in an under-emphasis on science and social studies. And because more than 40 percent of California schools have API scores at or above the state minimum, they no longer have to worry about helping students who are not yet proficient reach that goal. That means that schools that enroll more affluent and better performing students could rest on the laurels of their students and let the quality of teaching slide.

Now is the time to fix those problems because the current policy context presents California with an opportunity to make significant changes that might be less possible later on. One reason is NCLB’s looming deadline. By 2014, students in grades that are tested are supposed to be proficient in English and math. Already, 2,500 California schools have failed to hit

What Steinberg should propose instead is to scrap the API entirely as the student achievement measure and to replace it with a schoolwide value-added system that incorporates proficiency rates.

their marks for seven years and, under NCLB, they should be pressured to make dramatic changes.

California could seek a waiver from much of NCLB, and in May the state Board of Education voted to do that. But that request would have to receive special consideration because the state is resisting meeting the requirements set by the U.S. Department of Education. Specifically, the state would have to develop a system that identifies schools for either interventions or rewards, based in part on changes in individual student achievement. It would also have to put in place teacher and principal evaluations based in part on students’ test scores.

The state is not offering to make any of these changes. But failing to do so could be costly. Education spending in the state has already been cut dramatically—and will be further cut without tax increases. A waiver would free up the estimated \$350 million annually that school districts must now set aside for outside tutoring or to pay for students at low-performing schools to transfer elsewhere.

Michael Kirst, a professor emeritus at Stanford University and president of the state Board of Education, argues that too many other policy

The current policy context presents California with an opportunity to make significant changes that might be less possible later on.

changes are pending—such as the implementation of Common Core State Standards and new tests that are supposed to go into use in 2014—that it would be better to leave the API as it is for now.³¹ But looked at another way, the next two years provide the state with a great opportunity to begin an important transition.

Richard Wenning, an architect of the Student Growth Percentile model that was pioneered in Colorado and is now being adopted or explored by 20 states, recommends that a new accountability system based on a growth measure be phased in, with schools held accountable for producing gains before teachers or principals are. One way to do that would be to use a growth measure for a few years to identify schools, such as Audubon, that are improving quickly, even if their proficiency rates are low. A school could be helped by having strong growth, but it couldn't be hurt if it was judged not to be making sufficient progress. Both the API, and a new measure, could continue to be reported. The low-stakes pilot would allow the state to iron out any bugs that occurred, as is being done in Los Angeles.

One of the consequences of shifting to the new Common Core assessments is that trend lines could be lost; each school would start with a clean slate. Growth measures require two or three years of data to construct, so if the state switched to new assessments in 2014, it could be 2017 before the state had an accountability and evaluation system based in part on student achievement. If the state adopted a value-added system now, however, the transition would be much faster and smoother. Students' performance on the current tests could be used to predict their performance on the new assessments. Meanwhile, other steps, such as explaining the new system to teachers, principals, and parents, could be put in place. "It's about political will," Wenning said.³² "It's not technically hard at all, but you need to be committed to building a next-generation accountability system that produces a consistent source of information about how children are progressing."

Alternatively, if dropping the API and creating an entirely new mechanism is too big a step, the state Board of Education could create a state waiver, similar to what the federal government is offering.

Local school districts could seek exemptions from the API and be offered incentives to create their own growth measures. The state would continue to gather and report on proficiency rates, so it would still be possible to see how schools and districts are doing against a uniform objective benchmark. But the districts would also report on their value-added scores or other types of growth measures. The state, or private foundations, could fund an external evaluation of the different models so that California could learn lessons for rebuilding a statewide system.

Already, other schools besides the Los Angeles school district are exploring or implementing a value-added or growth measure of student progress. Four charter management organizations that together operate about 70 schools in Los Angeles are using student-level growth models in their teacher evaluations. Nithya Rajan, the interim executive director of the organization, known as The College-Ready Promise, said the organization is proceeding cautiously; before attaching consequences to the scores, it is trying to determine whether teachers who are highly rated by that measure also score well in observations of their practice. "This is a big change... so we wanted to be sure we had confidence in the data," she said.³³

Los Angeles is also part of a consortium called the California Office to Reform Education, which includes San Francisco, Long Beach, Fresno, and Sacramento, districts that are also interested in adopting a growth model. "I can't wait around for the state," says Los Angeles Superintendent Deasy. "I can make much more informed decisions because I have a trajectory. If I can keep improving I can get there. Then, it's about pace."³⁴

Deasy acknowledges that the AGT is far from perfect. There is no way, for instance, to factor in the

Alternatively, if dropping the API and creating an entirely new mechanism is too big a step, the state Board of Education could create a state waiver.

“dosage” of instruction that students are receiving. If students are absent a lot, it’s unfair to hold the school accountable for their learning. A related problem has to do with how to accurately attribute students’ learning to individual teachers. For example, students who are far behind may get extra tutoring or be grouped with others and taught by another teacher. Or students could be taught by teachers working in teams. How much did each teacher contribute? The growth model can’t tell. At the same time, students can show up on the wrong rosters or transfer from one class to another in the middle of the year.

Partly because of such problems—issues that Los Angeles is working out—Deasy thinks that data should account for only about 30 percent of a teacher’s evaluation; in his view, clinical observations by principals or other educators should carry the most weight. Yet without data about growth over time, he says, the picture is dangerously incomplete. “Without this added information...I am steering a ship with a major piece of radar not working,” he said. “Radar tells you what’s ahead and sonar tells you what’s beneath. Without both of them, I’m not helping the system and the system is not helping individual students.”

Notes

1. DeWayne Davis, in discussion with author, Spring 2011.
2. California State Board of Education meeting agenda May 9–10, 2012. Item 6. Accessed on May 21, 2012 at <http://www.cde.ca.gov/be/ag/ag/yr12/agenda201205.asp>
3. *Untapped Potential: The Status of Middle School Science Education in California*, Lawrence Hall of Science and SRI International (San Francisco, CA: The Center for the Future of Teaching and Learning at WestEd, 2012) http://www.cftl.org/documents/2012/SCI_fullReport.pdf
4. Kevin Carey and Robert Manwaring, *Growth Models and Accountability: A Recipe for Remaking ESEA* (Washington, DC: Education Sector, May 2011).
5. Dianne Piche, et al., “Title I in Midstream: The Fight to Improve Schools for Poor Kids,” *Report of the Citizens Commission on Civil Rights*, Corinne Yu and William L. Taylor, eds. (Washington, DC: Citizens Commission on Civil Rights, Summer 1999).
6. Alaska, Arizona, Arkansas, Colorado, Delaware, Florida, Iowa, Michigan, Minnesota, Missouri, North Carolina, Ohio, Pennsylvania, Tennessee, Texas
7. “A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act,” U.S. Department of Education, March 2010, <http://www2.ed.gov/policy/elsec/leg/blueprint/publication.html#part1>
8. President’s FY 2013 Budget Request for the U.S. Department of Education, available at <http://www2.ed.gov/about/overview/budget/budget13/index.html>
9. Rolf K. Blank, “State Growth Models for School Accountability: Progress on Development and Reporting Measures of Student Growth,” Council of Chief State School Officers, 2010. Accessed on October 28, 2011 at http://www.ccsso.org/Documents/2010/State_Growth_Models_2010.pdf
10. “Trends and Early Lessons on Teacher Evaluation and Effectiveness Policies,” (Washington, DC: National Council on Teacher Quality, October 2011) http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf
11. Douglass N. Harris, in discussion with author, Fall 2011.
12. Los Angeles Unified School District, “Quality Leadership and Teaching to Ensure a World Class Education for All,” available at <http://sae.lausd.net/sites/default/files/Teacher%20Effective%20Task%20Force%20Resolution.4.28.09.pdf>
13. Richard Buddin, “How Effective Are Los Angeles Elementary Teachers and Schools,” *Los Angeles Times*, August 2010. Accessed on October 31, 2011 at <http://documents.latimes.com/buddin-white-paper-20100908/>
14. The author was formerly the director of the Hechinger Institute on Education and the Media at Teachers College, Columbia University. The Institute helped offset the cost of Buddin’s analysis. The author was not involved in or consulted regarding the content of the stories made by the newspaper, including the decision to publish the names of teachers.

15. Robert H. Meyer, in discussion with author, Summer 2011.
16. Edward Haertel, in discussion with author, Fall 2011.
17. Reports available at <http://portal.battelleforkids.org/BFK/LAUSD/Home.html?sflang=en>
18. 2010 and 2011 AGT School Reports for Audubon Middle School, <http://portal.battelleforkids.org/BFK/LAUSD/Home.html?sflang=en>
19. Visit link to see interactive graphic, <http://matrix.escmatrix.com/lausd/swf/loader.jsp>
20. Joel Parks, in discussion with author, Spring 2011.
21. School level data for Charles Drew Middle School available online at <http://dq.cde.ca.gov/dataquest/Cbeds4.asp?FreeLunch=on&PctHisp=on&PctBlack=on&cSelect=CHARLES%5EDREW%5EMIDDLE--LOS%5EANGELES%5EUNI-1964733-6057962&cChoice=SchProf1&cYear=2010-11&cLevel=School&cTopic=Profile&myTimeFrame=S&submit1=Submit>
22. Delia Estrada, in discussion with author, Fall 2011.
23. Noah Bookman, in discussion with author, Fall 2011.
24. California Teachers Association Teacher Evaluation Workgroup, "Teacher Development and Evaluation Principles," May 2011. Accessed on November 1, 2011 at <http://www.cta.org/Issues-and-Action/Teacher-Quality/Teacher-Evaluation-Principles.aspx>
25. California Senate Education Committee analysis of Assembly Bill 1130 of 2009. Accessed on October 31, 2011 at ftp://leginfo.public.ca.gov/pub/09-10/bill/asm/ab_1101-1150/ab_1130_cfa_20090622_162536_sen_comm.html
26. Patricia Ann Rucker, in discussion with author, Fall 2011.
27. CA Senate Bill 547—2011–2012 Regular Session, as introduced, available at <http://e-lobbyist.com/gaits/text/179864>
28. Letter from Governor Edmund G. Brown to the members of the California State Senate, October 8, 2011. Emphasis in original. <http://toped.svefoundation.org/wp-content/uploads/2011/10/API-SB547VetoMessage100811.pdf>
29. Arun Ramanathan, in discussion with author, Fall 2011.
30. Senate Bill No. 1458, available at http://www.leginfo.ca.gov/pub/11-12/bill/sen/sb_1451-1500/sb_1458_bill_20120224_introduced.pdf
31. Michael Kirst, in discussion with author, Summer 2011.
32. Richard Wenning, in discussion with author, Summer 2011.
33. Nithya Rajan, in discussion with author, Fall 2011.
34. John Deasy, in discussion with author, Fall 2011.