



During the next two years, EdSource will develop a series of briefs exploring issues related to the development of a comprehensive data system in California and the opportunities and challenges the state needs to address. This brief is the first in that series.

California's Emerging Education Data System: A Status Report

HIGHLIGHTS

- Data based on individuals can yield more comprehensive and accurate information about school progress2
- The work of the national Data Quality Campaign highlights potential goals for California and points out places where the state's current plans for its data system may still fall short.....3
- The long-awaited CALPADS system that collects data on individual students is set for implementation in 2009-104
- Work is starting on CALTIDES, a similar system to be built using data on individual educators6
- Current implementation plans leave out some K-12 data and do not include linkages to other state systems, such as higher education8
- Issues related to data quality, local district capacity, and privacy protections present roadblocks to the state's future progress9

During the past decade, the academic progress of California's public school students has become the yardstick by which schools' success is measured. Each year, the state's assessment and accountability systems generate more and more data—results on multiple tests by subject and grade level, graduation rates, demographics, etc.—all disaggregated by student subgroups, tied to specific schools, and for multiple years.

This wealth of new data makes it possible for state leaders, educators, and researchers to more accurately answer many critical questions about the progress of students and the school conditions that support that progress. But to take full advantage of this, California must successfully develop new strategies, implement new systems and technologies, and develop new capabilities for compiling and analyzing those data.

The state is close to completing a major milestone in this effort, the creation of the California Longitudinal Pupil Achievement Data System (CALPADS). This system is based on data for individual students and

educators, an approach endorsed by the state more than a decade ago and required by the federal government in 2002. Its development has taken much longer than expected in large part because of fiscal and political constraints.

After several years of hard work and lots of perseverance, this long-anticipated student data system is now nearing completion. However, much work and many decisions remain as policymakers and educators move forward with its implementation and consider next steps toward making the state's education data system more robust and comprehensive.



Quality data can strengthen education improvement efforts

Accurate, timely, and comprehensive data about how a state's schools are doing helps policymakers, educators, and researchers understand what is working and what is not. With this information, they are better equipped to evaluate which policies and practices best support teaching and learning and to act on those findings.

Data based on individual students and educators provide more accurate information across systems and over time

States can collect education data at the institutional or at the individual level. In the first case, the state collects data on school or district characteristics. Local officials count how many students of each ethnicity are enrolled or how many teachers are fully credentialed, for example. They then submit aggregated data reports to the state. The summary data for institutions tend not to change much from year to year, but the people in them—students, teachers, administrators—do. For example, a school may report the academic improvement of its English learners each year, but these data are not likely to be based on exactly the same children. This approach to data collection results in disconnected databases that might satisfy mandated data reporting requirements but leave big gaps in information about how students are progressing academically.

By contrast, individual-level data are based on each student and educator—the people who make up the institutions. No matter where in the state a student goes, the data system will keep a confidential record of that student's school experiences, following his or her achievement on standardized tests, participation in special programs, and course-taking information over time and across institutions. Similarly, the data system can keep track of the credentials, experiences, and training of educators as they progress in their careers. When local officials submit this individual-level data directly to the state, the California Department of Education (CDE) can then aggregate the data by school, district, and the state over time and across data

elements. This approach offers a more accurate understanding of what is going on in the state's schools as well as the resources (such as teacher preparation programs) that support student learning.

Reliable and accurate student- and teacher-level data can provide information about how well instructional programs are working or how student achievement is improving or declining over time. The data could, for example, shed light on the effectiveness of intervention programs, how well students are progressing in a particularly rigorous course, or the relationship between expenditure decisions and student achievement.

With data documenting individual students' progress over the course of several years, educators and policymakers across the state can test long-standing hypotheses about student achievement. One example comes from a recent study that looked at whether student performance in the early grades can predict students' long-term success. Researchers with the Public Policy Institute of California were able to test this hypothesis at the district level with student-level achievement data from the San Diego Unified School District. With these data, they were able to reliably predict which

students would fail the California High School Exit Exam based on how well they had performed in the fourth grade.

Longitudinal data systems offer information that is more accurate

A "longitudinal" data system tracks individual students' enrollment history, program participation, and achievement on standardized assessments from year to year. These data are compiled in a way that allows for an evaluation of each student's progress, the impact of educational inputs on their progress, and an assessment of program quality. A longitudinal system can also follow teacher characteristics, including their training and longevity, and match those data with data of their students each year. This type of information can support efforts to hold schools, districts, and the state accountable—in a sound, fair way—for successfully educating their students over time.

A data system that collects and provides ongoing management of student-level data statewide enhances the accuracy of a state's existing school performance data. For example, when changes in school-level test scores from year to year are used to measure education progress, some of the changes result

More accurate answers to the following questions will be possible with a longitudinal education data system based on individual-level data

State Policymakers	Are there particular teacher preparation programs whose graduates are more likely to remain in the profession?
District Administrators	Have our students been improving in math at the same rate as students statewide? Are they on track for success in higher grades?
School Leaders	How does the academic progress of our English learner subgroup compare with EL progress in other schools with similar populations?

from the fact that different students take the test each year. This is especially true in schools with high student mobility. A statewide data system that links student records over time offers a way to compare scores for only those students who are present in the school over several years, which provides a truer measure of that school’s effectiveness.

Longitudinal, student-level data systems can also enhance researchers’ abilities to isolate the impact of a particular school or policy on student learning. An analysis that follows students over time and across various classrooms, schools, or districts, for example, makes it possible to determine how much of their progress can be attributed to a specific school experience. Local educators can use the same kind of data to evaluate their own programs and instructional strategies, including being able to follow students from when they enter school until they leave.

The benefits of longitudinal data systems have received additional attention with the implementation of the No Child Left Behind (NCLB) Act. The federal law requires states to report disaggregated data on student achievement and graduation rates in order to increase data transparency. These data are collected most accurately with electronic, longitudinal data systems, and California has been building such a system to comply with NCLB.

The Data Quality Campaign articulates a national vision

The Data Quality Campaign (DQC) is a national effort to encourage and support state policymakers in improving the collection, availability, and use of high-quality education data. DQC also helps states implement their longitudinal data systems in order to improve student achievement. The National Center for Educational Achievement manages the privately funded DQC, and the list of its founding organizations reads like a “who’s who” of national leaders in education. DQC’s founding organizations include Achieve Inc., Alliance for Excellent Education, Council of Chief State School

The Data Quality Campaign’s 10 essential elements

Implementation planned in California?	Elements of a longitudinal data system
Yes	1. A unique, statewide student identifier that connects student data across key databases over time. This is a single, nonduplicated number that is assigned to a student throughout his or her education career.
Yes	2. Student-level enrollment, demographic, and program participation information. This information is important for knowing which types of students (by ethnicity, gender, poverty status, etc.) are in certain programs (such as Special Education, GATE, National School Lunch Program, etc.) and for determining the effects of those programs.
Partially	3. The ability to match individual students’ test records from year to year to measure academic growth. Individual student performance on state exams should be maintained with the ability to disaggregate the results by test question and content standard.
Yes	4. Information on untested students and the reasons they were not tested. States need to find out why students are not tested and then match those records to separate enrollment and program participation databases in order to identify patterns associated with specific student populations.
Partially	5. A teacher-identifier system with the ability to match teachers to students. Matching teachers to students by classroom and subject is critical to understanding the connection between teacher training and qualifications and student academic growth.
Yes	6. Student-level K-12 transcript information, including information on courses completed and grades earned. With this information, researchers can monitor the impact of rigorous course-taking on preparation for postsecondary education and the job market.
No	7. Student-level college readiness test scores. Student performance on exams, such as the SAT college admissions test and end-of-course advanced placement (AP) exams, are important indicators of students’ college readiness.
Yes	8. Student-level graduation and dropout data. In order to calculate the four-year cohort graduation rates defined in the new National Governors Association compact, states need to be able to follow individual students over time.
No	9. The ability to match student records between the pre-K-12* and higher education systems. States and school systems need better data on student success when they leave the P-12 system and enter college.
Yes	10. A state data audit system assessing data quality, validity, and reliability. Valid and reliable reporting requires checks on the accuracy and quality of the data submitted so that the public can have confidence in the quality of the information coming out of the state’s public education system.

* The DQC includes preschool in its discussion of K-12 data systems. Currently, California’s system does not include preschool data. However, with the recent passage of Senate Bill 1298 (discussed on page 9), this may change.

Officers, The Education Trust, National Center for Educational Accountability, National Center for Higher Education Management Systems, National Governors Association's Center for Best Practices, Schools Interoperability Framework Association, Standard & Poor's School Evaluation Services, and State Higher Education Executive Officers. The campaign is funded by the Bill & Melinda Gates Foundation, the Casey Family Programs, and the Lumina Foundation for Education.

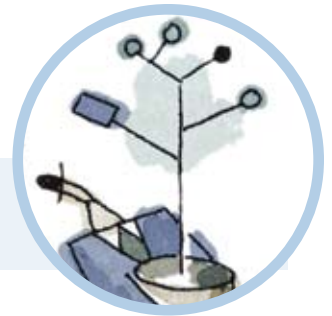
As a centerpiece of its work, the DQC has established a list of "10 essential elements" for high-quality statewide education data systems. (See the table on page 3.) These elements have received widespread approval by advocates for longitudinal data systems. They

also align with many of the data elements required for compliance with NCLB.

Proponents say that with these elements, a data system can provide the core information needed to support improved student achievement. However, they add that the 10 elements are essential but not sufficient for certain analyses. DQC says that states should also plan to collect financial information at the school and program levels that can be linked to individual student achievement.

The DQC elements provide a valuable lens for examining the status of California's emerging data systems as they are currently envisioned. At the center of California's approach are two key efforts: the California Longitudinal Pupil Achievement Data System

(CALPADS) and the California Longitudinal Teacher Integrated Data Education System (CALTIDES). When both are completed, the state will be fully implementing six of the 10 essential elements the DQC identifies. In addition, California has plans to implement some aspects of two other elements. However, the state has no immediate plan for implementing two of the DQC's essential elements. The rest of this report explores California's emerging data systems in the context of these 10 elements and other features.



California's data system is a work in progress

Many stakeholders concerned with California's public education system have been campaigning for a comprehensive data system for years, but they have had trouble securing adequate state funding to develop it. When the federal No Child Left Behind Act (NCLB) required states to assess individual student achievement over time and determine more accurate graduation rates, California became more serious about building a statewide system that could generate such information. To fully comply with NCLB accountability requirements, California had to be able to monitor individual student enrollment history and achievement data over time, which it could not do with its existing system.

Currently, CDE has 128 active data collections that it uses to populate 109 different databases. These multiple, separate data collections represent a substantial burden for the local education agencies (LEAs) responsible for submitting the data. Districts must keep track of all of these data submissions, which have differing deadlines throughout the year. Moreover, each collection is separately specified and reported. This lack of coherence also hinders the accessibility and transparency of the data for users—whether they are parents, teachers, administrators, researchers, or state policymakers.

Once CALPADS and CALTIDES, the state's emerging data systems, are fully

developed, California's method of collecting education data will change substantially.

CALPADS represents a major shift in California's approach

State lawmakers first authorized CALPADS in 2002 when they enacted Senate Bill (SB) 1453, which approved the creation of a database to house scores from multiple state tests for every pupil over time as the pupil moves up in grades and changes schools. This database for the K–12 education system makes it possible to monitor student progress on the California Student Testing and Reporting (STAR) program tests, the California High School Exit Exam (CAHSEE), and the

California English Language Development Test (CELDT).

SB 1453 was ambitious in dictating that the system should be used to accomplish four goals:

- Provide school districts and CDE access to data necessary to meet NCLB's reporting requirements;
- Offer a better means of evaluating education progress and investments over time;
- Give LEAs information that can be used to improve pupil achievement; and
- Provide an efficient, flexible, and secure means of maintaining longitudinal, statewide, pupil-level data.

After years of planning and development, CALPADS is close to becoming a reality. The state signed a three-year contract with IBM in December 2007 to develop the infrastructure and user interface for the CALPADS system. This work is currently in progress, and state officials expect full implementation in 2009–10.

However, the current work plan and project scope will not fully deliver the data capability described by the Data Quality Campaign or as originally envisioned by state lawmakers. A great amount of work could be done to further CALPADS's capacity. Without certain features and data elements, CALPADS's ability to provide sophisticated information to policymakers will be limited, as will the opportunity for educators and researchers to use the information at the local level.

California has assigned unique student identifiers—a crucial first step

As part of authorizing CALPADS, SB 1453 changed the state's model of collecting data from institutional-level to individual-level. The bill called for the assignment of a unique number for identifying every K–12 student enrolled in a California public school. Assigning an individual, yet nonpersonally identifiable, number to each student is a necessary first step for tracking students in a longitudinal system. California School Information Services (CSIS) oversaw the completion of this task and is also in charge of the annual student and educator data collection process.

The transition to student-level data submission was the first and most crucial step in moving toward the kind of system envisioned by state and federal leaders. With the help of CSIS, the state has now assigned to each student a unique number that “stays” with the student throughout his or her K–12 education experience in the state. These numbers are called statewide student identifiers (SSIDs). They will make it possible to link each student's information longitudinally but cannot be directly linked to students' names or any other personally identifiable information in the system. SSID numbers allow student-level data to be fully integrated into

Student data system: Progress and timeline for completion

	Date	Task
✓	1997	CSIS authorized to develop a voluntary electronic statewide school information system, which included individual student identifiers.
✓	September 2002	Senate Bill 1453 signed into law, authorizing CALPADS.
✓	June 2005	CSIS assigned statewide student identifiers (SSIDs) to all students.
✓	Fall 2006	Local education agencies (LEAs) used SSIDs to submit enrollment information to the state.
✓	Fall 2007	LEAs used SSIDs to submit information on graduates and dropouts to the state.
✓	December 2007	State awarded contract to IBM; CALPADS development began.
	2008–09	IBM completes CALPADS development and pilot testing begins.
	2009–10	California implements CALPADS statewide.

California School Information Services (CSIS)

The state launched CSIS in 1997 to establish an electronic data collection system that would make student- and teacher-level reporting more streamlined. CSIS works with local education agencies (LEAs) to build their capacity to collect, maintain, report, and use individual-level data. This work happens on a voluntary basis, so only a portion of districts and county offices participate. The agency also helps districts submit reports electronically to the California Department of Education and exchange student transcripts with other LEAs and postsecondary institutions.

In 2002, the state charged CSIS with working with LEAs to assign unique student identifiers to all students in the state. This task was completed in 2005.

CALPADS, which is the system that will maintain K–12 data at the state level.

Preliminary student-level data have already shed new light on important questions. For example, in the past, CDE only reported how many students in each grade in a given year took the California Standards Test (CST) in Algebra I and how they performed. In August 2008, CDE used student-level data to report how many students took the Algebra I CST for the first time, how many were repeat test-takers, and how the performance of these two groups differed. This information could have major policy implications given recent actions by the State Board of Education effectively requiring that the Algebra I test be the math CST for all 8th graders.

In the fall of 2006, LEAs for the first time submitted their enrollment data to the state using individual student records rather

than aggregate counts. CDE used this data to determine official enrollment counts for 2006–07. By the next fall, the state used SSID data to count 2007–08 enrollments as well as graduates and dropouts from the 2006–07 school year, resulting in more accurate data on dropouts. CDE released that official data in July 2008.

As part of their SSID data submissions, LEAs also include program participation and other NCLB-required information. However, CDE has yet to release official reports using this data. CSIS manages the annual submission of student-level data, which will eventually populate CALPADS when it is fully launched in 2009–10. In preparation, CSIS is working with local districts to improve the accuracy and reliability of this data collection. (For more details, see page 10.)

Mandated data collections are costly to reimburse

When California voters approved Proposition 4 in 1979, they enshrined in the state constitution a guarantee that the state would reimburse local governments, such as school districts and county offices of education, for the cost of implementing any new state-required program or higher level of service. In California's very constrained fiscal climate, school districts have actively pursued reimbursement for any and all state mandates. State officials have, in turn, been careful to avoid any new requirements that can be interpreted as a mandate.

However, the mandate reimbursement requirement does not apply if California merely codifies a federal requirement. At least partly for this reason, state officials, particularly the Department of Finance, have sought to limit the data elements in CALPADS to what is required in the federal No Child Left Behind Act, rather than adding elements that could be construed as a state mandate.

CALPADS is limited to data required for NCLB compliance

Despite its broad authorizing language, the actual scope of CALPADS is expected to fall short of what was envisioned under SB 1453, due in part to costs associated with mandates. State law requires the Department of Finance (DOF) to approve an expenditure plan for the project. According to a 2006 report by Janet Hansen (vice president and director of education studies with the Committee for Economic Development), the DOF has used its leverage to keep the project focused on narrow compliance with NCLB's requirements as a cost-saving measure. Requiring LEAs to provide additional information would impose greater costs on the districts that the state is not prepared to reimburse. Hansen also mentioned policymakers' concerns about data management and governance over the system as contributing to the data system's narrow scope.

Therefore, the CALPADS collection is mainly limited to those data elements specifically required by NCLB. Additional data elements that were previously included in the state's annual collection of basic student and staff data known as CBEDS (California Basic Educational Data System) will also be included in CALPADS. This includes staffing data on teacher demographics, type of assignment, and more. Lastly, course completion data will be included in order to meet reporting requirements for the federal Perkins grant.

The data elements that will make up CALPADS correspond to the essential elements of a longitudinal data system as articulated by the DQC. These include:

- Student-level demographic data
- Program participation data
- Student discipline
- Enrollment status
- Course-taking data
- Student assessment data

(For a more comprehensive list of the data that will be included in CALPADS, please visit the EdSource website: www.edsource.org/iss_datasy.html)

Two important elements are on the "to do" list

In addition to what is being done, California has plans to partially implement two more of the DQC's essential elements—matching student test scores from year to year and creating a teacher identifier system. But it remains unclear if and how these data will be used beyond meeting compliance requirements for NCLB. Although CALPADS will match individual student test scores from year to year, the method used to score the tests prevents using them as a measure of academic growth. The state also plans to implement an educator-level data system that can be linked to CALPADS. This system, CALTIDES, will match teachers to students longitudinally in order to study the effect of teacher preparation and other programs on student achievement.

Measuring individual students' academic progress

When CALPADS is fully implemented in 2009–10, the state will be able to track individual student test scores over time. However, the design of California's standards and assessment system does not allow the use of simple year-to-year changes in student test scores as measures of academic growth. This has prompted discussions of how progress will be measured using a longitudinal data system.

The state's academic expectations generally grow in complexity as students go through the grade levels. However, the rigor of the curriculum does not increase in a perfectly straight line. The state's content standards do not become, for example, 30% more difficult in each successive grade, and neither do performance goals on standardized tests. The California Standards Tests (CSTs) are not scored on a continuous scale from grades 2 through 11. All of this means that the state cannot use simple score differences over time to measure individual student growth. A student does not achieve, for example, a score of 250 on her second-grade math test and show "10 points of growth" by achieving a 260 on her third-grade math test the following year.

Despite this limitation, state leaders are hoping to find a way to examine academic growth with the existing test results. In the 2007–08 Budget Act, the Legislature set aside funding for the state's standardized testing contractor, Educational Testing Service (ETS), to explore options for measuring growth.

ETS evaluated various alternatives before recommending that the state match students' scores from one grade to the next and develop tables showing how, for example, students with a given score on a grade 3 test usually score on the grade 4 test. Policymakers could then look at an individual student's score in one year and use the tables to see whether she was on target to reach a certain score the following year. Officials could also look at test scores of a student and tell whether he had made average, above average, or below average progress. Programs could be evaluated based on the growth of participating students as compared with the growth in achievement among students who did not participate.

ETS did not specify dollar figures when they talked about the cost of various approaches to measuring growth, but they did discuss cost-related factors to consider. For example, with regard to the recommended approach, ETS says that the state would not have to collect additional data but would need to do additional straightforward analyses to produce the tables they proposed, which would require more resources. Furthermore, ETS suggests that the state would need to update such analyses each year. In addition, ETS assumes that the state would conduct focus groups to learn how best to communicate to educators and parents about the use and meaning of growth scores.

Tracking teacher characteristics over time

The state is also developing CALTIDES, a system to track teacher characteristics over time. The legislation authorizing the development of this system (SB 1614, passed in 2006) calls on the Commission on Teacher Credentialing (CTC) to assign Statewide Educator Identifiers (SEIDs) to all educators working in the K–12 public school system in a position that requires a credential or authorization granted by the CTC. This includes teachers, school principals, most district administrators, counselors, and librarians. CTC and CDE will jointly develop CALTIDES.

The first goal of CALTIDES is the same as that of CALPADS: to satisfy federal NCLB and state reporting requirements. It will also help meet requirements to monitor teacher assignments and conduct program evaluations. CALTIDES will primarily integrate two existing databases—CALPADS, which collects staffing and student information, and the teacher credential and authorization data collected and managed by CTC—to pull together longitudinal educator data. CALTIDES will also include information on internship and beginning teacher support programs.

CALPADS and CALTIDES data will be linked using SEIDs. CALTIDES will allow research on the effectiveness of teacher training programs as well as teacher retention and mobility. The state will also be able to study the training and experience of teachers in

Educator data system: progress and timeline for completion

	Date	Task
✓	July 2005	Legislature called for a feasibility study report to build an integrated teacher data system.
✓	September 2006	Senate Bill 1614 was signed into law authorizing the educator data system, CALTIDES.
✓	May 2008	The Commission on Teacher Credentialing (CTC) assigned statewide educator identifiers (SEIDs) to all certificated employees.
✓	June 2008	The California Department of Education released a request for proposal (RFP) to develop CALTIDES.
	February 2009	Final bids in response to RFP approved.
	January 2010	Contract for CALTIDES development begins.
	2011–12	California implements CALTIDES statewide.

hard-to-staff areas, such as math, science, and Special Education. SB 1614 prohibits the use of data in CALTIDES or CALPADS “for the purposes of pay, promotion, sanction, or personnel evaluation of an individual teacher, or groups of teachers, or any other employment decisions related to individual teachers.” Districts with sophisticated data systems can already do this if they and their teachers union agree that it is appropriate.

CALTIDES will not include information on classified employees—those individuals who do not hold a credential and are therefore not tracked by CTC. This includes teaching assistants, teacher aides, pupil services aides, library aides, office/clerk staff, custodians, bus drivers, and cafeteria workers. Further, current data collections (geared toward compliance with NCLB) are incomplete for some credentialed employees. Even with a longitudinal system in place, it will not be possible to answer critical questions about administrators, such as the length of a principal’s tenure at a school site or the frequency of superintendent turnover within a district.

CDE released a request for proposals (RFP) for a vendor to develop CALTIDES in June 2008, and final bids are projected to be approved in February 2009. The contract is scheduled to begin in January 2010. While vendors are developing their proposals, CDE is applying to the U.S. Department of

Education’s Institute of Education Sciences (IES) to provide funding to develop CALTIDES, as it did for CALPADS.

Implementation of several important features is further down the road

California currently has no plans to implement several features important to a robust data system as part of either CALPADS or CALTIDES. The system will not include data on students’ college readiness results, specific characteristics of students with special needs, or school site-level financial data. Additionally, CALPADS will not be immediately linked with the data systems in education sectors outside of K–12, though recently adopted legislation makes progress toward that goal.

Additional K–12 data elements are still in the future

For reasons explained in the preceding section, CALPADS will primarily be limited to those data elements required for compliance with NCLB. However, many stakeholders consider these K–12 data elements incomplete, leaving many important research and evaluation questions unanswerable. For example, the state has no plans for LEAs to submit students’ college readiness test results (such as SAT and AP scores). This information could show how student performance at the high school level predicts success in college.

Florida's education data system is particularly comprehensive

Many education stakeholders see Florida's data system as a national model. It not only has all 10 of the Data Quality Campaign's essential elements, but it also goes further. The state's Education Data Warehouse contains a wide array of information on schools, staff, and students. With information from multiple public education and social services agencies under one roof, Florida's data system can provide information that is a long way off for California.

Florida built its data warehouse from 2001 to 2003, but filled it with data going back to 1995. The database contains information on students' assessment results, course-taking, and demographics. Besides K-12 data, Florida tracks information on students enrolled in pre-kindergarten programs, adult education, technical schools, community colleges, and universities. These education data are linked to students' financial aid packages, the social services they receive, and their employment records.

The fact that Florida has a single governance and oversight structure for multiple agencies has greatly facilitated the building of a data system that works across all of the state's education segments. Data systems are linked so they can "talk to" each other. For example, the prekindergarten data system can be linked with the community college data much more easily than would be the case in California.

Many important research questions can be answered in Florida that would be impossible in California and most other states. For example, with Florida's data system, researchers can examine the long-term effectiveness of a school intervention program, showing whether students who participated in the program have lower dropout rates or greater success in college.

Being a pioneer creates certain limitations, however. For example, Florida collects data from local education agencies via databases built in the late 1980s. This has presented obstacles to web-based, online access for users statewide. However, the state has recently deployed a web-based system called Sunshine Connections that allows school leaders, teachers, and parents to access a wealth of student data and tools for improving instruction and learning—though some data, such as students' test results, are restricted. Another issue is that the data infrastructure uses Social Security numbers as individual identifiers, which presents challenges to preserving student and teacher privacy.

It will also take time for the state to incorporate data elements from other student-level databases within CDE into CALPADS. Other than program participation counts, data from the California Special Education Management Information System (CASEMIS) and the Migrant Student Information Network (MSIN) will not be integrated into CALPADS immediately. However, these two data systems will use SSIDs to collect data on Special Education and migrant students and their teachers, making it possible to link CASEMIS and MSIN data with CALPADS in order to analyze, for example, how student achievement is affected by different learning disabilities or by how often migrant students change schools.

Additionally, state officials have no plans to incorporate financial data into CALPADS

or CALTIDES. Linking students' performance with the resources available to their schools would clarify the ways in which those resources relate to student performance. Such data would also shed light on which groups of students benefit from increased spending in their schools, another question that cannot be answered by California's current statewide data system.

A key barrier interferes with the integration of school-level financial information in a data system. Currently, LEAs submit their data through a format called SACS (Standardized Account Code Structure). Although these data are compiled at the district level, the SACS structure was designed to also accommodate site-level financial data. Districts thus have the ability to report accounting data disaggregated down to the

school level, but they are not required to use the school-site account codes. Because it is optional, the availability of school-level data vary across districts and cannot be used to inform decision-making at the state level.

Whether school-level accounting should be required as part of SACS reporting—a necessary first step if such data are to be incorporated into a data system—is up for debate. Groups such as the Governor's Committee on Education Excellence have recommended that the state require the reporting of school accounting data rather than just district averages. They say that this information can inform the way the state makes investments in schools. On the other hand, Janet Hansen argues in her 2006 paper on education data in California that the benefits of obtaining this type of data may not be enough to justify the burden this data collection would place on LEAs and schools.

Proposals include linkages with other data systems and agencies

Several proposals call for the new student-level data system to go beyond the multitude of K-12 indicators and link to other data sets and agencies that offer telling data related to student achievement. These include data that follow children who participate in preschool and public health and welfare programs as well as data on young adults transitioning out of school into college and the workforce. Linking student data from preschool through higher education, as well as with other agencies, would provide a seamless network of information that proponents believe would shed light on both policy and practice related to how the state is serving its young people.

Currently, 18 states link their K-12 and higher education systems and some have also linked to the systems of other government agencies, according to a survey by the National Center for Educational Accountability (NCEA). Several advocacy groups, such as Children Now and the Education Trust-West, are calling for California to make this move as well, creating linkages among data sets from multiple education segments in order to have a truly comprehensive set of data on the state's students.

Their goal is to link the K–12 system with data on enrollments in early childhood education and postsecondary programs, such as adult education and technical training programs. An interagency-linked system would require building the appropriate systems and implementing the proper technological infrastructure to link the databases at CDE, the Health and Human Services Agency, the Employment Development Department, and others.

A major report by RAND on building a K–20 (kindergarten through graduate school) data system in California recommends developing such a system incrementally, working with the three public postsecondary systems in the state (community colleges, California State University, and University of California) to ensure that a common student identifier is used and that data definitions are

consistent. This would allow research on college readiness, relationships between high school coursework and success in college, and the effectiveness of college preparation programs. In Florida, these types of linkages have proved to yield a great deal of information, but they are far from implementation in California. One major obstacle here is that each of the systems operates under a different governance structure.

Some members of the education and research communities are interested in linking student data with teacher data. A report from the American Institutes for Research's District Practitioner Working Group says that this would enable the state to determine how training programs help improve teacher quality, ultimately improving student progress. The state would link individual teachers to individual students and track those

teachers, measuring how their students' achievement had changed over time.

Senate Bill 1298 (the Education Data and Information Act of 2008), recently signed by Gov. Arnold Schwarzenegger, makes substantial progress toward linking some of these data systems. Authored by Sen. Joe Simitian and Sen. Darrell Steinberg, the bill requires state- and federally-funded, center-based child care and child development programs, as well as all three segments of California's public higher education system, maintain SSID numbers for their students and use them to report data to the state. The state chief information officer would develop a plan to link these education data systems from pre-K through higher education.



Roadblocks to fulfilling a broader vision

CALPADS lays a solid foundation for longitudinal education data collection and its use in California. But the scope of CALPADS—the number of data elements it reports, the tools available to analyze and present the data, and the accessibility of the data—is less robust than many had originally hoped.

The state has no plans to integrate financial data into the database, which will mean that CALPADS itself will not help answer questions about the relationship between spending and achievement. Data relating to specific student populations, such as foster youth, will not be integrated or linked to the system. For the most part, CALPADS will not offer the public user-defined data queries, generally providing only prepared reports. However, these reports will utilize student-level information that far surpasses the scope and accuracy of data previously available to the public. (But LEAs will have more options for how to query and download their own data.)

The limited scope of CALPADS can be attributed to several factors. Political will on the part of the state government has been inconsistent, stalling the full implementation of a student-level data system for years.

Thus, the appropriation of state funding has been minimal. In the midst of challenging state budget deliberations in 2006, California received a three-year grant totaling \$3.26 million in 2006 from the U.S. Department of Education's Institute of Education Sciences to fund 50% of the costs of building the state's longitudinal data system.

California has continued to depend on federal funds to develop its data systems. In 2007–08 alone, federal funds spent on the project amounted to nearly three times the funding appropriated by the state itself, which was less than \$1 million out of the total \$3.9 million spent on the development of CALPADS that year. That is a minute fraction of state spending on K–12 education, which totaled more than \$50 billion in 2007–08, according to a May 2008 report from the Legislative Analyst's Office.

Several other obstacles hinder the effective implementation of CALPADS, CALTIDES, and any other data systems the state might develop. One particularly difficult roadblock is the limited capacity of local school district officials to compile accurate data. A data system is only as good as the quality of the data it reports, which depends on local officials' capacity to enter data accurately. Another notable challenge is privacy regulations that a state data system must adhere to because transforming data into useful information requires state officials, educators, and researchers having access to that data.

Local education agencies have limited capacity to ensure high-quality data

In order for a statewide student information system to be useful, the data entered into the system must be accurate. Unfortunately, the

data management capacity of California's nearly 1,000 LEAs varies substantially, which can result in data of inconsistent quality that impede informed decision-making.

The transition to individual data reporting for the new statewide systems relies on regular data collection and correction at the local level. This is a new expectation for the staff in many districts that requires staff training and time, and therefore, money. A 2008 policy brief from Policy Analysis for California Education (PACE) elaborates on this point, saying that the effective implementation of CALPADS and CALTIDES relies upon sufficient funding for the necessary training and infrastructure for local officials. A number of California districts have a head start on this, thanks in many cases to their participation in CSIS since 1997. Districts such as Long Beach Unified and Garden Grove Unified are well known throughout the state for their robust, comprehensive districtwide data systems. But some of the state's smaller districts did not even have electronic data systems in place when they were required to assign unique identifiers to their students in 2005.

Advocacy groups such as the Education Trust–West argue that LEAs may not have an incentive to report accurate, reliable data unless it affects their resources, such as enrollment and attendance numbers that determine per-pupil funding from the state. They also suggest that the state should help LEAs make use of their data so they get something back from their investment in data collection and reporting. With training and professional development for local officials, these groups say, LEAs would be better able to submit quality data to the state. In turn, the state could provide analyses of local data and reports that districts need to inform their decision-making.

State policymakers have provided funds for local capacity building, but the investment has consisted primarily of one-time, as opposed to ongoing, monies. In addition to providing funding to districts participating in CSIS's Best Practices Cohort, policymakers have, to date, appropriated about \$37 million to help non-CSIS districts with the start-up costs associated with participating in

California's current efforts to build local capacity

Part of the mission of the California School Information Services (CSIS) is to build the capacity of local education agencies (LEAs) to maintain quality information systems and provide the necessary professional development to do so. CSIS works with LEAs to help improve their local data management practices.

The Best Practices Cohort Project, run by CSIS, has made particular strides in this direction. Since 2006, it has been preparing volunteer LEAs for submitting student-level data to CALPADS and for using that data locally. CSIS provides training in meeting CALPADS submission requirements and data quality standards, and offers districts funding for local data management in exchange for completion of training and data submission.

In the 2008–09 state budget, policymakers appropriated almost \$8 million for the Best Practices Cohort Project, enough funding for all eligible LEAs in the state to participate if they so choose.

CALPADS. What has been largely lacking, according to CDE, is the commitment of significant ongoing funds, which are necessary to maintain accurate data over time. The state provides non-CSIS districts with \$0.25 per enrolled student to maintain the data; a figure that CDE asserts falls far short of what is needed.

In their contribution to the *Getting Down to Facts* research project in 2006, researchers at Springboard Schools note that without good quality data, schools and districts cannot make informed decisions. They suggest establishing a best-practices clearinghouse that would identify districts that are using data well and propose guidelines that local leaders could follow. They also recommend the implementation of a web-based training tool on how to use the newly emerging student-level information, building on CSIS's efforts with the Best Practices Cohort. (See the box on this page.)

In response to this and other research, the California Budget Project (CBP) has called on the state to invest resources at the local level so that teachers and administrators at both the school and district levels understand how to use data to inform instructional practices and program effectiveness. If data cannot be easily accessed and understood at the local level, CBP says, educators will not use it, thereby compromising critical reform efforts.

Legislation recently signed by the governor, Assembly Bill (AB) 2391, is aimed at building educators' capacity to analyze and use data in the classroom. The bill builds

upon the existing Math and Reading Professional Development Program, which provides incentive funding to LEAs for their teachers to participate in 40 hours of specified professional development and 80 hours of follow-up training. AB 2391 authorizes teachers to spend up to half of their 80 follow-up hours in areas that include data analysis, the implication of data analysis and its effect on increasing pupil achievement, and statewide and local data management systems, among other options.

New data systems heighten tensions between student privacy and access to data

Given that one of the key goals of a statewide longitudinal data system is to evaluate student progress and the effectiveness of specific programs, many experts argue that researchers must have access to student-level data. However, balancing this goal with the need to maintain student privacy—in compliance with federal and state law—is challenging.

Based on current laws and practices, researchers and evaluators will have only limited access to statewide information collected by CALPADS and CALTIDES when the systems are fully in place. Federal laws, such as the federal Family Educational Rights and Privacy Act (FERPA), impose limits on the disclosure of student records (without parental permission) by educational institutions that receive funds from the U.S. Department of Education. California state laws affecting access to student data generally mirror federal policies, according to the Legislative Analyst's Office.

The way state education agencies comply with FERPA and other privacy laws can vary. CDE has historically been conservative in its interpretation of FERPA, carefully balancing the right to personal privacy guaranteed in the state's constitution since 1972 with providing access to data. Under CDE's interpretation of FERPA, only the LEA or school—not the state—can authorize the disclosure of student-level information. For example, if a district submits student-level data to the state, CDE may not disseminate that LEA's data except in the form of aggregated or nonpersonally identifiable reports, such as those currently found on CDE's website.

Legal experts from the Data Quality Campaign have highlighted gray areas that would help state officials to “maximize the power of education data while ensuring compliance” with FERPA. For example, FERPA restrictions on data disclosure only apply to personally identifiable information on students. CDE is able to “scrub” the data so that a student's identity cannot be deciphered, and then the data may be disclosed. However, it is nearly impossible to effectively scrub all school-level data. For example, if there are

only one or two Native American children enrolled in a school or a district, it would be simple to determine the identity of the child in the data set.

FERPA does authorize disclosure of student-level data for studies aimed at improving instruction under strict conditions, such as research related to implementing district and school accountability measures or assessing graduation data. The state may grant access without parental permission to researchers conducting such studies as long as the data are determined to be nonpersonally identifiable.

Information on privacy and data systems

For more information and resources on the federal Family Educational Rights and Privacy Act (FERPA) and its implications for state longitudinal data systems and student privacy, visit the Data Quality Campaign's website: www.dataqualitycampaign.org/policy_implication/ferpa.cfm

The full text of the FERPA regulations can be found on the U.S. Department of Education's website: www.ed.gov/policy/gen/reg/ferpa/index.html

For more analysis on this issue, see the California Legislative Analyst's Office report, *Redefining Student Data Access Policy* (2008): www.lao.ca.gov

Federal policymakers are currently considering regulations that would revise some of those conditions, potentially improving researchers' ability to access the data. Meanwhile, Senate Bill 1298, signed into law by the governor, authorizes the state chief information officer—a member of the Governor's Cabinet who works to improve the state's technology infrastructure—to release nonpersonally identifiable student data to eligible parties.



How to get it done

When it comes to the visions, policies, priorities, and technical guidance for the development of a comprehensive data system in California, there are many areas of agreement among educators, researchers, and policymakers.

The need to continue current work to implement CALPADS and CALTIDES—and provide sufficient funding to do so—is a consistent conclusion. Laying a firm foundation and supporting LEAs through the transition are important first steps. Incorporating more data elements into the existing framework is also frequently recommended.

However, many nuts and bolts issues continue to challenge policymakers. They have found it difficult to prioritize expanding a data system and building local capacity in light of limited state resources. When the tradeoff is not spending that money on

instruction in the classroom, investing in improving data collection can be a difficult budgetary decision.

In her 2006 report on data systems prepared for the *Getting Down to Facts* research project, Janet Hansen identified California's lack of leadership—having no data system “champion”—at the state level as a contributor to the slow growth of the state's data system. She credits Florida's “long-time legislative emphasis on making data-driven decisions and...legislative mandates requiring (and funding) the creation of robust systems” for that state's successful data system. In

order to build support for a state data system that encourages data-driven decision-making in classrooms, district offices, and Sacramento, Hansen said California needed leaders who would navigate the political challenges and acquire the support and funding necessary for a more comprehensive system.

California has made progress during the past several years. State policymakers have sponsored pieces of legislation to advance the development and usability of CALPADS and CALTIDES. A number of nonprofit and advocacy groups have taken on this issue to raise its profile and build support. The public

release of 2006–07 dropout rates—the first release of student-level data—made it clear how such data can provide new insights. It also motivated the media to tell interesting stories about the state of California’s public schools. But that experience also raised questions about accurate interpretations of information.

Policymakers are also asking to learn more. “Improve data systems and use them wisely” was one of the four priorities of the Governor’s Committee on Education Excellence. Schwarzenegger released the committee’s recommendations on the topic in spring 2008.


Additionally, the governor and Superintendent of Public Instruction Jack O’Connell commissioned a report from consulting firm McKinsey and Company, which is expected to be released in the next few months. This forthcoming report will provide a long-term implementation roadmap for how California can transform the vision of a highly productive data system into a reality. The recommendations could help guide the state’s construction of the technological infrastructure necessary to support the ongoing development of these systems, with an eye to their future potential. The work is paid for by a grant from the Bill & Melinda Gates and the William and Flora Hewlett foundations.

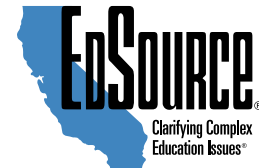
McKinsey’s report is consistent with the vision articulated by the DQC and described in this publication. Its goals include:

- Increasing the quality, accessibility, and basic use of current data;
- Expanding the use of the data; and
- Creating linkages with agency data systems outside of K–12 education.

The consulting firm will present its recommendations as an incremental three-step process, representing the system’s changing needs as it develops over time and incorporating factors such as available funding, evolving best practices, and changing needs. These steps are all based upon the foundation, and continued progress, of CALPADS and CALTIDES.

The report reflects the input of 200 individuals representing more than 100 organizations (including schools and districts) throughout the state, according to its authors. They conducted interviews with students, parents, community-based organizations, classroom teachers, school and district leaders and staff, state policymakers, researchers, and employers.

Perhaps as CALPADS continues to be implemented and its preliminary benefits realized, the state will solidify a “culture of data” that will support a broader and more robust system of education data in California. 



Trish Williams
EdSource Executive Director

2008–09 EdSource Board of Directors

Davis Campbell, President
President, California School Boards Association Governance Institute

Lawrence O. Picus, Vice President
Professor, Rossier School of Education, University of Southern California

Martha Kanter, Fiscal Officer
Chancellor, Foothill–De Anza Community College District

John B. Mockler, Secretary
President, John Mockler & Associates, Inc.

Susan K. Burr
Executive Director, California County Superintendents Educational Services Association

Carl A. Cohn
Distinguished Leader in Residence, College of Education, San Diego State University

Christopher Cross
Chair and CEO, Cross & Joftus, LLC

Kenneth F. Hall
Executive in Residence, University of Southern California

Reed Hastings
CEO, Netflix, Inc.

Gerald C. Hayward
Co-Director (Retired), Policy Analysis for California Education

Janis R. Hirohama
President, League of Women Voters of California

Santiago Jackson
Director, Office of Legislative and Governmental Affairs, Los Angeles Unified School District

Kelvin K. Lee
Superintendent (Retired), Dry Creek Joint Elementary School District

Jo A.S. Loss
President-Elect, California State PTA

Paul J. Markowitz
Teacher (Retired), Las Virgenes Unified School District

Amado M. Padilla
Professor, School of Education, Stanford University

Don Shalvey
CEO and Co-Founder, Aspire Public Schools

Gloria Taylor
Co-President, American Association of University Women–California

To Learn More

- For a list of the works cited in this report, go to: www.edsource.org/pub_datasys10-08.html
- For ongoing coverage of California’s development of CALPADS and CALTIDES and for a comprehensive list of the data elements to be included in those systems, go to: www.edsource.org/iss_datasys.html
- To review past and present bills related to this topic, go to: www.leginfo.ca.gov

Acknowledgments

This report was researched and written by:

Heather Barondess

With research support from:

Brian Edwards

Edited by:

Mary Perry

EdSource thanks the **Bill & Melinda Gates Foundation**, the **William and Flora Hewlett Foundation**, and the **James Irvine Foundation** for their investment in our core work.