



PPIC

PUBLIC POLICY
INSTITUTE OF CALIFORNIA

Designing California's Next School Accountability Program

October 2014

Paul Warren

Summary

California is in the midst of a major K–12 reform effort. In 2010, the state adopted the Common Core State Standards (CCSS), which outline what students should know in mathematics and English. In 2013, it adopted tests of the new standards developed by the Smarter Balanced Assessment Collaborative (SBAC). These tests will be administered beginning in 2015, replacing the California Standards Tests (CSTs). In addition, the state revamped its school-finance system in 2013, creating the Local Control Funding Formula (LCFF) to streamline local funding and increase support for disadvantaged students. The LCFF also requires districts to set performance targets on a range of school and student success indicators as part of a district Local Control Accountability Plan (LCAP).

Less attention, though, has been paid to how these developments affect state and federal accountability programs. There are now four K–12 accountability programs operating in California, each with its own strengths and weaknesses. The sheer multiplicity of goals and performance indicators is confusing. California can do little to change the federal accountability program, but it can—and should—revise the state’s accountability programs so they send strong, consistent signals that student achievement is the core objective of the K–12 system. Our analysis of the strengths and weaknesses of the current programs leads us to propose several steps that merge state and local accountability programs and create a more straightforward approach to improving schools and student outcomes.

First, California should create a new state measure that would align with the LCAP program. The new state measure should be simple, statistically valid and reliable, and create strong incentives for schools to focus on student success. We developed an option that assesses school and district performance from several perspectives. Almost all of these data are already in LCAP, and indicators used to construct the state performance measure would become LCAP priorities.

Our measure has five types of indicators. Each provides a different perspective on school and district quality and student success. The measure includes current achievement levels (student test scores) as well as indicators of persistence in school (attendance, dropout, and graduation rates). Also included are indicators of student readiness for kindergarten and for “reading to learn,” beginning in 4th grade. The measure also evaluates longer-term success, including whether academic gains persist over time and the track record of students beyond school in college and career. Finally, teacher and student survey data are used to evaluate whether schools are organized to promote student achievement. Together, these indicators create incentives for schools and districts to address student needs, teach in ways that promote long-term benefits to students, and stress good management of schools and the instructional process.

Second, the state should develop and fund a larger program of technical assistance to school districts. Currently, the state dedicates only \$10 million in federal funds to technical assistance. This amount should increase substantially. To address the issues involved in this effort, the California Department of Education (CDE) should develop a multiyear plan for the types of services districts need, the amount of funding the state should make available, and the delivery of assistance.

Third, the state needs to tinker with the governance arrangements of accountability programs. The LCAP process should grow into an annual local review of district strengths and weaknesses, with county offices empowered to prod districts to improve each year. For districts with moderate problems, technical assistance would be optional. For districts with more severe problems, technical assistance would become more directed

as needed to protect the interests of students. The state also should create consequences for low-performing districts (rather than schools) that fail to improve after significant investments of technical assistance.

Our proposal is only one of many possible ways to address issues with California's accountability programs. Our approach builds on recent reforms that emphasize local accountability and makes improving student achievement the priority. This would establish a focus for district LCAPs while also recognizing local priorities. Such a capacity-building strategy, though, will not offer quick results. Districts will have to be willing to consider new ways of operating, and county offices would have to accept new responsibilities in the instructional arena. But if our analysis is correct, approaching accountability as a learning process offers a greater likelihood of significant improvement over time than the current system.

Contents

Summary	2
Figures	5
Tables	5
Abbreviations	6
Introduction	7
Major Features of K-12 Accountability Programs	8
Goals and expectations	8
Communication, assistance, and consequences	10
Rethinking the Premises of NCLB	11
NCLB and AYP	12
Keep focus on the goal, measure with multiple indicators	13
Build short- and long-term quality indicators into the process	15
Use accountability programs to build capacity	16
Conclusion	18
California's Accountability Programs	20
Public School Accountability Act	20
Local Control Funding Formula	23
The CORE NCLB waiver	27
Where to Go From Here	31
Indicators of performance	32
Creating good incentives	37
Linking state and local accountability	38
Conclusion	41
References	43
About the Author	46
Acknowledgments	46

Figures

1. 4th grade NAEP scores do not show the same gains as the 4th grade CST 14

Tables

1. Major components of a K-12 accountability program 8
2. How AYP is calculated 12
3. API methodology 20
4. LCAP performance indicators 24
5. CORE performance indicators 28
6. State and local accountability indicators 32
7. Five elements of a new state accountability measure 38

Abbreviations

API	Academic Performance Index
AYP	Adequate Yearly Progress
CCSS	Common Core State Standards
CDE	California Department of Education
CORE	California Office to Reform Education
CST	California Standards Test
DAIT	District Assistance and Intervention Team
IIUSP	Immediate Intervention in Underperforming Schools Program
LCAP	Local Control Accountability Plan
LCFF	Local Control Funding Formula
NAEP	National Assessment of Educational Progress
NCLB	No Child Left Behind (Elementary and Secondary Education Act, 2001)
PSAA	Public School Accountability Act
SBAC	Smarter Balanced Assessment Consortium
SQII	School Quality Improvement Index

Introduction

For more than a decade, California K–12 school performance has been evaluated using two accountability programs: the Public School Accountability Act (PSAA) and No Child Left Behind (NCLB). These programs rely on two separate but related measures of school performance. The Academic Performance Index (API) is the primary *state* measure of school and district performance, in use since 1999. Operating as part of the PSAA, the API uses test scores from students in grades 2–11 to calculate the growth in student achievement in schools and districts. Under NCLB, California schools are also evaluated with a *federal* performance measure known as Adequate Yearly Progress (AYP). AYP measures whether sufficient proportions of students in grades 3–8 and one grade in high school achieve at a state-determined “proficient” achievement level on state tests. Both the API and AYP primarily rely on the California Standards Tests (CSTs), which assess student achievement in English and mathematics in grades 2–11 and science and history/social science in selected grades.

In 2013, two new K–12 accountability programs were established in California that will operate alongside the existing programs. The Local Control and Accountability Plan (LCAP) was created as part of the state legislation that established the Local Control Funding Formula (LCFF). In addition, the federal government tentatively approved a new federal accountability program for eight California school districts under a federal waiver of NCLB. The new federal accountability measure—called the School Quality Improvement Index (SQII)—would replace AYP as the federal performance measure for these districts if the waiver receives final federal approval.¹ Both LCAP and SQII include state test results along with a range of other student outcomes.

All four accountability programs will be affected by the implementation of the Common Core State Standards (CCSS). The new standards will be effective in California beginning in fall 2014, and new tests—known as the Smarter Balanced Assessment Consortium (SBAC) tests—that are aligned with the new standards will replace the CSTs in spring 2015. Because state test results represent core accountability data, the new tests will require modifications to all four programs.

As a consequence, this is a good time to reconsider the state’s long-term plan for holding schools accountable for the progress of students. The API is 15 years old, and the technology of accountability has improved significantly since it was first developed. At the federal level, AYP is being replaced by state-developed measures under a program of waivers from NCLB requirements. At some point in the future, California may be required to develop a new accountability measure under a reauthorized NCLB. With new tests on the horizon, the state could develop a new accountability measure with the long-term goal of replacing both the API and AYP.

This report reviews the state’s options for the next generation of K–12 school accountability programs. First, we describe the major elements of accountability programs. We then develop guidelines for accountability programs that address problems with the design of NCLB. We also examine the three additional programs currently used in California, highlighting the strengths and shortcomings of each. The last section outlines a state accountability measure and program that includes a broader range of student outcomes than just test scores and also aligns with our design guidelines.

¹ The SQII would apply to only seven districts, because Sacramento City Unified School District withdrew from the waiver in 2014.

Major Features of K–12 Accountability Programs

Accountability programs are complex, having many interconnecting components. The design of accountability programs begins with a “theory of change” — a set of facts and assumptions that outline a credible plan for how an accountability program will improve educational outcomes. As shown in Table 1, there are five main parts of K–12 accountability programs. Goals and expectations establish the measurement framework for evaluating the progress of students, schools, and districts. Communication, assistance, and consequences provide the incentives and tools to help educators take the steps necessary to improve student achievement.

TABLE 1
Major components of a K–12 accountability program

Goals. The system’s foundation is established by defining what the program seeks to accomplish and for which groups of students.

Expectations. The design of the accountability measure translates goals into outcomes, sets priorities among the goals, and defines expected progress for schools and districts.

Communication. Publicizing school and district accountability scores enlists the support of parents and local communities in school improvement.

Assistance. State funding and technical assistance support local efforts to improve schools and districts.

Consequences. The state uses its power to intervene in districts where local pressure and assistance prove insufficient to improve schools.

SOURCE: Adapted from Perie, 2007.

Goals and expectations

Goals and expectations define what policymakers expect of schools. Goals identify outcomes the program is intended to improve. Goals may include increasing academic proficiency, helping more students graduate, or improving success in college and postgraduation employment. NCLB, for example, sets a goal of academic proficiency in mathematics, English, and science. Reducing performance gaps among groups of students also represents a key goal. Goals must be translated into specific measurable objectives. All current accountability programs require schools to meet performance targets for subgroups based on race/ethnicity, education status (e.g., special education and English Learner), and economic status (e.g., low income). By clarifying objectives and priority populations, the program creates a basic matrix of desired outcomes.

Accountability Terminology Used in This Report

Program refers to the five components that typically are part of educational accountability systems—goals, expectations, communication, assistance, and consequences.

Measure is the formula used to assess school and district performance.

Indicator represents the individual outcome variables that are used to calculate school and district performance.

Validity and reliability. The development of an accountability measure or formula translates goals into concrete program expectations. This begins with data on school and district performance. These indicators, such as test scores and graduation data, must be *valid* and *reliable*. Valid data means that the numbers provide meaningful information on the activity being measured. For instance, cheating on tests makes scores invalid because the results no longer represent what students can actually do. Valid data also requires common definitions so that reported numbers are comparable from school to school. Reliable data means that the information is sufficiently accurate for using in a school or district performance measure. Because a significant number of schools are quite small, data reliability is a significant issue.

The concepts of validity and reliability also apply to accountability measures. Broadly speaking, validity is a judgment about whether specific data are appropriately used to draw sound conclusions about something that cannot be directly measured.² There are several types of validity, but for the purposes of this report we discuss two problems that threaten the validity of accountability measures. First, indicators of local performance that are not valid or reliable threaten the validity of the overall accountability measure. The second is the concept of *construct validity*, or coherence, which says that a test is valid to the extent to which it actually measures what it claims to measure (Brown, 2000). Imagine a measure of school quality that contains test scores, expulsion rates, and the number of teachers. Just how does the number of teachers contribute to school quality? Though school size may influence school quality, including the number of teachers in an accountability measure would make it difficult to interpret what the school scores on this measure mean about a school, that is to say, it threatens the measure's validity.

Accountability measures reflect policy. In addition to these technical issues, accountability measures also involve important policy decisions. For instance, measures may be designed as a series of performance tests that are separately assessed (known as *conjunctive* scoring).³ Alternatively, *compensatory* scoring combines several indicators into an index, which results in only one performance hurdle. In general, there is no “right” or “wrong” choice. Conjunctive scoring ensures that schools satisfy the policy objective of each performance test. Compensatory designs, by averaging the various indicators, allow high performance on one objective to offset substandard scores on another. Compensatory measures also require policymakers to set priorities by assigning weights to the different indicators. Another critical policy choice involves whether schools are

² Or, stated more precisely: “Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment” (Messick, 1990).

³ AYP, for instance, requires schools (and significant subgroups) to meet performance targets in mathematics and English. Failure to attain one of these subject-area targets means the entire school does not make AYP.

measured by the level of performance, or growth in performance—or both—and the standard for determining what is “adequate” performance.

The accountability measure represents a complex mix of technical and policy factors. The technical task is to develop an accountability measure that itself is a valid and reliable indicator of school and district performance. From a policy perspective, the measure is the state’s clearest expression of its aspirations for students. The goal is to make the technical and policy features work in harmony so that the measure creates strong incentives for educators to act in the best interests of students.

Communication, assistance, and consequences

The design of accountability measures also must work in concert with program supports and consequences. Accountability programs generally contain three components that shape how the program works at the local level. One component is communication: informing teachers, administrators, parents, and communities about school performance and rankings. All communities want children to attend good schools. Accountability data provides information about the quality of local schools and empowers parents and others to press for improvements.

A second component in raising local performance is assistance: helping low-performing schools or districts improve. This option assumes that schools or districts do not know how to meet the state’s performance goals, or have too few resources to make changes necessary to improve. Technical assistance or additional funding addresses these problems.

A third element imposes penalties on schools or districts that perform below expectations. Such consequences may include sanctions designed to increase external pressure for improvement. For instance, under NCLB sanctions include restricting local financial flexibility, providing tutors for students, replacing teachers and staff, and converting a school into a charter school. Consequences also can take the form of more-directed assistance, in which schools or districts are required to work with an outside advisor, implement a new curriculum, or take other specific steps to address issues that undermine student achievement.

Like accountability measures, the program of communication, assistance, and consequences requires balancing a variety of technical and policy issues. For example, are districts or schools the target of assistance and consequences? School-level reforms generally assume that teachers and principals are most directly responsible for the quality of classroom instruction. But programs also target *districts* due to the important governance and regulatory roles of district boards and administrations. Other factors also can enter into this strategy. In part, the use of sanctions depends on the program’s underlying assumptions about the willingness of local educators to align with the state’s goals. Another factor is the extent to which parents can successfully pressure educators to make needed changes.

The two main components—the accountability measure and the support and sanctions strategy—must work together. The program must create reasonable expectations for educators that help build a strong local constituency for improvement. This allows the “carrots” of funding and technical assistance to build the local capacity needed to deliver a quality product. It also reserves the “sticks” of sanctions for schools or districts where local accountability fails to overcome local forces that undercut quality.

Rethinking the Premises of NCLB

Unfortunately, there is little evidence to guide states about how to design the five components of accountability programs so they are most effective in improving student achievement. We have evidence that NCLB was partially successful. A national evaluation of the program found significant increases in 4th grade mathematics, with smaller increases in 8th grade mathematics. Gains for low-income students were significant in both grades. Researchers found no consistent impact on 4th grade reading (Dee and Jacob, 2011). However, despite these gains NCLB did not come close to reaching its goal of 100 percent proficiency, as more than half of all students scored below the proficient level in mathematics and reading on the 2013 federal National Assessment of Educational Progress (NAEP).⁴

Researchers also acknowledge that there is an inadequate understanding of how the various components of accountability programs work together to improve the condition of students. In reviewing the impact of test-based accountability on student achievement, the National Research Council's Committee on Incentives and Test-Based Accountability in Public Education concluded that "policy makers do not yet know how to use test-based incentives to consistently generate positive effects on achievement and to improve education" (Hout and Elliot, 2011).⁵

Deming and Fullan. Thus, a review of accountability programs needs to go beyond a discussion of the measures and indicators used to evaluate performance and examine each component, including the foundational assumptions (or theory of action), of the programs. In this section, we examine selected features of NCLB to develop principles for the design of accountability programs.

This analysis is influenced by the work of W. Edward Deming, who developed a set of principles and processes for restructuring organizations to increase quality and reduce costs. These principles do not constitute a cookbook for improvement. Instead, they identify the critical perspectives and attributes of organizations that boost quality by continually improving.

Deming's principles are founded on the idea that an organization creates rules and practices—a "system" in Deming's vernacular—to run its operation. That system is responsible for the success or failure of the organization. If the process results in bad decisions or faulty products, it generally reflects a problem with the organization's system. Deming's principles call for an organization to first clearly identify its goals. Then, the organization uses data and analysis to increase the quality and efficiency of each component of its system.

Deming's work provides the framework for a variety of quality-improvement efforts, including the federal Baldrige management awards. In the 1990s and early 2000s, a number of states encouraged school districts to use the Baldrige process to improve student achievement. Evaluations show that districts using the Baldrige guidelines achieve significant process improvements, such as higher attendance or fewer dropouts. Unfortunately, there are no rigorous assessments of the impact of these principles on student achievement (Walpole, 2002). Deming's guidelines have been shown to be effective in other service industries. For instance, hospitals successfully used the principles to improve service quality and customer satisfaction by creating a culture focused on continual improvement (Shortell, 1995).

⁴ The National Assessment of Education Progress is administered to a representative sample of students in each state (<http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf>).

⁵ Hout and Elliot, page 92.

Michael Fullan, an educational theorist from Canada, also is influenced by Deming’s principles. Fullan, a former dean of the education school at the University of Toronto and an author of many studies and books on systemic change in K–12 education, envisions a continual improvement process that, over time, generates significant gains (Fullan, 2011). Capacity building—helping administrators and teachers get better at their jobs—is the main avenue of improvement (Fullan, 2008). Increasing the quality of instruction is done in groups, getting teachers to collaborate within and between schools (Fullan, 2011). Fullan is currently advising the California Office to Reform Education (CORE), which works with the eight districts in California that were granted the NCLB waiver in 2013, as well as several individual districts in California.

Below, we discuss three lessons that emerge from an analysis of the problems with NCLB. Two provide guidance about the number and types of accountability indicators that are needed to balance incentives at the local level. The third finding discusses the role of sanctions—what level of schooling should be held accountable, and the role of building school capacity as a substitute for sanctions. First, we provide a brief description of the performance measure used in the federal program, Adequate Yearly Progress, known as AYP.

NCLB and AYP

AYP measures whether sufficient proportions of students in grades 3–8 and one grade in high school achieve at a state-determined “proficient” achievement level on state tests.⁶ NCLB required states to develop tests in English, mathematics, and science. States set a “proficient” level of achievement on those tests that satisfied state learning standards. Annual school targets are based on a set percentage of proficient students. To “make” AYP, the school-wide percentage and the proportion of each significant subgroup must meet or exceed the target each year (see Table 2). In addition, AYP uses conjunctive scoring, so schools must exceed the target for the school and all its subgroups in both English and mathematics. Federal rules require school and district targets to increase over time, ultimately reaching 100 percent in 2014.

TABLE 2
How AYP is calculated

School test results	
Performance level	Percent of students
Advanced	20
Proficient	30
Basic	20
Below Basic	10
Far Below Basic	20

Theory of action. NCLB is based on the idea that all students should perform at a proficient level. This goal is built into the program through its measure, AYP. To reach that goal, NCLB provided additional funding for districts and states. The program also contains very specific timelines for improvement and sanctions for failing to meet performance targets. Though the law gave states considerable leeway to define “proficient,”

⁶ AYP is somewhat more complex than described here. In addition to test scores, high schools also must improve graduation rates. Middle and elementary schools also must have an additional performance indicator. In California, these schools must make progress on the API. In addition, a school also can make AYP if it reduces the proportion of students scoring below the proficient level by 10 percent from the prior year. This provision is known as “safe harbor.”

California had already established its academic standards. As a result, proficiency in California was set at a relatively high level.⁷

Keep focus on the goal, measure with multiple indicators

AYP measures school academic performance primarily using annual state test scores. The simplicity of this measure creates strong incentives for teachers to develop ways to boost student scores. Unfortunately, these incentives do not always result in real increases in learning. To channel these efforts in more productive directions, policymakers need to consider how to protect the quality of data used in accountability measures.

Accountability pressure leads to distorted data. Deming’s work calls for organizations to maximize quality and minimize long-term costs. Deming believed that focusing on short-term profits encourages organizations to cut corners, which can reduce quality and lead to additional long-term costs (such as repairs under warranty). In K–12 education, NCLB implicitly identifies long-term costs as students with remedial needs (students performing below the proficient level) and students who drop out before graduating. Minimizing these costs seems like a reasonable long-term goal for the system.⁸

The mere act of measuring quality, however, can lead employees to take actions that make the data look better but do not result in actual quality increases. This problem has long been recognized as Campbell’s law: “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor” (in Nichols and Berliner, 2005).⁹ In education, studies have found that teachers often respond to high-stakes testing by focusing on the skills and knowledge tested each year (Shepard, 1990). This “teaching to the test” approach can generate score inflation, which is defined as growth in student scores that does not reflect gains in student skills and knowledge (Center on Educational Policy, 2008). In other words, if the practice of teaching to the test is widespread, CST scores will overstate what students actually know and thus are not a good reflection of long-term quality.

In fact, California scores show evidence of test inflation. Figure 1 shows CST scores in 4th grade English compared to results of 4th grade California students on the federal NAEP. Unlike the CSTs, teachers have little information about the NAEP test, and no consequences are tied to student performance on the federal test. The intense focus on the CSTs translates into large gains over time, whereas NAEP scores increase only marginally. Other factors may also affect the gap in scores: NAEP standards emphasize somewhat different content than California standards, for instance.¹⁰ But the very modest growth in NAEP results over the past decade are an indication of test inflation in CST scores.

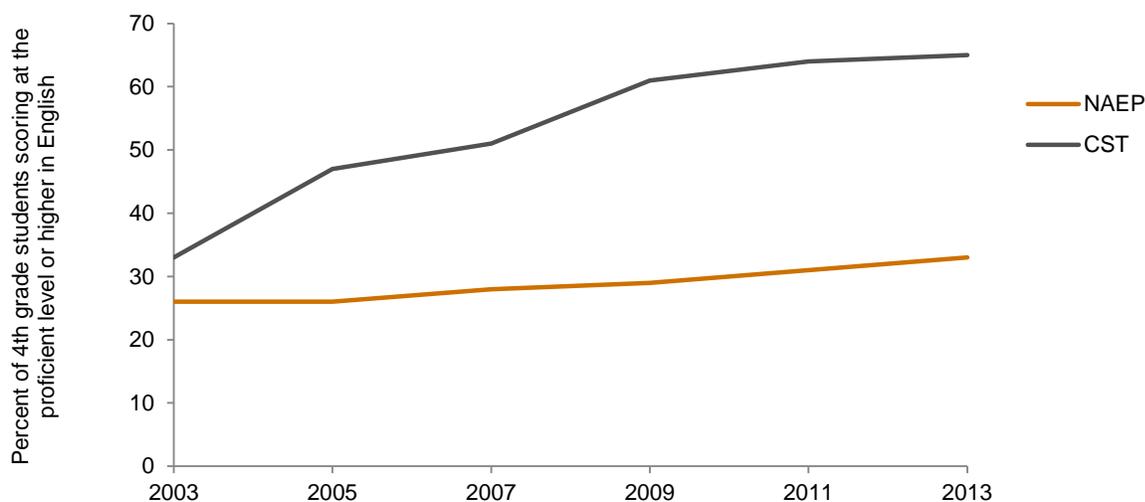
⁷ For instance, California was one of six states receiving a score of B or better on an evaluation of rigor by Paul Peterson and Frederick Hess (*Few States Set World-Class Standards*, Education Next, Summer 2008). The Fordham Institute also rated California’s standards highly (see, for example, *The State of State Math Standards*, by Klein, Parker, Quirk, Schmid, and Wilson at www.math.jhu.edu/~wsw/ED/mathstandards05FINAL.pdf).

⁸ Growth in achievement is also considered a long-term goal, both as a measure of progress towards proficiency and as a sign of narrowing gaps in performance among significant subgroups.

⁹ Nichols and Berliner, page 4.

¹⁰ The California Department of Education advises against direct comparisons between CST and NAEP results because of differences in the content and purpose of the tests. In this case, however, we are not comparing the two tests (and results) but, rather, the longitudinal trends in the scores of the two tests. Many researchers use NAEP data to assess the significance of state testing results (see Dee and Jacob, 2011; Nichols, 2007; and Center for Educational Policy, 2008).

FIGURE 1
4th grade NAEP scores do not show the same gains as the 4th grade CST



SOURCE: National Assessment of Educational Progress (NAEP) State Profiles, DataQuest: California Standards Test (CST) Results.

NOTE: NAEP tests reading skills and the CST measures English language arts, which includes reading and writing.

Options to protect accountability data. There are several ways the next generation of K–12 accountability programs can address this problem. First, performance indicators should be chosen or designed to resist local attempts to “game” them. In the case of state tests, for example, SBAC assessments are being designed to make it more difficult to teach to the test. Students take tests online, and the tests “adapt” based on which questions students get right or wrong. By reducing the amount teachers know about specific questions on the test, SBAC hopes to reduce their ability to teach to the test. However, given the pressure to do well on state tests, it remains uncertain if this will be sufficient to maintain the integrity of the data.

Another way to reduce incentives for teaching to the test is to use multiple indicators in the accountability measure, thereby reducing the emphasis on test results (or any one indicator). Multiple indicators also allow the state to recognize other major school outcomes or factors that reinforce the goals of the accountability program. For instance, student attendance at school is a strong candidate for including in accountability measures, as attendance reflects student work habits, it is closely linked to learning¹¹, and attendance data are reasonably reliable. Thus, attendance has all the attributes of a good performance indicator in a measure of student academic performance.

The downside to multiple measures is the threat to the clarity of the program’s goals. As discussed above, coherence is an important attribute of accountability measures. A measure that uses indicators that are not clearly linked risks sending mixed messages to educators in the field. Similarly, a measure with too many indicators risks diluting the focus of the accountability program. Imagine a measure that had 100 performance indicators! That many desired outcomes would give local educators considerable flexibility to emphasize the areas in which they were most likely to find success. In addition, the meaning of the accountability measure could be wildly inconsistent across the state.

¹¹ National Center on Educational Statistics, *Every School Day Counts: The Forum Guide to Collecting and Using Attendance Data*, available at <http://nces.ed.gov/pubs2009/attendancedata/chapter1a.asp>.

As with many areas of accountability policy, there are no clear rules for establishing the appropriate number of indicators in an accountability measure. Instead, policymakers must balance the need for multiple measures with sending clear messages about the most important educational outcomes. Because states have limited experience using data other than test scores in accountability formulas, beginning with a relatively small set of well-understood indicators provides clear signals to educators and minimizes the risk of using indicators that result in unintended consequences.

AYP can be insensitive to growth. AYP provides information only on the proportion of students that score at or above proficient. As a result, a school can boost the achievement of lower-performing students significantly, but if those students do not score at or above the proficient level, AYP will not reflect that improvement. Because of this problem, AYP creates a limited perspective on the performance of all students at a school. In addition, AYP excludes students that move during the school year, and as a result about 325,000 students, or 7 percent of all those tested in 2011, were excluded from a school calculation.¹² Excluding students from accountability measures exposes those measures to potential bias—in this case the AYP determination for schools with a significant number of mobile students may be affected.

The Obama administration’s NCLB waiver program requires states that receive waivers to include a measure of growth as well as proficiency in accountability measures. This allows lower-performing schools that advance achievement significantly to make adequate progress. Including both levels of achievement and growth in achievement is an example of balancing incentives created by the measure. The single focus on proficiency can undercut the goal of helping lower-performing students improve. Including growth along with proficiency creates incentives to address the needs of all students.

Build short- and long-term quality indicators into the process

The ultimate goal of the educational system is to graduate students who are prepared for the rigors of higher education and the workforce. Although these are important outcome measures for high schools, elementary schools need measures that show whether students are meeting grade-level expectations and are “on track” to meet the long-term goal of graduation with “college and career” skills. AYP, however, encourages teachers to focus only on current achievement—and not on whether students are successful in future grades. Because this reinforces teaching to the test, policymakers should consider a range of short- and long-term indicators of achievement in accountability measures.

Short-term data for improvement. Schools and teachers need timely information to assure that students are mastering grade-level material. Students who miss portions of grade-level material usually move on to the next grade, and the cost of helping such students recoup this material can be readily identified. For instance, districts may incur the cost directly by paying for summer school or other supplemental classes. Or, parents may bear this cost by hiring tutors to help their children make up what was missed. If deficits are not remediated, then teachers at the next grade bear the cost—in the form of lost instructional time spent addressing learning deficits. (Of course, students ultimately bear the costs of receiving a poor education.)

One problem is that state testing alerts teachers to student performance problems only *after* the school year is over, too late to give teachers the information they need to help students. However, many districts use *formative* tests—standardized assessments that are integrated into the curriculum during the school year—to provide feedback to teachers and students on what they have learned and what they have failed to grasp.

¹² Author’s calculation using 2011 API data from the California Department of Education.

With this information, teachers can adjust lessons or instructional approaches and reteach portions of lessons as necessary. The 2014–15 California budget includes funding to purchase SBAC formative tests for districts. Thus, the notion of using formative tests as timely feedback is taking hold with educators.¹³

Longer-term data strengthens accountability measures. Accountability measures also can be designed to stress long-term quality. AYP focuses only on performance in each grade. As a result, the measure does not reflect the future “costs” of teaching practices designed to generate high test scores at the expense of true learning. As a result, long-term data on student success is needed to see if students are being prepared to succeed in the future, as they move on to higher grades and more difficult material. Thus, including longer-term indicators of success will reinforce the goals of the program.

The system needs three different types of achievement data:

- **Short-term feedback on whether students have learned the current material.** These “formative data” provide the quick information to help teachers improve and help ensure that students are learning the skills and knowledge called for in the standards. Formative assessments, though, are for improvement purposes only and should not be used for accountability.
- **Annual information that measures student progress toward the long-term goals.** These data can measure the status and growth of students over the past year. With the availability of formative data, annual test results constitute a validation of sorts that students mastered the standards.
- **Data on the longer-term success of students in staying on track for meeting state expectations.** This reinforces the program’s goals and creates incentives for schools (such as elementary and middle schools, or middle and high schools) to work closely together to ensure instruction is aligned and students are learning the prerequisite skills for success in the long run.

Use accountability programs to build capacity

NCLB relies on a variety of sanctions to motivate educators. Most of these sanctions penalize schools for performance issues. However, most educators do not need threats to spur improvement. By redefining the “theory of action” for accountability programs, the state can use accountability programs to promote a system that builds competence and stresses continual improvement. There is still a role for sanctions, but it should be limited to school boards that refuse to take the steps necessary to remedy deficiencies in their educational programs.

Holding districts responsible. For an accountability program to be successful, it must hold the right actors accountable. In California, school districts have the greatest ability and authority to effect change. Districts determine funding, programs, and policies for local schools. Thus, holding a school responsible for student outcomes when it controls very little of the inputs violates a basic principle of effective accountability.

NCLB partly violates this accountability principle by placing schools at the forefront of accountability consequences and creating rigid, punitive penalties for failure. Though school employees deliver education to students, districts make most decisions that shape school operations. School boards and top

¹³ According to the Great Schools Partnership, “While the formative-assessment concept has only existed since the 1960s, educators have arguably been using ‘formative assessments’ in various forms since the invention of teaching. As an intentional school-improvement strategy, however, formative assessment has received growing attention from educators and researchers in recent decades. In fact, it is now widely considered to be one of the more effective instructional strategies used by teachers, and there is a growing body of literature and academic research on the topic. Schools are now more likely to encourage or require teachers to use formative-assessment strategies in the classroom .” (See <http://edglossary.org/formative-assessment/>.)

administrators allocate funds, set hiring policies and practices, adopt instructional materials, and provide leadership that sets expectations. In California, evaluations of school interventions found that districts play a major role in creating an environment in which schools can succeed (Parrish et al., 2005, Westover et al., 2012). Westover and colleagues found superintendent turnover; contentious relationships among board members, administrators, or union officials; or an entrenched culture of low expectations reduces district support for achievement. Thus, districts that are unable to generate the consensus for improvement at the district level are also unable to help schools make progress. Similarly, Parrish and colleagues recommended that, because of its critical role, the state should hold districts accountable for helping their low-performing schools improve.

Under NCLB, schools are the first entities to feel the consequences of failing to make AYP. Although both districts and schools are accountable, low-performing schools usually miss making AYP well before districts. As a consequence, a “failing” school could be well into the more significant sanctions by the time its district begins program improvement. In districts that do not adequately support the improvement process, schools face an uphill battle to raise student achievement.

There are precedents for holding districts responsible for inadequate performance. Massachusetts uses a concept similar to bankruptcy as a last-resort sanction, after capacity-building activities have failed. The state administers a multistep process of support and intervention, using regional teams to work with challenged districts to develop better systems and services for improving low-performing schools. If this help fails, the state may put schools into “receivership,” assuming responsibility from district boards for the operation and support of very low-achieving schools. In extreme cases, the state places chronically low-performing districts in receivership and appoints an executive to run the district (Massachusetts Department of Elementary and Secondary Education, 2014).

In California, boards are held accountable for severe fiscal problems. California law eliminates the powers of boards when school districts receive significant emergency loans from the state.¹⁴ This action reflects the idea that the board failed in its fiduciary responsibilities. Creating similar accountability consequences for school boards that fail to address educational problems would better align local responsibility and authority. As discussed above, however, school board sanctions would occur only if the state found it unwilling to take the steps needed to improve district quality.

Accountability as a process for system improvement. NCLB’s automatic sanctions also unnecessarily put a punitive spin on accountability. The sanctions approach undercuts the trust employees need to try new things and to take risks that may have long-term payoffs. Rather than blame employees, administrators need to determine what exactly is not working at a school and take steps to improve the situation. NCLB sanctions seem to do the opposite, encouraging parents to move their children to other schools and directing districts to contract out to provide supplemental tutoring.

Research also suggests that educators in lower-performing schools and districts often lack an understanding of how to build successful educational programs. A recent study of NCLB “program improvement” looked at schools in a large urban district that were under pressure to improve student achievement (Finnigan and Daly, 2012). The study found a variety of problems that limit improvement efforts:

¹⁴ See Education Code section 41326.

- **A failure to focus on the causes of low performance.** Teachers and principals looked for ways to increase test scores rather than try to understand the problems students face in the classroom.
- **A lack of knowledge about changes that might improve instruction.** School-improvement teams tended to recycle ideas that had been tried in previous improvement efforts.
- **A lack of trust needed to take actions that could lead to better outcomes.** As in Deming’s principles, researchers recognized the key role trust plays in improving schools and its lack in low-performing schools.
- **District assignment practices that resulted in significant teacher and principal turnover.** District practices that result in significant school turnover complicate the effort to get buy in for an improvement plan that “stays the course” over time.

Teachers and principals do not always have the organizational savvy needed for significant school improvement. Though the study did not examine in detail the district role in these activities, it seems evident that the central office also does not understand how to support these schools, or does not have sufficient incentive to provide the necessary support.

These findings suggest that technical assistance should be the primary consequence of low performance. There is also evidence that such directed capacity building can make a difference in school districts. For instance, California operates a support program for districts that have been in NCLB program improvement for at least three years. The state establishes independent District Assistance and Intervention Teams (DAIT) to assess district programs for supporting schools and help craft and monitor plans to address district weaknesses.

An evaluation of the DAIT program found that teams helped low-performing districts build capacity in a number of areas, leading to gains in achievement and smaller gaps between student subgroups (Westover et al., 2012). The report found that student achievement in districts given *intensive* DAIT services increased by 20 percent of a standard deviation after three years, a large increase. Activities that made the most difference include improving the use of data to inform instruction, making professional development more effective, and getting districts to promote high expectations and within-district accountability.

The idea that schools should be required to take special steps to improve seems like a reasonable consequence from low scores on an accountability measure. Harsher sanctions, though, suggest an unwillingness on the part of the K–12 system to try to improve. While this may be a problem in a few places, issues of capacity and conditions for effective reform appear to be the larger problems. Eliminating school-level sanctions and substituting various levels of capacity-building assistance is a strategy that has several benefits. First, it directly addresses barriers to improvements in most schools. Second, a focus on capacity building reduces the stigma of accountability on teachers, and it provides the strongest avenue to making accountability a positive learning opportunity for the many teachers who want students to succeed. Third, by reducing teacher stress, revising consequences can lessen the incentive to “teach to the test.” This in turn could reduce threats to the validity of accountability data and the corruption of the accountability process.

Conclusion

Our analysis suggests that significant changes are needed to ensure that the next generation of accountability programs works as desired. Not surprisingly, the indicators of performance matter—a lot! The best use of data is to improve quality and efficiency, and not to assign blame or credit. As NCLB has shown, using test results as the primary data source for accountability can distort the process it measures and corrupt the very

data that is designed to help administrators, parents, and communities gauge the effectiveness of their schools as well as provide an independent assessment of each student's academic status.

Districts are the key accountability points, since they control most variables that affect school performance. The DAIT program demonstrates that, with the right combination of help and incentives, in most cases school leaders can build the capacity of low-performing districts to raise student achievement. Districts should be held accountable for low-performing schools, but sanctions should be limited to those districts that are unable to build the consensus necessary to improve student achievement. The Massachusetts model suggests a way to deal with these districts.

The DAIT program also suggests the need for an ongoing program of pressure and assistance designed to push districts to improve. It does not make sense to wait for districts to qualify for DAIT—the state should have a process that helps strengthen the focus on quality in every district, helping lower performers improve and encouraging midlevel performers to move up. As mentioned earlier, state law empowers county superintendents of schools to monitor and intervene in districts with multiyear financial problems. By inserting county review and approvals into the annual budget process, the state created stronger incentives for districts to take the sometimes-difficult steps needed to maintain fiscal health. The new LCAP process could easily be adapted to resemble the local budget-review process in its system of support and incentives that would provide a modest level of ongoing pressure to improve.

The idea behind these changes is to move away from blame as a motivating factor and to focus accountability programs on creating an environment and a system of information and training that helps educators become more successful. It asks administrators and teachers to get better at their craft. For example, districts need high-quality employee-evaluation programs to help teachers improve and to encourage those who cannot meet minimum standards to seek other work. Furthermore, organizational improvement does not exclude other strategies. Charter schools remain a viable option for expanding educational opportunities for students or evaluating innovations. By increasing the quality of administration and teaching, we think accountability can be a potent long-term component of the state's K–12 improvement strategy.

California's Accountability Programs

This section analyzes the three California accountability programs and measures. Each relies on a different set of indicators and different combinations of growth and performance levels. Each measure also has shortcomings. The API was designed for conditions that existed in 1999, and new methodologies exist that are more useful and accurate. LCAP and SQII use performance data that may not be valid and reliable. In addition, the two measures use multiple performance indicators, which raises the issue of whether the two programs create a coherent definition of performance. Each program, though, also has important strengths that are useful to understand when considering the next generation of accountability measures. We begin with a discussion of the state API. Then we turn to the two new measures, LCAP and SQII.

Public School Accountability Act

The Academic Performance Index (API) has been the primary state measure of school and district performance since 1999. Operating as part of the Public Schools Accountability Act (PSAA), it uses test scores from students in grades 2–11 and the high school exit examination to calculate the growth in student achievement in schools and districts. It includes scores from English, mathematics, science, and history/social science.¹⁵ PSAA called on schools to make “growth,” which is defined as shrinking the gap between an API score of 800 and the schools’ prior-year score. Schools with an API of 800 or higher were required to make a one-point gain to make “growth.” Table 3 shows the API methodology, which creates a weighted average of student test scores. The formula uses “progressive” weights: that is, schools gain more points for improving the achievement of lower-performers than students working at higher levels.¹⁶

TABLE 3
API methodology

School test results			
	Number of students	Weights	Points
Advanced	120	1000	120,000
Proficient	170	875	148,750
Basic	200	700	140,000
Below Basic	110	500	55,000
Far Below Basic	45	200	9,000
Subtotal	645		472,750
API			733

$$\frac{472,750}{645} = 733$$

SOURCE: California Department of Education 2012–13 Academic Performance Index Reports: Information Guide available at www.cde.ca.gov/ta/ac/ap/documents/infoguide13.pdf.

NOTE: The API blends the results of several tests into its average. The different subject area test results are weighted, with English getting the largest weight.

Theory of action. The API promoted “continual improvement” in K–12 education. Because CDE was unsure how quickly schools could be expected to improve, it created a system where *growth* was the annual goal.

¹⁵ The API is slated to change in the future, however, as state law requires adding other school outcomes, such as high school graduation, to the API by 2016.

¹⁶ Students who improve from Proficient to Advanced add 125 points to the API. Students moving from Far Below Basic to Below Basic add 300.

An API of 800 was established as the long-term goal, although that was somewhat arbitrary. All schools were encouraged to grow each year, even high-performing schools. PSAA also required schools and districts to make similar growth for its major ethnic/racial subgroups.

PSAA did not provide additional funds for improvement or impose consequences on schools that did not improve sufficiently. Instead, the state created a voluntary program that supplied additional funds in exchange for the possibility of consequences in the future. This program, known as the Immediate Intervention in Underperforming Schools Program (IIUSP), assisted schools with below-average API scores that failed to grow in the previous year. IIUSP did not result in consistent gains in participating schools (Parrish et al., 2005). This was followed by the High Priority Schools Grant Program, which provided additional funding to low-performing schools in California. But this program also found that participating schools did not improve performance faster than comparison schools (Harr et al., 2007).

PSAA-API emphasizes continual growth

The API provided a methodology that measured growth in school performance at a time when the state's testing program was under development and it had no data system for following student progress over time. The API is a better measure than the "percent proficient" calculation of AYP because it captures the achievement of almost all students in a single performance rating. In addition, its progressive weighting formula affords all schools a similar chance to make adequate progress each year and creates incentives for schools to focus on the needs of low-achieving students.

The push for continual improvement also is a positive feature. All schools, both low and high performers, were expected to make progress each year. Unfortunately, the long-term impact of the API and its continual improvement ethos is unknown, as the program was displaced by NCLB after operating for only a few years. However, PSAA was never eliminated from the Education Code, and CDE continued to publish API scores until 2013, when state tests were suspended.

Program has several weaknesses

Like NCLB, PSAA has several problems that have become apparent with the passage of time. While the API appears simple, it actually is quite complex, and its problems are a reflection of this complexity. In addition, though the state continues to use the DAIT model as part of NCLB sanctions, the state never embraced technical assistance as an important state-level strategy for improving the K–12 system or created district-level consequences when officials could not or did not take steps needed to support school improvement.

Validity issues. The API is not based directly on the growth of *individual* students (even though longitudinal data are available). Instead, it compares the scores of all students at a school from one year to the next. To make the comparison as valid as possible, the API excludes students who leave or arrive at a school during the school year. Students who move during the summer, however, are counted. As a result, the API formula contrasts the scores of somewhat different groups of students. In addition, like AYP, the API also excludes students who move during the school year. Systematically excluding certain types of students opens the door to error in the API.

The design of the API makes it challenging to interpret, potentially creating misleading signals about quality. API growth measures whether schools got *better* at educating students compared to the prior year. But this is not the same as measuring the relative effectiveness of schools. Research has shown that API growth is inconsistent with gains as measured by a value-added analysis (Fagioli, 2014; and Glazerman, 2011).

Value-added scores generally evaluate student growth at a school compared to the growth of similar students at other schools.

Measuring whether schools became more effective in the past is useful, but it can mischaracterize school performance in certain instances. For example, a school that raised student achievement substantially each year might show no API growth because its students were growing at roughly the same rate each year. That is, the school was very effective (or added significant “value”) but did not get better from one year to the next.

The opposite also can occur: the API can show growth when new students arrive at a school better prepared than the prior year. Because the measure assumes that this year’s students are exactly the same as last year’s, the API interprets the higher scores as growth in school effectiveness, even though it really reflects better preparation in the previous school. Changing student demographics at a school can have the same effect. Thus, the measure can show growth even when a school did not truly improve.¹⁷ Though value-added data is a more useful indicator of school performance for accountability purposes, the API embodies what the state could create in 1999, given the lack of tests and a data system that supported a better measure.

Transparency issues. Because the API is deceptively complex, the measure is not well understood by nontechnical users. There are several important ways that the API can mislead:

- **Progressive weights make growth comparisons invalid.** The progressive weights are an important policy feature of the API, but it also makes simple API growth comparisons invalid. Because schools earn more points by helping lower-achieving students improve, schools at the lower end of the API scale have more opportunity to grow than schools at the upper end of the scale. As a result of this feature, though, comparing the point growth of schools with different API scores is not valid. This makes the API less useful for comparing the annual performance of schools.
- **Changes in the calculation of the API affect school and district rankings.** When changes occur, CDE ensures that the *state average* API does not change. The API for individual schools, however, can change significantly. As a consequence, school APIs go up or down over time simply due to changes in the composition of the API—and not because of changing student achievement.
- **Comparisons of growth over multiple years are not possible.** CDE’s website makes it clear that API data should be compared only by subtracting the base API of one year from the growth API of the subsequent year. But because the makeup of the API and demographics of the student population change over time, the meaning of the API also changes over time. As a result, the API cannot be used to track individual school performance over many years.
- **API targets suggest high schools underperform.** Average elementary school scores were 811 in 2013 while high school APIs averaged 757. This difference, however, is an artifact of setting the API target for all schools at 800, but failing to ensure that the tests are of equal difficulty across grades. As a result, achieving a score of 800 is easier in elementary and middle schools than in high schools.

No assistance plan for districts. The initial design of PSAA included the IIUSP. When that program failed to help schools improve, the state tried a second school-focused assistance program. That program also failed to show positive results, and the state then turned to DAIT. The district program got better results. Unfortunately, the state never capitalized on these findings. The supply of funding for technical assistance (provided through county offices) remained at \$10 million a year, and the state never created district-level consequences to help spur school boards to address difficult local problems that undermine school improvement.

¹⁷ This problem makes interpreting middle and high school APIs more difficult. If elementary scores are increasing, a middle school API could increase even though the school did not add more value.

API has outlived its usefulness

The API has many attractive features. Its emphasis on growth gives all schools a chance to succeed in the accountability system. The measure's progressive weighting ensures that schools must address the needs of lower-performing students. PSAA also encourages continual growth—even highly rated schools are required to show some growth each year. The state's assistance and intervention processes also were developed under PSAA, and they are now in use as part of NCLB.

Though the API was a useful measure in 1999, other measures are now available that provide a more accurate estimate of the gains made by students and the contributions of schools. The API was designed to do one thing: measure increases in annual school growth. It cannot measure value added by schools or shed much light on longer-term performance trends, such as the relative growth of schools or districts, or the relative progress of elementary and high schools. The measure also is easily misunderstood. Though the API is very complex, the state has taken pains to make it seem simple by creating the appearance of consistency across time and across different types of schools. Because of this misleading simplicity, the measure can easily be misused or misinterpreted.

State law requires CDE to add new indicators into the API, such as graduation rates, attendance, and preparation for college and employment. The new SBAC tests also will require changes to the API. Thus, given the fundamental shortcomings of the API, this is a good time to reconsider its future. The state needs a good school-performance metric—educators, policymakers, and the public need credible information on the progress of the system. The API served California education well for many years, but the state can now do better.

Local Control Funding Formula

The Local Control and Accountability Plan (LCAP) was created in 2012 by state legislation establishing the Local Control Funding Formula (LCFF). The LCFF revamps the state's K–12 finance system, merging a large number of state categorical programs into two funding streams.¹⁸ As districts implement the new finance system, LCAP is designed to help districts connect funding decisions to school-performance issues and provide a level of accountability for student success. LCAP employs 19 indicators of school or district performance in eight “priority” areas, and it requires districts to make progress on at least a few each year. Unlike the other accountability programs, LCAP does not include a formula that summarizes a school's performance. As a result, all data elements are equally important. Table 4 displays the eight priority areas and the specific performance indicators.

¹⁸ A base grant that is provided for all students and an additional amount that is made available to meet the needs of low-income, English learner, and foster care students.

TABLE 4
LCAP performance indicators

<p>Academic achievement</p> <ul style="list-style-type: none"> -Standardized test scores -Percentage of students that score a 3 or higher on Advanced Placement exams -Percentage of students determined ready for college by the Early Assessment Program -API scores -English Learner (EL) reclassification rate -Percentage of EL students that become proficient -Percentage of students earning “A-G” university course requirements <p>Basic educational inputs</p> <ul style="list-style-type: none"> -Credentialed teacher -Adequate supply of standards-aligned textbooks -Well-maintained facilities <p>Parental involvement</p> <ul style="list-style-type: none"> -Efforts to involve parents 	<p>Student engagement</p> <ul style="list-style-type: none"> -Attendance -Chronic absenteeism -Middle school and high school dropout rates -High school graduation rates <p>School climate</p> <ul style="list-style-type: none"> -Suspension and expulsion rates -Parent, pupil, and teacher surveys <p>Implementation of Common Core State Standards</p> <ul style="list-style-type: none"> -Implementation of standards, particularly for English learner students <p>Course access</p> <ul style="list-style-type: none"> -Access to and enrollment in required areas of study <p>Other student outcomes</p> <ul style="list-style-type: none"> -Outcome data on the required areas of study
---	--

SOURCE: California Education Code section 52060.

NOTE: LCAP = Local Control Accountability Plan.

Theory of action. LCAP promotes local accountability, stressing that parents and communities can positively influence decisions about how best to use funds to improve schools. Central to LCAP is parent and community input on district priorities. LCAP also puts a spotlight on district priorities and specific goals for student success in the eight performance areas. It also uses county offices of education to ensure that districts are complying with the requirements of the law and for providing technical assistance to struggling districts. Only the state superintendent of public instruction can impose consequences, using the law’s limited authority to intervene in “failing” districts. The new law also provided a one-time \$10 million allocation for technical assistance. However, this assistance is not yet available, as the process for allocating these funds remains a work in progress.

An emphasis on “local”

The strength of LCAP is that it places a spotlight on school and district performance as budget decisions are made. In addition, the law attempts to give parents and communities additional opportunities to shape local education and expenditure plans. District budgets are complex, making participation of nonexperts difficult. Time will tell whether these provisions to strengthen the community voice will have long-term impacts on school and district governance.

LCAP also creates a new *local* emphasis on accountability. The program stresses the importance of districts in setting priorities and developing plans for improvement. In addition, the broad range of outcome indicators gives districts great flexibility to shape local priorities. The participation of county offices of education could be a particularly strong program element, as it introduces an annual review of district problems, programs, and expenditures. Until now, most districts have operated without any serious external assessment of the effectiveness of their academic programs. In addition, county offices may provide technical assistance as the primary strategy for improving district performance.

LCAP is weak on accountability

Unfortunately, there are also serious problems with LCAP. In fact, LCAP fails to conform to most of the basic design elements of accountability programs. The 19 indicators are not well defined in statute. Plus, there are too many goals, and no hierarchy of importance among the goals, which makes it difficult for districts to focus their efforts. Finally, with so many performance indicators, locally established goals, and no clear priorities among them, there is no way to determine whether districts make sufficient progress. We discuss these issues in more detail below.

Data issues. The LCAP contains several important data issues that could seriously undermine the program if they are not addressed:

- **Priorities that are not clearly defined.** One example is the implementation of the Common Core State Standards (CCSS). These standards guide curriculum design in mathematics and English beginning in 2014–15. State law does not indicate how implementation will be measured, and therefore it is not apparent how districts would be held accountable for this task.
- **Indicators that have no statutory or regulatory link.** Terms such as “chronic absenteeism” and “pupil suspension” are not linked to an Education Code definition. As a result, districts appear free to determine their meaning. Lack of consistent definitions could undermine the validity of LCAP.
- **Indicators that are influenced by local practices or administrative decisions.** This problem makes data vulnerable to actions that distort the data. For instance, local standards for English Learner (EL) reclassification vary significantly (Hill, 2014). This means that schools and districts looking to increase the reclassification rate under LCAP could simply relax EL standards—with unknown consequences for students.
- **Indicators that could create unintended consequences.** Like our example of EL standards above, LCAP creates incentives that push educators to do things that are not necessarily in the best interests of students. A review of California’s alternative programs, for instance, found that some districts use alternative schools to serve students that are disruptive or behind in their studies (Warren, 2007). By including suspensions and expulsions as one LCAP indicator, the program could increase the incentive to transfer students as a way of “improving” school outcomes.

Validity issues. As discussed earlier, unreliable data leads to questions of validity. If indicators are defined inconsistently or if administrative actions can alter school data, then the measure of performance will not be valid. LCAP is vulnerable to this problem.

LCAP also is threatened by poor-construct validity. This means that not all of its indicators are consistent with the objective of a broad definition of student success. One state priority calls for adequate facilities and a credentialed teacher in each classroom.¹⁹ Research, though, has yet to document that schools cannot provide a good education without these inputs. In fact, there is considerable research showing that credentials have little or no impact on teacher quality (Walsh, 2001; Aaronson et al., 2003; Kaine et al., 2008).

The nonacademic outcomes in LCAP also are potentially problematic. Although research generally links these types of outcomes—such as student expulsions—to academic success, the evidence of a direct nexus between some outcomes and higher achievement is thin. For instance, reducing expulsions by itself may have little impact on the underlying problems that make the school’s climate less-conducive to learning. In

¹⁹ The budget includes \$10 million in federal technical assistance funds for schools and districts that have not made AYP in at least three years and are in “program improvement.”

¹⁹ Adequacy of these inputs was the subject of the *Williams* lawsuit and a subsequent program that resulted from settlement negotiations.

addition, if expelled students are given in-school detention (where students are not allowed to attend class) in lieu of expelling or suspending them, the policy would make the school's data look better but do little to improve the school culture or reduce the amount of lost class time experienced by students. There are only a few examples in which states have used these nonacademic indicators for accountability purposes (Schwartz, 2011). Thus, there is little information about the potential for unintended consequences associated with these indicators.

Absence of a central goal or clear priorities among its goals makes it hard to understand what the state is trying to accomplish with LCAP. In fact, as noted above, the program gives districts considerable latitude to shape the program to meet local priorities. But this situation means that "success" can be defined many different ways. If LCAP is designed as a supplement to state and federal accountability, then allowing local variation makes sense. But if the program is intended to substitute for a statewide accountability measure, LCAP sacrifices comparability in its measures of performance. Without a consistent measure of performance, LCAP creates a very weak accountability tool.

Requiring districts and schools to monitor eight state priorities and up to 19 performance indicators also makes it difficult for schools and districts to focus the improvement process. Currently, districts are in the midst of implementing the CCSS in English and mathematics. The multiple goals in LCAP encourage districts to address school climate and safety, high school graduation and college preparation, parental and community input, and facilities—all at once. And though these are all important school outcomes, addressing them simultaneously risks diluting the system's capacity for effective change.

Transparency issues. The design of LCAP and its data problems lead to transparency problems. Clearly, the lack of clear definitions and the ability to influence school outcomes through administrative policies create a potential barrier for parents and the public to gauge school or district performance. LCAP also will give communities a huge amount of data in a way that is very difficult to digest; the first-year experience with LCAPs show districts using one or two indicators to measure each state priority. As a result, a school with, say, five subgroups would list 40-50 different data elements in its LCAP. Providing so much data without any comparative information on similar districts seems likely to overwhelm most parents and community members.

County office reviews. Limits on the county office role undercut the impact of local reviews on district quality. The program defines the county office review process as a rather ministerial function. County duties are limited to determining whether district plans meet the requirements of state law: for example, does the plan conform to the state-approved template, and is the budget sufficient to implement the plan? County offices also are required to offer technical assistance to districts when plans are not approved (California County Superintendents Educational Services Association, 2014).

LCAP does not grant county offices clear powers to evaluate the quality of a district plan or reject a plan because it fails to address significant areas of district weakness. The lack of broader authority potentially reduces the value of the local review process. LCAP allows county offices that want to avoid controversy the ability to maintain a passive role as a plan checker, rather than as a quality checker. Alternatively, a county office that takes seriously its role in promoting quality could find districts ignoring its advice.

LCAP needs attention

The accountability features of LCAP resemble more a data report than an accountability program. And even as a data report it has major problems. Performance indicators are inadequately defined, and several could

lead to unintended consequences. The eight state priorities also fail to create a coherent set of objectives for schools, as evidence is lacking of a direct connection for several LCAP indicators. Finally, there are far too many indicators to draw conclusions about the “bottom line” implications of the data for schools. Thus, in its current form, LCAP may create more confusion than clarity.

Early reports on LCAPs submitted by district in July 2014 also indicate that reviewers are finding the plans difficult to understand. Consistency across districts appears to be a problem. Much of the concern centers on the budget section in LCAP, in which districts describe the cost of activities that are proposed in the plan. Specifically, LCAPs do not provide a clear picture of how funds will be used to promote the achievement of students, especially of low-income, English Learner, and foster care students (Hahnel, 2014).

With some changes, though, LCAP could complement the state accountability program. LCAP has several strong elements. First, LCAP recognizes that accountability works best when the driving force for improvement is local and the focus is on districts. In conjunction with a state measure that creates clear priorities for student achievement, the wider range of outcome indicators and greater involvement in planning could strengthen local accountability. If its data problems are resolved, LCAP could highlight student and school outcomes in more detail than state or federal accountability programs. In a sense, LCAP data allow communities to “look under the hood” of districts to develop a more-detailed sense of where district performance needs to improve. This could also allow LCAP to spur the development and validation of new indicators in areas of special interest and concern, such as non-cognitive skills.

The review authority given to the county superintendent of schools could transform that position into a “critical friend” that pushes for continual district improvement. Like the fiscal-accountability process created by AB 1200, LCAP could empower the county superintendent to create a process of continual improvement, working with districts each year to address weaknesses in local programs. With these changes, LCAP would complement the state’s formal accountability system with a local system of data, assistance, and oversight.

The CORE NCLB waiver

In August 2013, the Obama administration approved a new federal accountability program for eight California school districts under a waiver of NCLB. The eight districts work together through the California Office to Reform Education (CORE) and include K–12 districts in Los Angeles, San Francisco, Oakland, Long Beach, Sacramento, Fresno, Santa Ana, and Sanger.²⁰

CORE’s accountability measure—called the School Quality Improvement Index (SQII)—combines the results of state tests with a range of student behavior and other school-outcome data. In total, SQII contains nine performance indicators. Table 5 illustrates the basic structure of the CORE measure for high schools. State tests will compose 40 percent of a school’s score (growth and levels of achievement), and high school success indicators account for 20 percent. Nonacademic factors account for the other 40 percent, including socioeconomic factors and indicators of school culture and climate.

²⁰ Clovis Unified and Garden Grove Unified school districts are members of CORE but not members of the group that received the federal waiver. Sacramento Unified did receive the first-year waiver but did not reapply for a second year.

TABLE 5
CORE performance indicators

Academic achievement (40%)

- Percent proficient
- Annual growth in achievement

Social and emotional factors (20%)

- Chronic absenteeism
- Suspension and expulsion rates
- Non-cognitive skills (Unspecified)

High school success (20%)

- Graduation rates
- Early high school persistence rates

School culture and climate (20%)

- Student, staff, parent surveys
 - Special education identification rates
 - English Learner reclassification rates
-

SOURCE: CORE Waiver Application (ESEA Flexibility: Request for Window 3, May 24, 2013).

NOTE: CORE = California Office to Reform Education.

The plan indicates that the SQII will operate somewhat like the API. Schools will be held accountable for reaching a status target (a score of 90 out of 100) or making annual growth. Growth targets call for schools to improve at least two points in two years and four points in four years. The SQII replaces AYP as the performance measure under NCLB for the eight districts. All other districts in California remain subject to AYP. At the time this analysis was written, CORE was in the process of refining its accountability measure, and the group's final accountability measure may differ somewhat from its initial proposal.

Theory of action. The NCLB waiver program does not completely revise the program's theory of action, although it attempts to soften the act's most problematic features. The waiver stresses growth in achievement as well as proficiency, and the requirement to achieve 100 percent proficiency is waived. This allows low-performing schools to show progress and reduces the number of "failing" schools. Many sanctions are eliminated, easing restrictions on the use of funds. In addition, mandated school interventions are targeted only to the 5 percent of schools with chronic underperformance and to an additional 10 percent of schools that struggle with particular subgroups. CORE assists these schools by pairing them with similar schools that have found success. The paired school provides technical assistance and guidance on the improvement process. With these changes, the theory of action underlying the waiver program looks more like California's original "continual improvement" program, with the addition of technical assistance for a relatively small number of low-performing schools.

By broadening the range of indicators, CORE makes the case that these factors are so important to the success of students that schools should be held accountable for them. The CORE districts' accountability measure includes both the growth and level of achievement on state tests and adds a variety of other indicators, most of which are linked to achievement. These include indicators of social and emotional well-being and school climate. Thus, the design of the SQII suggests that preparing students to be college- and career-ready is dependent on improvement in both the academic and nonacademic areas.

SQII has innovative features

The strengths of the measure lie in the range of indicators that are merged into the index. On academics, the measure recognizes the complementary perspectives that growth and status provide. High school graduation and persistence are given a high priority in SQII. On the nonacademic side, SQII incorporates a broader range of school indicators than are contained in LCAP, and it avoids most of LCAP's most significant data issues (discussed further below).

In addition, SQII merges the nine indicators into a single measure, setting achievement as the highest priority and clearly identifying the priority of the other outcomes. Though complex, SQII provides pretty clear signals about the goals of CORE in designing the measure. Academics and graduation rates receive the bulk of the weight in the formula, totaling about 60 percent. This leaves the nonacademic measures at 40 percent of the weight (and perhaps higher in elementary schools, where there is no graduation or persistence measure). As a result, no single factor will have a dominant impact on SQII.

Two additional features distinguish SQII from the other accountability measures. First, SQII includes data on the persistence of students from grades 8–10 as a middle school indicator. That is, it holds middle schools accountable for the success of its students in early high school. This measurement serves as long-term indicator of quality that reinforces the program's goals for students. In addition, the indicator also creates incentives for middle school educators to work closely with its high school counterparts to improve the preparation of students for high school.

Second, SQII incorporates subgroup data quite differently than other accountability programs. Specifically, the measure builds subgroup performance directly into the performance index rather than calculating separate scores for each subgroup. By using this compensatory scheme, high scores from one group can offset lower scores from another group. As a result, a low subgroup score does not prevent a school from making adequate growth so long as the overall average of the school and subgroup scores indicates adequate improvement. CORE also set the minimum size of significant subgroups at 20 students (compared to 100 students under California's current federal accountability workbook rules). This change, which increases the number of subgroups at many schools, strengthens the subgroup protections in the CORE districts.

SQII raises data and validity questions

SQII has several issues similar to those discussed above with LCAP—potential data and validity issues. The complexity of SQII also raises questions about coherence and transparency.

Data issues. Data used in SQII may have similar issues as LCAP. Like LCAP, several indicators in SQII are affected by administrative actions. The CORE plan recognizes the potential for inappropriate incentives from including EL reclassification and special education identification rates, and it proposes ways to mitigate the problem. However, other potential problem areas, such as suspensions and expulsions, are not addressed. CORE staff advise that it is working to define these terms and plans to implement uniform definitions when it begins analyzing data in 2014–15.²¹ In addition, similar to API, many CORE indicators are affected by changes in the underlying demographics of the student population that occur each year.

Validity issues. Assuming data issues are addressed, construct validity is an issue with SQII. The question is, do the various indicators work together to measure a coherent objective? Based on its first-year waiver, it is not

²¹ Phone interview with Noah Bookman of CORE, February 27, 2014.

evident how the socio-emotional factors and the relationship of the percentage of special education students connect to school performance. Like LCAP, SQII also includes suspension and expulsion rates as indicators. And as with LCAP, including these data in SQII suggests that simply by limiting suspensions and expulsions—with no other changes—schools will generate better student outcomes. To its credit, the CORE plan acknowledges that research has not documented a direct link between the rate of school disciplinary actions and achievement. Including these data, however, creates the potential for schools to show growth on SQII by taking actions that have little effect on student achievement.

Transparency issues. If SQII is developed as planned, it will have some of the positive communication attributes of the API. The SQII formula will calculate a score for each school, and adequate progress will be defined as reaching a certain score or growth target. The school score will make it easier for educators and parents to grasp a school’s bottom-line performance. The measure, though, is extremely complex, which will impede understanding of what growth means regarding school performance. But as we saw with LCAP, too much undigested information also can create communication problems. SQII represents a different communications challenge: how to explain the meaning of “growth” in a multidimensional index of outcomes.

Evaluate the CORE waiver program

We are wary of drawing too many definitive conclusions about SQII, given that it is a work in progress. In general, however, SQII represents the type of accountability measure most states have adopted under the NCLB waiver program, although, with its nine performance indicators, it is more complex than most. Yet SQII has been developed in a thoughtful way. For that reason, SQII should garner consideration as California’s accountability measure when the federal act is reauthorized. Although more development is needed to fully flesh out the measure, SQII marks an interesting new direction. Thus CORE and CDE should use this opportunity to evaluate the impact of the waiver program. Tracking the impact of nonacademic influences on schools and students would protect the interests of students in the CORE districts and help decision-makers understand how these indicators function in an accountability context. There are also other important issues to monitor, including the impact of SQII on achievement and the effect of its subgroup methodology on protecting the interests of important groups of students.

Where to Go From Here

A great deal has been learned about the strengths and weaknesses of K–12 accountability programs over the past 15 years. From our analysis of NCLB, we concluded that multiple measures of student performance can strengthen local incentives for good instructional practices and reinforce the state’s long-term goals for students. We also learned that bypassing districts in the accountability process ignores the critical roles and responsibilities of school boards and district staff in the school-improvement process. And by focusing sanctions on school boards, the state can free teachers to use accountability data to improve curriculum and instruction, as well as hold school boards accountable for building school and district capacity.

Our review of the California programs illustrates the urgent need to simplify and clarify the state’s accountability programs. The complexity of having four different programs operating simultaneously is daunting, and the message it sends to educators and school boards is confusing. Unless the state obtains a waiver from NCLB, or the act is reauthorized, there is little the governor or legislature can do to directly address the problems created by NCLB and the school-level sanctions it employs. However, the state can address problems with API and LCAP in a way that prepares for the reauthorization of NCLB in the future.

California’s programs also illustrate the challenges of designing measures that use multiple performance indicators. Though policymakers seek to use accountability to address a wide range of outcomes, policy interests in LCAP outstrip the availability of valid and reliable data. Much of the data are cross-sectional, which cannot adjust for changing student demographics and student mobility. For that reason, longitudinal student-level data are needed. Multiple indicators also make accountability programs significantly more complex. Such complexity makes communication with parents and the public difficult.

Our analysis points to several short- and long-term actions the state should take to improve its accountability system. First, the state Board of Education and CDE should work with the CORE districts to evaluate the effectiveness of SQII and its plan to provide technical assistance by pairing schools. The evaluation could begin by gauging the impact of SQII on school planning and assessing the effect of school pairing on creating an environment in the low-performing schools where student achievement and instructional improvement become the focus. As testing resumes in California, the state could then evaluate the effect of CORE’s waiver program on student achievement compared to districts not in CORE.

Second, LCAP’s shortcomings should be addressed by statute. The state should make clear whether the program is intended as an alternative or complement to existing state and federal accountability programs. As our analysis suggests, we think the program would work well in conjunction with the state accountability program. Consequently, our suggestions would move the program in that direction, by defining performance indicators, deleting those that create problematic incentives, identifying priority outcomes that will be included in a new state measure, collecting and posting district data on the state’s K–12 website, and clarifying the roles of the county offices.

In the longer run, the state should develop an alternative to the API that clearly expresses the state’s goals for student achievement. Along with a new measure, the state should strengthen its program to encourage school boards and district staff to manage schools more effectively. Technical assistance would be the primary vehicle. School board sanctions, similar to those in AB 1200, would also be part of the package. Finally, county offices would be empowered to work with districts to critique local plans, identify weaknesses in local programs, and help districts find solutions to problems of low achievement.

In designing a new state accountability program, the state has many options to choose from. There are no “right” or “wrong” options, only choices that reflect different trade-offs and judgments about how best to create positive local incentives. The option we developed reflects the lessons of existing accountability programs and builds on LCAP to create an integrated state and local accountability system. We start by identifying the performance indicators that are valid and reliable and that measure the capacity of schools and districts to organize and help students achieve at the highest possible levels. We then describe how these indicators are used to evaluate performance in each level of the K–12 system: elementary, middle, and high school.

Indicators of performance

Our approach to developing a new state accountability measure recognizes the need for a relatively simple program and data that is valid and reliable and that provides different perspectives on school performance. LCAP also requires good data, but because LCAP would no longer trigger consequences the criteria for LCAP indicators may be somewhat less strict. Figure 6 illustrates our division of outcome indicators for use in a new state measure and in LCAP.

The new state measure would include five performance indicators. Achievement data—test scores—are an obvious component. We also include preparation for kindergarten, because it is a key factor in long-term student success. Success in college and employment is perhaps the best performance measure for the K–12 system and is included as a high school indicator. Persistence (attendance and dropout data) also is an important indicator because it is a leading metric for student achievement. The fifth indicator is school environment, as reported by teacher and student survey results. Teacher attitudes reveal whether educators are working together to improve curriculum and instruction, and whether the district is providing the support needed for improvement. Student attitudes provide a second perspective on the school environment.

TABLE 6
State and local accountability indicators

Measure	Indicators
New state accountability measure	Achievement, kindergarten preparation, persistence, postgraduate success, school environment
Revised Local Control Accountability Plan	Long-term EL rate, percent disabled students, suspensions and expulsions, college preparation, teacher credentialing, textbook supply, facility quality, parental engagement

LCAP would retain most of the remaining indicators that are currently in law. The revised list in Figure 6 excludes several that do not work well as performance indicators, including implementation of CCSS. We also replaced the two existing EL indicators with the long-term EL rate (students that take more than five years to transition to fluency). We also added an indicator of special education performance (percent disabled students), as LCAP did not include a measure of district performance in this area outside of test scores.

Below, we provide a more detailed justification for the four non-test performance indicators we suggest for the state measure. Most of the data used in these indicators are collected by districts and are included in LCAP and SQII. A few indicators are not, such as kindergarten preparation and success in college. In

addition, several are not collected at the state level. Thus, our proposal would result in a significant new data collection effort for CDE.

Kindergarten preparation

Students who are prepared academically and socially for the rigors of education create multiple benefits for schools. First, these students are likely to do well in class. Second, because they are prepared, they do not create the need for teachers to concentrate on material that should have been mastered previously. When most students have the needed prerequisite skills, teachers have more time to assist lagging students and then to take the entire class to a deeper level of learning. As a result, the entire school benefits. These positive “externalities” suggest that the system should place a high value on students arriving in kindergarten relatively well prepared for instruction.

Research has shown that preparation for school begins in the very early years of a child’s life. Indeed, the benefits of early childhood education and preschool are well documented. Children who attend preschool generally experience higher rates of academic success in the early elementary grades and score higher on social and emotional development measures (Yoshikawa et al., 2013).

Thus, readiness for kindergarten can be measured in two ways. One, the state could develop a kindergarten evaluation tool that assesses student preparedness. Ohio has implemented an assessment for entering kindergarten students that covers physical, social, and academic skills. The assessment is not used to exclude children from school but to provide information to teachers and parents.²²

An alternative measurement is to determine whether students attended a preschool or child care program that included an educational component. This option is likely to be substantially less accurate than the readiness assessment, as attendance in such programs reflects neither the quality of it nor the intensity of the child’s participation. In addition, parenting style also is a factor in readiness that is not measured by program-participation data. However, unless the state adopts an assessment program (such as Ohio’s), preschool participation is the only existing way to gauge any preparation for kindergarten.

Ensuring that students are prepared for kindergarten puts an impossible burden on educators because it depends on events that occur outside school. However, the idea is simply to encourage educators to work with parents, preschool programs, and city and county governments to promote preparation, not to make the K–12 system responsible for guaranteeing preschool to all children.²³ There are ways to boost preparation activities—cooperative parent preschools, parenting education classes, training to increase the quality of existing center- and home-based care—that do not require the expense of full-time classes. Districts in poorer parts of the state may find this coordination constitutes a bigger challenge than districts in more affluent areas, but low-income students benefit more from attending preschool, which makes the need for coordination that much more important.

Persistence

Persistence—the quality that allows someone to persevere in the pursuit of a goal—is a meaningful indicator for schools because it reflects a student’s commitment to education. Attendance is one measure of persistence. Most students need to attend school every day to learn at high levels. Most curricula are

²² See <http://education.ohio.gov/Topics/Early-Learning/Guidance-About-Kindergarten/Ohios-New-Kindergarten-Readiness-Assessment>.

²³ In 2014, the legislature proposed to make pre-kindergarten available to all four-year-olds. This class would essentially make preschool universal. See <http://sbud.senate.ca.gov/sites/sbud.senate.ca.gov/files/SUB1/05222014Sub1PartA.pdf>.

designed to deliver content to students in the classroom, with homework largely reinforcing skills acquired during the day. As a result, attendance signals that students are receiving this content. Students who do not attend class regularly may never learn what they missed, and probably develop problems in the future.

There also may be a deeper significance to attendance: that it signals whether the entire system is functioning to ensure that students get a good education. Attendance and persistence problems reveal the potential for long-term costs in the form of academic deficits, remediation needs, and even long-term societal costs. For older students, failure to attend regularly suggests that they may be disengaged from school, and it is a leading indicator for dropping out. When students miss classes and fall behind in their studies, it significantly reduces the odds of success in school (National Research Council, 2004). Emphasizing attendance and persistence also encourages schools to devote more attention to both the personal and academic issues students face. Research shows that many students drop out because they fall behind in their studies. A wide range of personal and social issues also contribute to dropping out (National Research Council, 2004).

Using attendance and persistence data as performance indicators, therefore, strengthens incentives for schools to reduce costs. Still, even though districts have strong financial incentives to maximize attendance, district data from 2012–13 indicate that attendance is a significant problem in California.²⁴ Using financial data reported for K–8 districts in the state, we calculated an absence rate of about 8.5 percent—a rate that is close to the 10 percent standard for chronic absenteeism. High school district data show absence rates exceeding 12.5 percent. Thus, the data suggest there is significant room for improvement in this area.²⁵

Data collection needs to improve. Both SQII and LCAP include indicators of attendance. Currently, though, only dropout and graduation data are valid, reliable, and collected at the school level. Truancy rates (the proportion of students with three or more days of unexcused absence) also are collected at the school level, but the accuracy of the data is unknown. In addition, truancy rates may not be as good an indicator as chronic absenteeism (percentage of students absent more than 10% of the time). Attendance is collected for fiscal purposes, but the state does not collect school- or student-level attendance. Chronic absenteeism data are not collected. In the near term, therefore, the state’s options are limited regarding dropout, graduation, and truancy data.

Postgraduation success

Student success after graduation is a reflection of the initiative and accomplishments of students. Neither SQII nor LCAP include data on student success after high school. However, LCAP includes several measures of college preparation, including SAT test results and scores on Advanced Placement and International Baccalaureate courses. LCAP also includes the proportion of students taking all of the “A–G” courses that students interested in attending state university in California must pass.

Actual data on college success is preferable to indicators of preparation for college. In 2008–09, 74 percent of high school graduates in California were enrolled in college or university. Of those attending one of the three state postsecondary systems, however, only 55 percent earned one year of college credits during the first two

²⁴ State funding is based on attendance (not enrollment).

²⁵ Elementary and high school district data from 2012–13, author’s calculation using revenue limit data from Ed-Data website (www.ed-data.k12.ca.us). Mobility into and out of public schools (including dropouts) affects our estimates. Districts with attendance rates of less than 50 percent are excluded from this calculation on the assumption that there are errors in the data.

years after graduation.²⁶ Students failed to earn more credits, in part, because they dropped out or were required to take remedial coursework. A 2005 study found 40 percent dropout rates among recent high school graduates that attended community college part time (Warren, 2005). California State University finds that passing the “A–G” courses does not ensure that entering freshmen possess college-level skills in mathematics or English (Legislative Analyst’s Office, 2011).

In addition, academic preparation ignores other critical nonacademic factors that affect college success. For example, students must navigate the application process to one or more colleges or universities, obtain the necessary financial resources, enroll in classes, and then negotiate the personal challenges that come with attending college. Especially for low-income students, academic preparation alone does not automatically lead to college or university success, and each step in the arduous process reduces the percentage of students who attend college and successfully complete their first year in higher education (Roderick et al., 2011). The 2005 study mentioned above also found that 25 percent of recent high school graduates attending community college did not list a goal for what they wanted to achieve in college, suggesting they had not decided why they were attending college (Warren, 2005).

The existing measures of preparation primarily reflect the academic skills needed for study at California’s two state university systems. However, most students do not attend a university right after graduation. Many attend community college or work. In general, community colleges do not have well-specified entrance requirements for either academic or vocational studies programs communicating to students the skills needed to take college-level courses.²⁷ Such lack of sound information on preparing for community colleges is a significant problem for high schools in motivating students and advising them (Kirst and Venezia, 2006).

To address these data problems, our state measure would use data that are available through the National Student Clearinghouse on all California high school students attending a private or public university or college. This data would allow the state to track the success of students in completing the first year of college-level coursework. The preparation indicators—the percentage meeting the “A–G” requirements and the percentage that are successful on AP tests—would remain in LCAP.

Data on actual attendance and success in college are not perfect indicators, as some factors fall outside of a district’s control. For instance, state budget cuts can reduce postsecondary opportunities, making school and district performance appear worse than it is. In addition, low-income students may have fewer higher-education alternatives than students whose families can afford private-school tuition. On balance, however, attendance and success data create important incentives for educators to consider the broader range of information and skills students need for college.

The state’s K–12 system also has no standards for the skills needed in the labor market. Schools and districts maintain vocational or other career-oriented programs, which may impart technical skills and information useful in specific industry sectors. But the baseline academic and non-cognitive skills expected by employers are not explicitly part of the K–12 curriculum. The CCSS and SBAC tests may contribute in this area, as the 11th grade tests are expected to identify academic achievement levels needed for “college and career.” Still,

²⁶ From CDE’s DataQuest website.

²⁷ Community colleges also provide pre-collegiate courses, and many recent high school graduates are required to take these “remedial” courses before taking college-level classes.

despite the fact that most K–12 students do not end up graduating from college or university, preparing for working after high school remains an underdeveloped part of the K–12 infrastructure.

Measuring the success of schools in helping students get good jobs also has received little attention previously. Other states calculate employment rates and wages of recent graduates using state employment data. But at the current time, California’s wage database cannot link with K–12 data and, therefore, is not useful for evaluation purposes (Warren and Hough, 2013). CDE is working on the issue of measuring preparation for careers as part of a statutory requirement to add indicators to the API, and it has identified several potential indicators (such as completing a sequence of vocational courses). In the meantime, preparation indicators will have to suffice until the state develops a process of linking K–12 data with employment data.

School environment

Both LCAP and SQII include surveys of parents, students, and teachers as indicators. LCAP includes survey results from the three groups as indicators of school safety and school connectedness. SQII also surveys students and parents on school climate. Teachers are queried about the school environment as well as support from their principal and district administrators. Neither program identifies specific surveys. So, in the case of LCAP, this means there will likely be little uniformity across the state in how safety and school connectedness are measured and evaluated.

The interest in this area stems from research that shows that school environment affects student achievement. Students with low school or social connectedness are more likely to drop out of high school and have mental or substance-abuse problems as adults (Cohen, 2009; Bond, 2007). Many states and districts have developed surveys on these topics.²⁸ Chicago public schools use survey data to evaluate and report whether schools have the “essential” supports that are needed for academic success. The University of Chicago Consortium for School Research, which developed these reports, found that school organization plays a key role in the success of schools. In fact, it determined that schools that contain the five supports—school leadership, collaborative teachers, involved families, supportive environment, and ambitious instruction—are much more likely to substantially increase student achievement than those with only a few of these key supports (Byrk, 2010).

The survey responses generate a score on each of the five supports. Interestingly, these supports overlap in many ways with the Deming guidelines: leadership, collaboration, support for improvement, and high goals are all key elements of Deming’s system. This link between essential supports and student achievement makes the Chicago survey very attractive for accountability programs. The survey results could provide a different perspective on the quality of schools. And because the five supports predict higher achievement, the survey results would provide useful diagnostic information for administrators and teachers, helping to focus school-improvement activities.²⁹

²⁸ CDE has developed survey instruments to measure student health issues (the “healthy kids” survey), the school learning environment (the “school climate” survey), and a survey of parents on the topics in the healthy kids and school climate surveys. CDE has promoted its surveys to districts as a way of determining whether schools maintain the conditions and supports needed to improve achievement. See *Helpful Resources for Local Control & Accountability Plans and School Safety Plans, School Climate, Health, and Learning: California Survey System*, Wested, 2014, available at http://cscs.wested.org/resources/LCAP_Cal_SCHLS.pdf.

²⁹ See <https://uchicagoimpact.org/5essentials> for more information on the research, surveys, and actual school ratings using the survey data.

As with test scores, including these indicators in a school- or district-accountability measure could threaten the validity of the data: teachers and students could inflate their assessment of learning conditions to get a higher accountability score. But getting honest answers could hinge on teachers and students believing that their responses would be used to improve conditions at the school, and not lead to future sanctions. Therefore, it would be important to emphasize that obtaining accurate survey data is in the best interests of students and teachers, and it could serve as a valuable indicator of school progress.

Creating good incentives

The next step in the development of our state accountability measure is to use the five performance indicators strategically to create positive incentives for student achievement. The indicators used to develop the accountability measure are: school readiness, current achievement, school environment, persistence, and long-term success. Table 7 displays the specifics of the measure.

The measure adapts the lessons from our analysis of NCLB and the three state programs. It uses valid and reliable data, minimizes the use of any one indicator to protect the validity of the information, and creates incentives to focus on both immediate and long-term student achievement. All of the indicators have a close relationship with student success, creating a coherent set of outcomes for schools and districts to concentrate on.

Readiness. The readiness indicator creates an incentive for schools to help ensure that kindergarten students are academically ready when they enter school. Specifically, each elementary school's readiness rating would reflect two elements. First, it would include a state-developed readiness assessment, or administrative data on the percentage of students who attended preschool. Second, changes in the percentage of 3rd grade students who were performing at "grade level" in mathematics and English also would be added to the readiness indicator. This would create a performance target for the early elementary grades that is based on research showing that children who cannot read fluently by the end of 3rd grade often struggle throughout their years in school.

Current achievement. Testing data on the level of achievement and individual growth from the previous year helps maintain a school's focus on academic performance. Both measures of the level and growth in achievement are needed to adequately evaluate school quality. Similar to the API, the indicator of current achievement would assign extra points for raising the scores of lower-performing students. Thus, the three measures would provide a balanced appraisal to educators.

School environment. Surveys of teachers and students would be used to calculate a school environment rating, which would promote good management of schools and the instructional process. Data on the five essential supports would constitute an important complement to testing data as an indicator of a school's capacity to function at high levels. The survey data would provide diagnostic information on the operation of the school, indicating whether teachers and students feel supported, and how teachers view the quality of the school's leadership.

Persistence. Evaluating schools based on attendance and dropout rates puts a spotlight on getting students in their seats on a daily basis. However, because of data limitations, current-year truancy and dropout rates would be included in middle and high school accountability measures. Elementary schools would include only current-year truancy rates. In the longer run, the state should study whether collecting daily attendance data is necessary, and whether truancy or chronic absenteeism (or a combination of them) creates strong incentives for schools to address attendance problems without the cost of collecting daily attendance data.

Long-term success. As discussed earlier, including long-term success indicators in an accountability measure reinforces the academic goals of the program. So, we assign school-testing data from former students at the next level of schooling. For instance, an elementary school rating would include test data on 8th grade students that previously attended the elementary school (the specific indicator would be based on achievement growth or the percentage of students who performed in the lowest quintile or performance level). High schools would use postgraduation college success rates and selected employment-preparation data. This indicator accomplishes two ends: it creates an incentive to teach in ways that build the longer-term success of a school’s students, and it generates incentives for schools to work together to ensure student success.

TABLE 7
Five elements of a new state accountability measure

Elementary	
Readiness	Change in the percentage of kindergarten students who attended preschool; the percentage of 3rd- grade students reading below “grade” level.
Achievement	Levels and growth of student achievement in grades 4 and 5; raising the performance of low-achieving students.
Environment	Survey data on essential supports in grade 5.
Persistence	Truancy rates.
Long-term success	Achievement growth from grades 5–8. The percentage of low-achieving 8th grade students of those that attended the elementary school.
Middle	
Achievement	Levels and growth of student achievement; raising the performance of low-achieving students.
Environment	Survey data on essential supports in grades 6–8.
Persistence	Truancy rates and middle school dropout rates.
Long-term success	High school exit examination passage rates; proportion of “on track” 10th graders.
High	
Achievement	Percentage of students meeting college and career achievement levels.
Environment	Survey data on essential supports in grades 9–12.
Persistence	Truancy rates; high school dropout rates; graduation rates.
Long-term success	The percentage of students who complete the first year of college; the percentage earning a vocational certificate or completing a sequence of vocational courses.

Linking state and local accountability

Once indicators for the statewide accountability measure have been identified, LCAP should be modified to give them priority. Because almost all of the indicators in our design are currently in LCAP, this means making it clear in statute that districts will be held accountable for progress based on indicators in the new state measure. The remainder of LCAP indicators are provided as secondary indicators, although they could represent important local outcomes.

The state should also begin the process of collecting data for LCAP and the new state accountability measure through the state’s student-level database, the California Longitudinal Performance Assessment Data System

(CALPADS). Clearly defining the indicators and collecting the data at the state level will have several benefits. Most obviously, it will improve the quality of the data and increase the consistency of definitions of variables used in the state-local accountability program. It also will allow the state to post on its website the accountability data for districts and schools. This would reduce the burden of data reporting on districts, create one source for local information on both state and local accountability measures, and allow CDE to post with each district's data comparable statewide school and district data.

Local reviews and technical assistance. As suggested earlier, the state accountability program would take advantage of LCAP's innovative use of county offices of education to oversee the local improvement process. The local plans would give priority to the indicators in the state's accountability program, but other LCAP indicators also could be included, and county superintendents would use LCAPs as a monitoring instrument.

County office roles also would be broadened. Currently, LCAP limits the scope of county office plan reviews to the requirements of state law. County offices may recommend changes to a plan or offer technical assistance to districts, but the county has no authority to require specific district actions. This setup is similar to the 1980's prior to the passage of AB 1200: county offices reviewed local budgets but had no ability to halt faulty district financial practices. AB 1200 gave county offices the *responsibility* to head off local financial problems, as well as the power to do so. LCAP could be structured similarly. For most districts, the LCAP review process would provide an external assessment of a district's educational and financial plan, and county recommendations would be advisory only.

County superintendents would be given authority to require changes in LCAPs in districts that are very low performing, have very low-performing subgroups, or have declining scores on the state's accountability measure over several years. We hope the use of that power would be rare, as imposing changes in academic programs that are not supported by districts do not have a high probability of success. But giving county offices this authority is necessary for underperforming districts to seriously entertain the county's feedback. That would give county offices traction to perform their function most effectively.

State technical assistance. CDE would have several roles under the linked state-local program. Most important, it would oversee the quality of county-operated programs and build the capacity of county offices to monitor academic quality. This will require the department to understand the improvement process from both the county and district perspective so that it can develop and disseminate materials and training that meet county needs.

CDE also is a member of the California Collaborative for Educational Excellence, which was created to provide advice and assistance to districts as a part of LCAP. The collaborative could play a role similar to the one the Fiscal Crisis and Management Assistance Team (FCMAT) performs in the fiscal arena. FCMAT serves as a management consultant on fiscal and administrative issues, and it also represents the state as fiscal experts in districts that are in financial peril. In addition to its current duties, the collaborative could be charged with working with county offices and districts to evaluate and improve LCAPs for the lowest-performing districts in the state. The collaborative also could provide a neutral assessment of local plans when counties and districts disagree.

CDE also would monitor the need for technical assistance in districts and county offices. Currently, the state spends \$10 million in federal funds for technical assistance. This amount should be increased significantly. To flesh out the issues involved in this effort, CDE should develop with the county offices a multiyear plan

for the types of services districts need, the amount of funding the state should make available, the types of services needed by districts, and the share that should be administered through county offices as opposed to the California Collaborative for Educational Excellence.

The last component of the linked-accountability program is creating district sanctions in state law. One option is to follow the Massachusetts model, in which schools and districts that are unable to improve performance after receiving significant assistance from the state are assigned a receiver who assumes control from the school board (this, too, is a feature of AB 1200). A second approach is to replace the school board if district-wide performance fails to respond to technical assistance. This would maintain local control over the district but with a new group of leaders who would be charged with taking the necessary actions to move the district in a positive direction.

Conclusion

California is beginning a transition in its curriculum, its tests of student achievement, and its funding and local-accountability system. Specifically, the current school year finds districts implementing the Common Core State Standards in English and mathematics. Student mastery of these standards will be assessed with tests developed by the Smarter Balanced Assessment Collaborative (SBAC). The Local Control Funding Formula provides districts with new funding and flexibility—and new performance mandates through LCAP.

Less attention, though, has been paid to how these developments affect state and federal accountability programs in the state. With the addition of LCAP, there are now four K–12 accountability programs in California. The multiplicity of goals and performance indicators is confusing. The state continues to refine the API despite the fact there are more accurate accountability models available. LCAP establishes a renewed focus on local accountability, but its 19 indicators and lack of priorities obscures the state’s goals for schools rather than clarifying them. Two new programs—LCAP and CORE’s waiver program—expand the range of performance outcomes, but they rely on data that has uncertain effects when used as accountability indicators.

The state can do little to change NCLB and its emphasis on sanctions until it obtains a federal waiver or the law is reauthorized. Still, the law’s measure of performance and its emphasis on school sanctions are problematic. In fact, California’s experience suggests that the federal accountability program should be reframed around the need to build local capacity, enlisting teachers in dialogues about improving curriculum and instruction. The CORE districts’ waiver program also is off-limits to state policymakers. Its program, though, introduces several interesting innovations that could benefit the state in the long run. So, rather than ending the program, the state should assess whether CORE’s innovations are effective in boosting student achievement and success.

However, California can—and should—revise the state’s accountability programs. The API is outdated and needs to be replaced. Its design addressed the desire to measure growth in school success at a time when the state’s educational technology was not sufficiently developed. In 2015, the new SBAC tests will be able to measure achievement growth at the student level, which undercuts the need to keep the API model. In addition, other states have developed measures that are much more accurate and useful than the API. The need to work on a new measure will become more pressing, as reauthorization (or a federal waiver) of NCLB will likely facilitate a move away from one federal performance measure in favor of individual state measures. This gives the state several years to develop a measure and test it for use after reauthorization of NCLB.

LCAP also needs attention. The statute is not clear whether the program is a substitute or complement to state accountability. Moreover, there are too many goals, and no clear way to determine whether districts have made sufficient progress improving student outcomes. These issues are compounded by performance indicators that are either not defined or create the potential for negative local incentives.

Nevertheless, the problems with LCAP should not overshadow its important contributions. The program attempts to strengthen local accountability by bringing parents and community groups into budget and planning discussions. In addition, using county offices of education to act as a quality check for districts also represents an important innovation. Indeed, in this large and diverse state, county offices seem to be the most logical avenue for providing annual high-quality feedback and technical assistance that many districts need to function more effectively.

Our suggestions for addressing these complex and multifaceted issues have three parts. First, the legislature and governor should merge a new state performance measure into the LCAP program. The new state measure should be simple, statistically valid and reliable, and create strong incentives for schools to focus on improving student success. Second, the administration should organize and fund a larger program of technical assistance. Currently, the state dedicates \$10 million in federal funds to technical assistance. This amount should increase substantially. To get a handle on the issues involved in this effort, CDE should develop a multiyear plan for the types of services districts need, the amount of funding the state should make available, and the share that should be administered through county offices as opposed to the California Collaborative for Educational Excellence.

Third, the legislature and governor need to address governance arrangements of accountability programs. County offices should receive more authority to require changes to district plans. The LCAP process would grow into an annual local review of district strengths and weaknesses, with county offices empowered to prod districts to improve each year. For districts with moderate problems, technical assistance would be optional. For districts with more severe problems, technical assistance would become more directed as needed to protect student interests. The state also should create consequences for districts that fail to improve after significant investments of technical assistance. CDE would be the intervener of last resort, when districts are unable to develop the consensus to focus on student success.

Our proposal represents one of many possible options for revamping California's accountability arrangements. It builds on recent reforms that strengthen local accountability. Our suggested state accountability measure makes improving student achievement the priority. This would help districts formulate local plans while also giving them flexibility to address local priorities. Our plan also increases the role of county offices in becoming an important accountability checkpoint. By also giving county offices additional technical assistance resources, our plan envisions accountability as a process of building capacity in school districts to better educate students.

One drawback of our plan is that it does not offer quick results. Districts would have to be willing to consider new ways of operating, and such change often starts small and takes time to develop. County offices would have to grow into new responsibilities in the instructional arena. But getting away from school-level sanctions would reduce teacher anxiety and increase their willingness to try new instructional approaches. Also, emphasizing technical assistance directly would address the problem of administrators often not knowing how to improve low-performing schools. Therefore, if our analysis is correct, approaching accountability as a learning process offers a greater likelihood of significant improvement over time than the current system.

References

- Bond, Lyndal, Helen Butler, Lyndal Thomas, John Carlin, Sara Glover, Glenn Bowes, and George Patton. 2007. "Social and School Connectedness in Early Secondary School as Predictors of Late Teenage Substance Use, Mental Health, and Academic Outcomes." *Journal of Adolescent Health* 40.
- Brown, James Dean. 2000. *What Is Construct Validity? Questions and Answers about Language Testing Statistics*. University of Hawai'i at Manoa. Available at <http://jalt.org/test/PDF/Brown8.pdf>.
- Byrk, Anthony S., and Barbara Schneider. 2002. "Trust in Schools: A Core Resource for Schools." In *The Four Elements of Trust*, Devin Dovicka. National Association of Secondary School Principals. 2006. Available at www.nassp.org/portals/0/content/54439.pdf.
- Byrk, Anthony S., Penny Bender Sebring, Elaine Allensworth, Stuart Luppescu, and John Q. Eason. 2010. "Organizing Schools for Improvement, Lessons from Chicago." University of Chicago Consortium on Chicago School Research.
- California Department of Education. 2011. Standardized Testing and Reporting research database. Available at <http://star.cde.ca.gov/>.
- California Department of Education. 2013. "2012–13 Accountability Progress Reporting System: Summary of Results." Available at www.cde.ca.gov/nr/ne/yr13/yr13rel78attb.asp.
- California County Superintendents Educational Services Association. April 30, 2014. *Local Control Accountability Plan (LCAP) Approval Manual, 2014–15 edition*. Available at http://ccsesa.org/wp-content/uploads/2014/04/CCSESA-LCAP-Approval-Manual-2014-15_May22.pdf.
- California State Auditor. March 2012. "High School Graduation and Dropout Data: California's New Database May Enable the State to Better Serve Its High School Students Who Are at Risk of Dropping Out." Report 2011-117. Available at www.bsa.ca.gov/pdfs/reports/2011-117.pdf.
- Cohen, Jonathan, Elizabeth M. McCabe, Nicholas M. Michelli, and Terry Pickeral. 2009. "School Climate: Research, Policy, Practice, and Teacher Education." *Teachers College Record*. January. Available at <https://schoolclimate.org/climate/documents/policy/School-Climate-Paper-TC-Record.pdf>.
- Dee, Thomas S., and Brian Jacob. 2011. "The Impact of No Child Left Behind on Student Achievement." *Journal of Policy Analysis and Management*. Available at http://deepblue.lib.umich.edu/bitstream/handle/2027.42/86808/20586_ft.pdf?sequence=1.
- Deming, William Edward. 1982. *Out of the Crisis*. MIT Press.
- Fagioli, Loris P. 2014. *A Comparison Between Value-Added School Estimates and Currently Used Metrics of School Accountability in California*. Springer Science.
- Fletcher, Stephen, and Margaret Raymond. 2002. "The Future of California's Academic Performance Index." Hoover Institution, Stanford University. April. Available at <http://credo.stanford.edu/downloads/api.pdf>.
- Forte Fast, Ellen, and Steve Hebbler. 2004. "A Framework for Examining Validity in State Accountability Systems. Council of Chief State School Officers." February. Available at www.ccsso.org/Resources/Publications/A_Framework_for_Examining_Validity_in_State_Accountability_Systems.html#sthash.OR0YTSNz.dpuf.
- Fullan, Michael. 2006. "Change Theory: A Force for School Improvement." Center for Strategic Education. Seminar Series Paper No. 157. Available at www.michaelfullan.ca/media/13396072630.pdf.
- Fullan, Michael. 2008. "The Six Secrets of Change." Available at www.michaelfullan.ca/images/handouts/2008SixSecretsofChangeKeynoteA4.pdf.
- Fullan, Michael. 2011. "Choosing the Wrong Drivers for Whole System Reform." Center for Strategic Education. Seminar Series Paper No. 204. Available at www.michaelfullan.ca/media/13501655630.pdf.
- Glazerman, Steven M., and Liz Potamites. 2011. "False Performance Gains: A Critique of Successive Cohort Indicators." Working paper, Mathematic Policy Research. December. Available at http://www.mathematica-mpr.com/~media/publications/PDFs/Education/False_Perf.pdf.

- Goldschmidt, Peter, Pat Roschewski, Kilchan Choi, William Auty, Steve Hebbler, Rolf Blank, and Andra Williams. 2005. "Policymakers' Guide to Growth Models for School Accountability: How Do Accountability Models Differ?" Council of Chief States School Officers. October. Available at www.ccsso.org/Documents/2005/Policymakers_Guide_To_Growth_2005.pdf.
- Hahnel, Carrie. 2014. "1,000 LCAPs Later, Let's Hope We Learn Something." *EdSource*, July 9. Available at <http://edsources.org/2014/1000-lcaps-later-lets-hope-we-learn-something/65263#.U9lXnqP5dLc>.
- Harr, J. J., Parrish, T., Socias, M., and Gubbins, P. 2007. "Evaluation Study of the High Priority Schools Grant Program: Final Report." American Institutes for Research. Available at <http://www.air.org/publications/FinalHPRReport>.
- Hill, Laura, Margaret Weston, and Joseph M. Hayes. 2014. *Reclassification of English Learner Students in California*. Public Policy Institute of California. Available at www.ppic.org/main/publication.asp?i=1078.
- Hout, Michael, and Stuart W. Elliott, eds. 2011. *Incentives and Test-Based Accountability in Public Education*. National Academies Press. Available at www.nap.edu/openbook.php?record_id=12521.
- Hughes, Teresa A., and William Allan Kritsonis. 2006. "A National Perspective: An Exploration of Professional Learning Communities and the Impact of School Improvement Efforts." *National Journal for Publishing and Mentoring Doctoral Student Research*. Available at <http://files.eric.ed.gov/fulltext/ED491997.pdf>.
- Kirst, Michael, and Andrea Venezia. "Improving College Readiness and Success for All Students: A Joint Responsibility Between K–12 and Postsecondary Education." Issue brief for the Secretary of Education's Commission on the Future of Higher Education. U.S. Department of Education. Available at www2.ed.gov/about/bdscomm/list/hiedfuture/reports/kirst-venezia.pdf.
- Legislative Analyst's Office. 2011. "Are Entering Freshmen Prepared for College-Level Work?" Legislative Analyst's Office, Higher Education. Answers to Frequently Asked Questions, Issue 2 (updated): March. Available at www.lao.ca.gov/sections/higher_ed/FAQs/Higher_Education_Issue_02.pdf.
- Linn, Robert L., 2006. "Educational Accountability Systems." CSE Technical Report 687. National Center for Research Evaluation, Standards and Student Testing (CREST). June. Available at www.cse.ucla.edu/products/reports/r687.pdf.
- Massachusetts Department of Elementary and Secondary Education. 2014. "Accountability, Partnership, and Assistance: Level 5 Districts." Available at www.doe.mass.edu/apa/sss/turnaround/level5/districts/default.html.
- Meyer, Robert H., 2008. "Value-Added and Other Methods for Measuring School Performance." National Center on Performance Incentives. Working paper 2008–17. February. Available at https://my.vanderbilt.edu/performanceincentives/files/2012/10/200817_MeyerChristian_ValueAdded.pdf.
- Messick, Samuel. 1990. "Validity of Test Interpretation and Use." Educational Testing Service August. Available at <http://files.eric.ed.gov/fulltext/ED395031.pdf>.
- Murnane, Richard J. 2013. "U.S. High School Graduation Rates: Patterns and Explanations." National Bureau of Economic Research. Working paper 18701. January.
- National Research Council. Committee on Increasing High School Students' Engagement and Motivation to Learn. 2004. *Engaging Schools: Fostering High School Students' Motivation to Learn*. The National Academies Press.
- Nichols, Sharon, and David C. Berliner. 2005. "The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing." Education Policy Research Unit (EPRU), Arizona State University. Available at <http://files.eric.ed.gov/fulltext/ED508483.pdf>.
- Nichols, Sharon. 2007. "High-Stakes Testing: Does It Increase Achievement?" *Journal of Applied School Psychology* (The Haworth Press, Inc.) Vol. 23, No. 2. Available at <http://scottbarrykaufman.com/wp-content/uploads/2012/01/High-Stakes-Testing1.pdf>.
- Parrish, Tom, Catherine Bitter, Maria Perez, and Raquel Gonzalez. 2005. Evaluation Study of the Immediate Intervention/Underperforming Schools Program of the Public Schools Accountability Act of 1999. American Institutes for Research. Available at www.air.org/sites/default/files/downloads/report/IIUSP_Report_FINAL_9-30-05_0.pdf.
- Perie, Marianne, Judy Park, and Kenneth Klau. 2007. "Key Elements for Educational Accountability Models." Council of Chief State School Officers. December. Available at www.ccsso.org/documents/2007/key_elements_for_educational_2007.pdf.

- Polikoff, Morgan S., Andrew McEachin, Stephani L. Wrabel, and Matthew Duque. 2013. "The Waive of the Future: School Accountability in the Waiver Era." (Presented at the Association for Education Finance and Policy Annual Conference, March.)
- Ragosa, David. 2005. "A School Accountability Case Study: California API Awards and the Orange County Register Margin of Error Folly." In *Defending Standardized Testing*, ed. Richard Phelps. Lawrence Erlbaum Associates Inc.
- Riddle, Wayne. 2012. "What Impact Will NCLB Waivers Have on the Consistency, Complexity and Transparency of State Accountability Systems?" Center on Educational Policy, The George Washington University. October. Available at www.cep-dc.org/displayDocument.cfm?DocumentID=411.
- Roderick, Melissa, Vanessa Coca, and Jenny Nagaoka. 2011. "Potholes on the Road to College: High School Effects in Shaping Urban Students' Participation in College Application, Four-year College Enrollment, and College Match." University of Chicago Consortium on Chicago School Research. July. Available at https://ccsr.uchicago.edu/sites/default/files/publications/SOE_Potholes.pdf.
- Schwartz, Healthier L., Laura S. Hamilton, Brian Stecher, and Jennifer L. Steele. 2011. *Expanded Measure of School Performance*. RAND Corporation. Available at www.rand.org/pubs/technical_reports/TR968.html.
- Shepard, Lorrie A. 1990. Inflated Test Score Gains: Is It Old Norms or Teaching the Test? CSE Technical Report 307. UCLA Center for Research on Evaluation, Standards, and Student Testing. Available at www.cse.ucla.edu/products/reports/TR307.pdf.
- Shortell, Stephen M., James L. O'Brien, James M. Carman, Richard W. Foster, Edward F. X. Hughes, Heidi Boerstler, and Edward J. O'Connor. 1995. "Assessing the Impact of Continuous Quality Improvement/Total Quality Management: Concept versus Implementation." *Health Services Research*. June.
- United States Department of Education. NCLB Flexibility (updated June 7, 2013). Available at www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html.
- Usher, Alexandra. 2012. *AYP Results for 2010–11* (November 2012 update). Center for Educational Policy, Education Commission of the States. Available at www.cep-dc.org/publications/index.cfm?selectedYear=2012.
- Walpole, Mary Beth, and Richard J. Noeth. 2002. "The Promise of Baldrige for K–12 Education." ACT Office of Policy Research. Available at www.act.org/research/policymakers/pdf/baldrige.pdf.
- Walsh, Kate. 2001. "Teacher Certification Reconsidered: Stumbling for Quality." The Abell Foundation. Available at www.nctq.org/dmsView/Teacher_Certification_Reconsidered_Stumbling_for_Quality_NCTQ_Report.
- Warren, Paul. 2005. "Improving High School: A Strategic Approach." California Legislative Analyst's Office. May. Available at http://lao.ca.gov/2005/high_schools/improving_hs_050905.pdf.
- Warren, Paul. 2007. "Improving Alternative Education in California." California Legislative Analyst's Office. February. Available at http://lao.ca.gov/2007/alternative_educ/alt_ed_020707.pdf.
- Warren, Paul. 2013. *California's Changing Accountability Program*. Public Policy Institute of California. Available at www.ppic.org/main/publication_quick.asp?i=1043.
- Warren, Paul, and Heather Hough. 2013. "Increasing the Usefulness of California's Education Data." Public Policy Institute for Research. August. Available at www.ppic.org/main/publication.asp?i=1067.
- WestEd. 2009. "Helping Students Who Transfer to New Schools: An Annotated Bibliography." Regional Education Laboratory West. May. Available at http://relwest-archive.wested.org/system/memo_questions/7/attachments/original/Helping_20students_20who_20transfer_20schools_20May_202009_1_.pdf.
- Westover, Theresa, Katharine Strunk, Andrew McEachin, Amy Smith, Shani Keller, and Mary Stump. 2012. *AB 519 Evaluation: Final Report*. School of Education Center for Education and Evaluation Services, University of California at Davis. May. Available at <http://education.ucdavis.edu/select-publications-and-reports>.
- Wilson, Mark. 2010. "Assessment for Learning and for Accountability." Center for K–12 Assessment & Performance Management. Available at www.k12center.org/rsc/pdf/WilsonPresenterSession4.pdf.
- Yoshikawa, Hirokazu, Christina Weiland, Jeanne Brooks-Gunn, Margaret R. Burchinal, Linda M. Espinosa, William T. Gormley, Jens Ludwig, Katherine A. Magnuson, Deborah Phillips, and Martha J. Zaslow. 2013. "Investing in Our Future: The Evidence Base on Preschool Education." Society for Research in Child Development. October. Available at <http://fcd-us.org/sites/default/files/Evidence%20Base%20on%20Preschool%20Education%20FINAL.pdf>.

About the Author

Paul Warren is a research associate at PPIC, where he focuses primarily on K–12 education finance and accountability. Before he joined PPIC, he worked in the California Legislative Analyst’s Office for more than twenty years as a policy analyst and manager. He also served as deputy director for the California Department of Education, helping to implement the state’s testing and accountability programs. He holds a master’s degree in public policy from Harvard’s Kennedy School of Government.

Acknowledgments

The author would like to acknowledge the time and assistance of Rob Manwaring, Jim Soland, and Rick Miller. The report also benefited from the comments and feedback of Hans Johnson, Niu Gao, and Lynette Ubois and Martin Aronson provided excellent editorial input.

Research publications reflect the views of the authors and do not necessarily reflect the views of the staff, officers, or Board of Directors of the Public Policy Institute of California. Any errors are my own.

PUBLIC POLICY INSTITUTE OF CALIFORNIA

Board of Directors

Donna Lucas, *Chair*

Chief Executive Officer
Lucas Public Affairs

Mark Baldassare

President and CEO
Public Policy Institute of California

Ruben Barrales

President and CEO
GROW Elect

María Blanco

Vice President, Civic Engagement
California Community Foundation

Brigitte Bren

Attorney

Louise Henry Bryson

Chair Emerita, Board of Trustees
J. Paul Getty Trust

Walter B. Hewlett

Member, Board of Directors
The William and Flora Hewlett Foundation

Phil Isenberg

Vice Chair
Delta Stewardship Council

Mas Masumoto

Author and Farmer

Steven A. Merksamer

Senior Partner
Nielsen, Merksamer, Parrinello,
Gross & Leoni, LLP

Kim Polese

Chairman
ClearStreet, Inc.

Thomas C. Sutton

Retired Chairman and CEO
Pacific Life Insurance Company

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research on major economic, social, and political issues. The institute's goal is to raise public awareness and to give elected representatives and other decisionmakers a more informed basis for developing policies and programs.

The institute's research focuses on the underlying forces shaping California's future, cutting across a wide range of public policy concerns, including economic development, education, environment and resources, governance, population, public finance, and social and health policy.

PPIC is a public charity. It does not take or support positions on any ballot measures or on any local, state, or federal legislation, nor does it endorse, support, or oppose any political parties or candidates for public office. PPIC was established in 1994 with an endowment from William R. Hewlett.

Mark Baldassare is President and Chief Executive Officer of PPIC.

Donna Lucas is Chair of the Board of Directors.

Short sections of text, not to exceed three paragraphs, may be quoted without written permission provided that full attribution is given to the source.

Research publications reflect the views of the authors and do not necessarily reflect the views of the staff, officers, or Board of Directors of the Public Policy Institute of California.

Copyright © 2014 Public Policy Institute of California
All rights reserved.
San Francisco, CA

PUBLIC POLICY INSTITUTE OF CALIFORNIA
500 Washington Street, Suite 600
San Francisco, California 94111
phone: 415.291.4400
fax: 415.291.4401
www.ppic.org

PPIC SACRAMENTO CENTER
Senator Office Building
1121 L Street, Suite 801
Sacramento, California 95814
phone: 916.440.1120
fax: 916.440.1121