



This brief is the second in a series exploring issues related to the development of a comprehensive data system in California.

Making California's New Data System Work: Quality Is Key

HIGHLIGHTS

- Schools and districts face increased challenges as they make the changes necessary to implement a student-level data system 2
- CALPADS implementation is under way..... 3
- State officials are racing to meet a tight timeline 5
- Strategies for improving local capacity provide important guidance 6
- Other states are meeting data quality challenges in innovative ways.... 8
- Local data capacity is central to a larger vision for the state's data system 8

California's K–12 public school system is in the midst of a fundamental change in the way it collects and reports student data. Experts hope that this transformation will provide educators, policymakers, researchers, and the public with more robust and more accurate information with which to evaluate the progress of schools and their students. However, the challenges involved in implementing a new statewide data system are substantial. And ultimately, the quality of information will depend on the capacity of California's thousands of schools and school districts to accurately track and report data about each of the millions of students they collectively serve.

As 2009 begins, the state has taken major steps toward the implementation of the California Longitudinal Pupil Achievement Data System (CALPADS) by specifying precisely the student-level data that all schools and districts must begin submitting next school year. School districts have also been selected to test certain features of CALPADS, a process that will begin this spring.

Currently, school districts vary substantially in their readiness to collect, maintain,

and report these data reliably. They also vary in the extent to which they have made data quality a priority, due partly to the pressures they face and partly to the voluntary nature of state programs up to this point. This report explores the challenges everyone from the data clerk at the local elementary school to officials in the California Department of Education will face as they work to ensure the quality and accuracy of data used to build the state's new student information system.

EdSource thanks the Bill & Melinda Gates Foundation for its investment in our core work.



Collecting and maintaining data has been a long-standing challenge for schools and districts

Collecting basic data about an individual student presents many opportunities for human error. It begins when a parent registers his or her child for school and submits several handwritten forms to the school staff. Schools must also confirm enrollment of returning students and flag, or “exit,” students who have left the school. In most schools, a secretary or clerk enters the student data into an electronic student information system (SIS). Typically, data entry is one of many tasks for which he or she is responsible.

Local education agencies (LEAs), such as school districts and county offices of education, oversee this school-level data collection, ensuring its accuracy and submitting the data to the California Department of Education (CDE). CDE compares the data to data from the previous year and alerts districts to potential errors. Districts must then resolve anomalies in the data (if a student wasn’t properly “exited,” for example) before a statewide education data set can be completed.

For years, district officials have taken a census of their students on the first Wednesday in October, known as Information Day. On that day, a school’s enrollment counts were reported in groups that were aggregated by basic demographic information such as grade, gender, and ethnicity. The total number

of students enrolled on that day served as the “official” enrollment count for the year. However, the aggregate data were limited and inflexible. The school may have reported the number of first graders by gender and ethnicity. But if CDE wanted other information—say the number of first graders by gender and English learner status—CDE would have to request other reports from the school district. Furthermore, aggregate data did not allow the state to accurately track students’ enrollment or achievement over time because there was no way to be sure that the students included in the aggregate figures had not changed significantly.

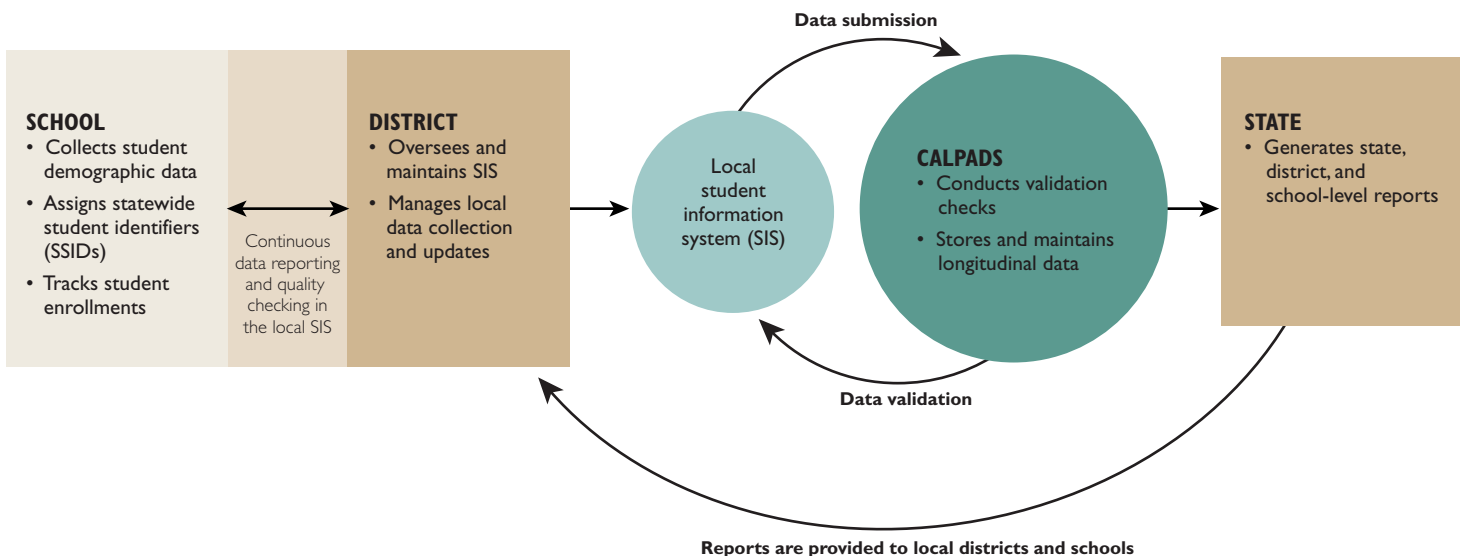
California shifts to student-level data collection

State policymakers have recognized the limitations of such aggregate reporting for years.

In 1997, the Legislature started providing funds to support the development of a California School Information Services (CSIS) system to streamline the collection, management, and reporting of data to the state and to facilitate the exchange of school data among LEAs. CSIS had begun in 1992 as a project of CDE and research organization WestEd to study the feasibility of electronically sharing student-level information. School districts collaborated in the effort; and after several years, CSIS was found to be a successful model for automating the exchange of school information, transferring student records, and reporting to the state.

Since its statewide implementation in 1997, CSIS has been working with volunteer school districts, assigning nonpersonally

California’s Student-Level Data Submission Process



identifiable numbers to their students and helping them maintain their student information in local databases and submit their individual records to CSIS. CSIS then aggregates the student-level data from participating districts into the aggregate state reports required of districts and submits them to CDE. The number of participating LEAs has grown to approximately 250, encompassing about 38% of the state's students.

In 2002, Senate Bill 1453 authorized the development of CALPADS and charged CSIS with assigning nonpersonally identifiable student numbers to all students in California's public schools. This work was completed in June 2005. Since 2006, CDE has been using the data based on these statewide student identifiers (SSIDs) to generate official enrollment counts. The SSID will remain with students throughout their academic career in the California public school system. The SSIDs are the foundation for CALPADS to be able to track individual student enrollment over time and to ultimately link enrollment and achievement data.

The way LEAs submit data to the state is changing

Although collecting and submitting data to the state has never been easy for districts, the assignment of SSIDs and the collection of certain data elements at the individual level instead of in the aggregate are likely to present new and intimidating challenges for many LEAs. It remains to be seen to what extent the benefits of a statewide student-level data system (e.g., longitudinal student data available to inform decision making, timely information on transferring students, fewer CDE requests for information) compensate for the costs of its implementation.

Student enrollment data are no longer just a snapshot of the students in a school on a single day. CALPADS is intended to be a "transactional" data system, in which enrollment changes are reported to the state on an ongoing basis. Other data, such as course enrollments or program participation, do not have to be transactional but are often reported more than once during a year.

This change will help California meet federal No Child Left Behind (NCLB) reporting requirements, including calculating

CALPADS implementation is under way

Spurred in part by federal requirements and funding provided in the No Child Left Behind Act (NCLB), California has been developing a comprehensive data system for many years. To fully comply with NCLB's accountability requirements, the system must be able to track students over time, including their enrollment status and achievement. Until recently, California's method of data collection did not do this. The state's emerging data system, California Longitudinal Pupil Achievement Data System (CALPADS), is based on data on individual students. When it is fully operational, it will likely enable the state to more accurately answer questions about the academic achievement of specific groups of students, including their progress over time. A complementary system, the California Longitudinal Teacher Integrated Data Education System (CALTIDES), will integrate existing databases to generate longitudinal information on educators, satisfying other federal and state reporting requirements.

In the past, data collection in California meant that school districts aggregated data about groups of students into summary reports. The implementation of CALPADS requires each local educational agency (LEA) to submit data to the state that reports on each individual student. Although California has made efforts to prepare LEAs to collect, maintain, and submit this type of data to the state on a regular basis, it is not clear whether all schools and districts will have the capacity to submit accurate and reliable student-level data to CALPADS when it launches in 2009-10.

For more information on CALPADS, see the October 2008 EdSource publication, *California's Emerging Data System: A Status Report* at www.edsource.org/pub_datasys10-08.html. It provides background information on CALPADS and student-level data and discusses the elements of a successful data system.

CALPADS Implementation: Scheduled Timeframes and Benchmarks

2008

✓	October	California Department of Education releases the draft "file formats" to LEAs, outlining the data to be collected by CALPADS.
---	----------------	--

2009

✓	January	LEAs that will test CALPADS' user acceptance are identified.
	May	LEAs begin testing features of the system.
	September	Existing statewide student identifiers (SSIDs) are switched to CALPADS.
	October	All LEAs submit official student enrollment, graduate/dropout, course enrollment, and teacher assignment data to CALPADS for the first time.
	December	CALPADS begins generating pre-ID files for LEAs to use with their test vendors in preparation for the 2010 State Testing and Reporting program (STAR).

2010

	March	All LEAs submit official immigrant counts and English learner status to CALPADS for the first time.
	June–September	All LEAs submit official course completion, student program participation, and student discipline data to CALPADS for the first time. (This end-of-year data submission may be optional in the first year.)

DATA: CALIFORNIA DEPARTMENT OF EDUCATION (CDE), 1/28/09

EDSOURCE 2/09

The initial student-level data release sheds light on dropout rates and data challenges

In July 2008, the California Department of Education (CDE) released the 2006–07 graduate and dropout data—two months later than expected. This data release was the culmination of a two-year process, beginning with local education agencies (LEAs) submitting individual student records to the state for its official enrollment counts in fall 2006. In fall 2007, enrollment counts were once again based upon individual records; and for the first time, the state was able to more accurately distinguish students who had dropped out of school from those who had re-enrolled elsewhere in the state. For students who graduated or did not re-enroll in their district that fall, LEAs were required to provide an “exit code” explaining why the student had left school.

However, this dropout data release presented challenges for both LEAs and CDE. Although LEAs were encouraged to provide an “exit code” in 2006—the pilot year for reporting graduate and dropout data based on individual student records—many did not. For the first time in 2007, official state graduate and dropout counts were based upon exit data reported for individual students. These data dictate the dropout rate, a statistic that is highly scrutinized by the public and by policymakers. Many LEAs did not devote sufficient time to verifying and entering exit reasons for students during the school year. As a result, they had great difficulty submitting and certifying the accuracy of their exit data.

In general, some districts do not sufficiently review their data before submission to the state and only realize the inaccuracies of their data once they are posted on the CDE website. In response, CDE allows LEAs to correct errors in their data for a month after the initial, preliminary data release. For 2006–07 dropout rates in particular, the re-released data resolved important inaccuracies in districts’ dropout rates. For example, Long Beach Unified originally reported 2,206 dropouts and lost transfers (students who said they were transferring to another school but never showed up) for 2006–07. After months of tracking down individual students to confirm their exit status, they revised their final estimates to 1,441 dropouts and lost transfers—a 35% reduction.

more accurate graduate and dropout rates. LEAs must follow individual students and update their data continuously, using the unique, nonpersonally identifiable SSID number assigned to the student when he or she enters school, typically in kindergarten. (The SSID is the key that will allow student demographic data to be merged with that student’s achievement information and other data when CALPADS is fully operational.) Schools and districts must account for the whereabouts of each student using this number, which allows students to be tracked when they change schools, graduate, or leave the public school system.

Maintaining unique student identifiers represents a new workload for local education agencies

Since 2006, all LEAs have been required to complete a student-level data submission

each fall. This submission includes basic enrollment and demographic information, and it serves three key purposes. First, it keeps track of each of the state’s students, confirming his or her enrollment in California’s education system. It also records where each student is enrolled in school, which provides the information needed to calculate accurate graduate and dropout counts. Finally, it allows LEAs to annually update data on their students—their eligibility for and participation in certain programs, such as free and reduced-price meals. The state requires LEAs to certify the accuracy of their data because funding for certain programs is based on the enrollment figures.

Managing student-level data through an SSID number is an ongoing task that can be time-consuming and complicated. As school employees learn to use SSIDs to report

student information, they are finding that the quality of the data in the state’s database depends upon the combined effort of all LEAs. When a student moves to a school in a new district, the receiving school district works with CSIS to find the student’s SSID. Based on basic demographic information submitted by the district, CSIS returns potential SSID matches. The district must then take care to investigate potential matches and select the SSID previously assigned to the student; or if there is no suspected match, request a new SSID. Many districts are careful and deliberate about SSID selection, but others tend to make errors.

Another opportunity for error lies in data entry about the student’s identifying demographic information. If either the sending or receiving district gave the wrong gender or date of birth, the student’s SSID may be difficult to transfer properly. If the student records are not transferred and the receiving district mistakenly assigns a new SSID to the student, all previous information about him or her is lost. Alternately, the receiving district may locate a student record of a different student and erroneously assign that student’s SSID to the district’s new enrollee, thereby hijacking another student’s SSID. Each of these “anomalies” must later be resolved, requiring LEAs to communicate with each other to clear up the mistake. Particularly in areas with high student mobility, maintaining SSIDs represents a significant workload that districts did not previously face when they simply reported aggregate enrollment data once each year.

Under CALPADS, districts will also have to submit student-level data that were previously only reported in the aggregate. Student course enrollments and course completion data, information about teacher assignments, the status of English learners, immigrant student counts, and student discipline data are each submitted at different points throughout the school year.

Currently, the 250 districts that have volunteered to work closely with CSIS—known as “State Reporting” districts—report course enrollments for individual students. But for the remaining three-quarters of the state’s

LEAs, this task represents a major learning curve. These LEAs are accustomed to mapping course codes used locally onto the state's course codes as part of aggregate reports. The accuracy of this process varies among LEAs. But when they must submit course enrollments for individual students, it becomes critical that the data are accurate. Many districts are unprepared for this task. In addition, LEAs will soon be required to report on whether students completed the course, data that no LEA has submitted before.

Once CALPADS is fully implemented, CDE will be able to discontinue a number of current aggregate collections, which should offset some of the increased workload for districts. Nonetheless, CALPADS represents a dramatic shift from a set of annual collections to the need for constant maintenance and updates.

LEAs statewide face varying data management challenges

Although many districts have emphasized data accuracy and use internally for years, others have had difficulty placing a high priority on data quality amid other demands. As the state continues to finalize policies and practices, districts struggle to keep up.

Most districts are facing serious budgetary and fiscal challenges as the state's budget crisis continues. Many LEAs fear that they will have insufficient staff and resources to do the data work required. Representatives from CSIS report that some district data coordinators (employees responsible for maintaining and submitting student database files) struggle with inconsistent support from district leadership or face problems getting schools to submit data to the district in a timely fashion.

Under the state's previous data collection model, aggregate data submission allowed for individual inaccuracies to be buried in the group and go largely unnoticed. The requirement that local data managers maintain students' unique statewide student identifiers (SSIDs) represents a major new workload. This can stretch a district's limited resources, particularly under the immense financial strain LEAs statewide are feeling.

Even though LEAs receive some of their funding based on official enrollment counts that are reported to the state, some district leaders do not prioritize timely submission of accurate, reliable data. In a recent book published by the Thomas B. Fordham Institute, *A Byte at the Apple*, Margaret Raymond notes that the people responsible for submitting the data (school secretaries, local data managers) are often not involved in using the data and typically do not have the motivation to make sure their work is accurate and complete. Raymond further points out that the consequences of bad quality data are often not felt until long after they have been collected, so incentives to collect and submit timely data can be weak.

Limited human, organizational, and financial capacity at the local level threatens state-level goals

Even in districts with a strong culture of data, several factors can challenge their capacity to manage data well.

Frequent turnover of staff in data-entry positions, for example, exacerbates the input errors that are common at the school level. Raymond and her team of researchers from Stanford University are monitoring the accuracy of data that 300 schools report to the state. They have found staff turnover to be a major reason for imperfect data reporting. The research team reports that data entry and management positions turn over, on average, once every two years for most schools. For about 30% of schools, however, the position turns over every year.

Additionally, the student information systems (SIS) that districts purchase from vendors to manage their data vary in their sophistication, ease of use, and how well they integrate state regulations and policies. Some

districts use dated systems, and some small districts still rely on simple spreadsheet programs to manage their student-level data. CDE encourages districts to use state-funded resources, such as the California Resource Learning Network and the California Technology Assistance Project, to help improve the features of their student information system.

Some districts do not leverage all of the functionality in their SIS to help prevent data entry errors. For example, data entry clerks in Modesto City School District had for years typed in student information for every data field in their system. District officials eventually discovered 23 different spellings of the word "Modesto" in that system. Problems such as misspellings or blank data fields impede the school district's ability to submit its data to the state. (Modesto City School District eventually reduced the opportunity for error by instituting a set of drop-down boxes in its student information system, allowing, for example, district staff to select the name of the city rather than having to type it.)

State-level obstacles to improved data quality exist

State officials are struggling to balance the demand for school districts to meet reporting requirements with the need to provide them with sufficient time, technical assistance, and resources to prepare quality data. The launch of CALPADS may exacerbate this. According to officials at both CDE and CSIS, many districts statewide will likely not be ready to provide timely, quality data to CALPADS in its first year. According to a CSIS participant from Newark Unified School District, even districts with excellent data practices are intimidated by what CALPADS may mean for their district.

State officials are struggling to balance the need for districts to meet reporting requirements with the need to provide them with sufficient time, technical assistance, and resources to prepare quality data.

Successful data management: Districts statewide implement promising practices

California schools not only vary in size and configurations, but also in their data management competencies and challenges. The following case studies highlight some of the barriers districts have faced and the local data management strategies that have proven successful in overcoming those obstacles.

Case Study: Etiwanda School District provides automated feedback on data quality to school staff

Etiwanda School District in San Bernadino County needed to clean up its school information system. In particular, school support staff did not consistently fill in data fields, such as students' home language, ethnicity, and parent education level. In response, the district launched its Data Integrity and Anomaly System (DIAS) in late 2007. Developed by a programmer within the district, DIAS sends e-mail alerts to school secretaries or clerks every week, telling them where there is a blank field that must be resolved. The alert includes the student's name and SSID, making it easier for school-level staff to resolve the problem.

One year after its implementation, district CSIS Coordinator Mary Harker reports that most schools in the district have just one or two corrections to make each week. The DIAS system not only makes it easier for secretaries and clerks to do their jobs well, but it has also improved the overall quality of district data. When the Etiwanda School District submitted its data to the state this past year, district officials were able to confirm the accuracy of their data against CSIS data files smoothly and quickly.

At this time, Etiwanda School District does not have any systems in place to share DIAS with other interested districts because it was designed specifically to work with Etiwanda's own student information system. However, other districts—or potentially the vendors of their information system—could replicate the idea of DIAS with their own systems.

Case Study: Novato Unified School District identifies “data stewards”

Marin County's Novato Unified School District has been working to improve its data management practices since 1997, with help from CSIS. As a result of being able to commit valuable resources and staff training to this goal, this district already has substantial experience and good success in submitting much of its data on individual students and teachers to the state electronically.

Thanks to this training and experience, Novato Unified is relatively well prepared for the launch of CALPADS. But district leaders have been challenged along the way to establish a “culture of data” within the district. In the past, the Human Resources Department was in charge of personnel data, school site officials managed student demographic data, other staff oversaw program participation designations, and so on. When the district began submitting all of its data electronically, however, the entire burden of data collection and submission shifted to the district's Information Technology Department.

Other staff members relinquished ownership of data that was previously under their purview. Data quality was compromised, and communication among the stakeholder groups evaporated.

Connie Benz is currently coordinator of communications for the district and is also a trainer with the CSIS Best Practices Cohort. She recognized the need to change the district culture to value data. She established a data management team comprised of “data stewards” from throughout the district. Each of Novato's 16 school sites has an appointed data steward who is responsible for ensuring the accuracy and reliability of the data submitted. At the district level, a data management team works collaboratively, with members holding one another accountable for the data they manage. Data are no longer the responsibility of a single department: the whole team has ownership and works together to advance the district's culture of data.

Case Study: Long Beach Unified School District tracks down transfer students

Tracking the 88,000 students in Long Beach Unified School District has proven to be a challenging task for district officials. A large urban district with a highly mobile student population, Long Beach faces challenges in accounting for the whereabouts of each individual student over time.

If a student's SSID is properly transferred to a new school, Long Beach counts that student as a transfer. However, if the new district erroneously assigns a new SSID to that student, the student then has multiple identifiers, which causes an anomaly in the data system. Last year, Long Beach was faced with 9,000 anomalies.

Long Beach had so many anomalies partly because it lacked good mechanisms to house knowledge about where students go after they leave the district. This resulted in an over-reporting of students as dropouts. To correct this, John Novak from the Research, Planning, and Evaluation Department developed a web-based tool that makes it easier for the district to monitor requests for students' cumulative files from other districts and store results of attempts to contact former students. In yet another approach, the district tries to trick the system by submitting new identifier requests for students with anomalies. The district then uses a tool to examine the results and locate matches for former Long Beach students assigned incorrect SSIDs.

Last year, these methods enabled Long Beach to find about 300 students enrolled in other districts in California who would have otherwise been counted as dropouts. (For more information on dropout data, see the box on page 4.)

State policies may temporarily complicate local data management efforts

State education officials at CDE and CSIS are limited by several state policies that prevent them from taking actions that might improve the transition to CALPADS for school districts.

Although the state implemented programs such as CSIS and its Best Practices Cohort Project to improve data quality, it did so on a voluntary basis to avoid mandate reimbursements. These voluntary programs are likely to improve the quality of data submitted by participating districts, but they do nothing to address issues of data quality in nonparticipating districts. (See the box about mandates on this page. For more information on the Best Practices Cohort Project, see page 8.)

According to officials from the Data Management Division of CDE, the agency cannot create incentives for districts to submit accurate data, nor can it issue sanctions to districts that repeatedly submit their data in an unreliable or untimely fashion. Additionally, CDE cannot take any regulatory actions that could be interpreted as a mandate to deal with late or incomplete reporting. Without that governing tool, CDE officials have explained, they have trouble motivating certain districts to report their data completely or on time.

CALPADS requires lots of work to be done in a short time

Designing, developing, testing, and implementing a system as massive and complicated as CALPADS poses a major challenge to all involved. In addition, the rollout of CALPADS is occurring on a relatively short timeline. (See page 3.) This is a major task for state officials charged with its design and implementation, as well as for districts that must deal with a steep learning curve. After a long procurement process, the state contracted with IBM in 2008. The contract is for 2 1/2 years, through June 2010, and includes the first year of CALPADS implementation. This is an aggressive schedule for designing and building a complex system that involves more than 1,000 LEAs and millions of student records.

This compressed timeline also means that LEAs are likely to be “learning on the job” as they submit their data to CALPADS.

Some districts may not have sufficient opportunities to learn how to work with the new longitudinal data system prior to their first full data submissions, which could affect the timeliness and accuracy of state education data in CALPADS’ first year.

CDE officials say that when CALPADS is up and running, they expect that districts will learn quickly. Districts that are already stretched thin may not have the time or resources to proactively prepare for CALPADS. But when the system is in place, local and state leaders will have a better idea of what it will take to submit timely, quality data. At that time, CDE officials say, districts will be more likely to ramp up their local data management capacity in order to meet state goals. This may be a challenge, however, because of the state’s fiscal crisis and the scarcity of district resources to fund such efforts.

Districts will be feeling the effects of change soon

CDE is working with school district leaders to make the transition to CALPADS as seamless as possible. With the tight timelines that CDE, CSIS, and LEAs each face, all parties are working to make sure nothing falls through the cracks. However, state officials’ ongoing decision making as to how the system will work is beginning to have reverberations at the local level.

CDE is getting critical information to school districts as quickly as possible, but officials acknowledge that they have not yet delivered final versions of all of the

documentation LEAs need to get a head start on their transition to using CALPADS. For example, CDE released a draft of file specifications and data definitions in October 2008 and finalized them at the end of January 2009. These state policies dictate which data elements must be on a school’s enrollment forms. However, some districts may not have time to revise those forms before they begin registering new students for 2009–10. And others may still not realize or take seriously the need to create and print new forms. If local agencies fail to collect the right types of data from their students when they enroll, this could represent a major setback for districts and ultimately for CALPADS.

Reporting student enrollment in courses and their completion of those courses presents another major challenge that districts will face in the coming months. With codes representing each course, schools track the courses students enroll in and those that they complete. The codes will also eventually be used to monitor educator’s teaching assignments. LEAs have local course codes, which they must map onto standard statewide course codes. Many of these course codes changed and were released in October 2008, and further modifications may still occur. These small details have the potential to complicate local data management. In order to protect their limited time and resources, some districts are waiting to take action until final data definitions and course codes are released.

Designing, developing, testing, and implementing a system as massive and complicated as CALPADS poses a major challenge to all involved.

California is required to reimburse districts for mandated costs

Proposition 4, passed by voters in 1979, requires the state to reimburse local governments (such as school districts and county offices of education) for costs they incur fulfilling state-mandated activities. With California’s tight fiscal climate, school districts have pursued reimbursements for all expenditures they consider to be state-required, and state officials from the Department of Finance have been careful to avoid any new requirements that might be interpreted as a mandate. Thus, lawmakers have not made the work related to data quality a clear mandate with which districts must comply.



Strategies for improving data quality are emerging

No single measure will improve local data management capacity statewide because districts vary in what they need. Many state leaders believe that the best general approach to building local data capacity is ongoing training and professional development that is flexible enough to meet the changing needs of a district. California's policymakers have made efforts to support programs that serve this goal, though a variety of perspectives exists on which strategies work best and how much they might cost.

Best Practices Cohort Project represents the state's current major investment in local capacity building

In response to the legislative requirement that all districts participate in CALPADS, the state funded the CSIS Best Practices Cohort Project in 2006–07 to help LEAs prepare for CALPADS implementation. The Best Practices Cohort targeted districts not already participating in the CSIS State Reporting program. Originally termed “CSIS Lite,” the project is a less intensive version of CSIS State Reporting: it helps districts improve their local data management practices and prepares them to complete electronic data reporting when CALPADS launches in 2009. In January 2009, 768 LEAs were participating in the Best Practices Cohort. (For CSIS funding information, see page 10.)

The Best Practices Cohort curriculum is designed to meet the varying needs of participating LEAs. Participants are required to complete five deliverables, including attending four professional development sessions focused on improving local data management. Prior to each session, district participants complete a needs assessment to guide their learning. They then articulate the next steps to implement improvements to their practices. In some cases, CSIS helps districts create a data calendar that establishes a common reference for benchmarks and deadlines. In other cases, CSIS works with districts on internal governance—creating collaborations and data sharing among different departments to eliminate duplication of effort and reduce the incidence of error.

Many districts see the Best Practices Cohort as a win-win program: the project not only guides and supports their local data management practices, but also provides much needed additional resources. Districts receive funding for participating in the Best Practices Cohort—at least \$8.51 per student and possibly more depending on their enrollment. Funding is distributed quarterly if requirements are met.

A consultant's report advises the state to focus on quality

In 2008, Gov. Arnold Schwarzenegger and Superintendent of Public Instruction Jack

O'Connell commissioned a report from consulting firm McKinsey and Co. to provide the state with a long-term implementation plan for a comprehensive education data system. The work was paid for by a grant from the Bill & Melinda Gates and the William and Flora Hewlett foundations.

The report—*Framework for a Comprehensive Education Data System in California: Unlocking the Power of Data to Continually Improve Public Education*—was released to the public in December 2008. It recommends linking data from the K–12 public education system to existing and emerging data systems in higher education, social services,

California is not alone in facing data quality issues

Other states that have already implemented their longitudinal data systems report that poor data quality is their biggest challenge. In some cases, state leaders have improved local capacity to submit quality data in innovative ways:

- The Arkansas Department of Education spent more than a year training school and education service employees to use its longitudinal data system, called Triand. Arkansas was one of the first states to implement an education data system that incorporated all 10 “essential elements” of a state system endorsed by the national Data Quality Campaign—a collaborative effort encouraging the development of state longitudinal data systems. The campaign also works to improve the collection, availability, and use of high-quality education data.
- In 1991, Florida began awarding state funding to districts solely on the basis of information that had been submitted electronically to the state's pioneer data system, then known as FIRN (Florida Information Resource Network). This served as a high-stakes incentive for the state's 74 districts to make sure their data were accurate and consistent.
- Kansas uses a voluntary professional development program to help school employees use the state data system, Kansas Individual Data on Students (KIDS). Kansas created the program—the Data Quality Curriculum and Certification—because of a need expressed by users in the field. The program addresses data quality at the point of entry by working with data clerks to improve their practices.

prekindergarten, and the workforce. It also spells out specific strategies and benchmarks that would help improve the use of data for instruction and inform policymakers' decisions about education policy. Some observers have speculated that state leaders will use McKinsey's recommendations as the guiding framework for California's data systems moving forward.

The report proposes an incremental, three-step process based on the state's existing work on CALPADS. Before broader, more comprehensive steps are taken, however, McKinsey recommends that the state should "enhance the quality, accessibility, and use of K–12 data."

The report reflects the input of 200 individuals representing more than 100 organizations statewide. At least some of these individuals cited concerns about the accuracy of the information in education data systems. In response, McKinsey recommends several specific steps that the state should employ to increase accuracy in its system. These steps, they say, would increase the timeliness of data submission and reporting.

The report divides its recommendations regarding improving data quality into three areas:

1. Data, information, and tools;
2. Governance, policies, and funding; and
3. Culture, training, and incentives.

The report's authors say that improving the accuracy of education data in California cannot be accomplished without an appropriate level of investment in data quality initiatives. They note the plan they recommend would cost the state between \$32 million and \$66 million in one-time costs and \$4 million to \$8 million in ongoing annual costs. Both the one-time and ongoing costs are similar to those currently spent on IT projects, the report points out.

As part of its one-time cost estimates, McKinsey expects \$10 to \$30 per student to be spent to improve the quality and timeliness of existing data collections through local initiatives recommended in this report. McKinsey notes that some of this money could potentially come from existing state programs supporting data quality initiatives, such as CSIS Best Practices Cohort.

Data, information, and tools

In studying how state data are collected and compiled, McKinsey consultants identified several inefficiencies. They say that the back-and-forth between the state and LEAs to certify data "creates delay, wasted efforts, and increased opportunity for errors." Instead, McKinsey encourages the state to focus on providing LEAs with advanced data quality tools, such as automated error-checking tools, to be used when the data are first entered at the school level. These tools could also automate some of the data entry: for example, automatically filling in a city and state when the zip code is given.

Improving data at the source of entry for key data elements has the dual benefit of enhancing accuracy and saving time and money, the consultants say. They point out that some large districts may already have advanced error-checking tools as part of their student information systems provided by outside vendors, but acquiring such tools can be cost prohibitive for small districts. Purchasing these tools on a statewide basis could be less expensive; and if most districts across the state are using one tool, it could also add the benefit of compatibility among systems.

Governance, policies, and funding

McKinsey recommends several improvements to the way the state manages its data. For example, although the state has anomaly-detection measures in place and requires districts to certify the accuracy of their data at the end of the collection process, the report recommends that the state also audit the tools districts use to do this work. Field visits to a sample of LEAs determined by a sampling methodology would help ensure that

the tools and other processes that districts have in place are resulting in accurate data.

McKinsey also weighs in on some of the processes that districts should have in place to ensure accurate data, such as an integrated calendar of data collections. The report suggests that it should be the state's role to provide LEAs with this documentation, which should include interim deadlines and goals.

Culture, training, and incentives

The report also makes the case that incentives for local agencies to produce high-quality data are important. They note that the most powerful incentive for districts is the usefulness of the information. If educators at the local level utilize the data reports prepared by the state and grasp the value of high-quality data for improving student progress, they will begin to see timely submission of data to the state as a mutually beneficial process. To change the culture of data use, the report suggests that the state should provide analytical reports to local school districts. The state should also develop rewards for schools and LEAs that maintain quality data and consequences for those that do not, though McKinsey does not specify what those rewards or consequences might be.

The report also stresses the importance of training and building skills to increase school district capacity to maintain quality data. The consultants suggest that California develop effective data-quality trainings and a certification process. The state could also include collecting, maintaining, and using data as part of a pre-service training and ongoing professional development for educators and administrators.

Many districts see the Best Practices Cohort as a win-win program: the project not only guides and supports their local data management practices, but also provides much needed additional resources.



The state's commitment to data quality remains uneven

For years, policymakers have struggled with prioritizing local capacity building while appropriating limited state funds. A proposal to support local data collection and reporting by providing an ongoing incentive grant of \$5 per student to LEAs is often talked about but has yet to be approved by lawmakers.

Legislative efforts for ongoing funding have failed

In the 2006–07 and 2007–08 legislative sessions, two bills were proposed that would have provided annual funding to support local data management activities at \$5 per pupil. Assembly Bill (AB) 2167, filed in 2006 by Assemblymember Juan Arambula, would have provided \$5 per pupil to LEAs enrolling at least 1,000 students and a flat rate of \$5,000 to districts with fewer students. These funds would have helped LEAs update enrollment counts, resolve anomalies, and facilitate information transfer between CDE and the LEA. This provision was eliminated from the final bill, however, in part because one-time funding was provided to prepare LEAs for CALPADS. Policymakers provided a total of \$29.5 million over three years so that all eligible LEAs could participate in the Best Practices Cohort. They allocated an additional \$7.9 million in 2008–09 for the same purpose.

Although the one-time funding for the Best Practices Cohort has been important, it is not intended to support the ongoing workload of all LEAs to collect, maintain, and submit student-level data. In 2007, Assemblymember Mike Feuer initiated a second legislative effort to provide such support for LEAs. Unwilling to commit ongoing funding to local data activities, the governor proposed in the May budget revision that the state allocate \$65 million in one-time funds to CSIS. The funds would have been distributed over two years to school districts based on CSIS's assessment of each district's training and other needs prior to the implementation of CALPADS. However, this funding was rescinded at the last minute. Assemblymember Feuer's legislation was also unsuccessful.

CDE sees investing in local districts' data capacity as a proactive approach to ensuring quality data and has long advocated for \$5 per student in ongoing funds to be spent on quality initiatives. The Legislative Analyst's Office (LAO) recommends funding local quality initiatives to prepare for CALPADS, though they support providing \$2.50—not \$5—per student, saying in part that the per-pupil amount will likely increase over time. And although the LAO recommends funding training initiatives such as the Best Practices Cohort, they do not support ongoing funding for data quality, arguing that the value of the data should provide enough motivation to ensure quality.

Other sources of funding for data quality improvements could soon become available.

As Congress develops an economic stimulus package, additional federal funds for education data systems may be included. The House version of the bill passed on Jan. 28, 2009, included \$250 million to help fund the development of state data systems, \$5 million of which may be used to fund efforts to improve data coordination. California is expecting to receive 10% to 12% of those funds and may use them to start implementing recommendations in the McKinsey report, according to CDE. However, the Senate version of the stimulus package passed on Feb. 10, 2009, does not provide money for education data systems. At the time of this report's publication, it remains to be seen how a federal stimulus package will affect California's data system.

The challenge of local data quality should get easier as everyone involved gets used to the new system.

How much will it cost to prepare districts for CALPADS and improve local data quality?


There are several proposals and estimates, based upon different sets of assumptions:

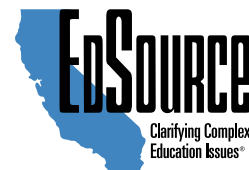
- Superintendent of Public Instruction Jack O'Connell is calling for ongoing funding of \$5 per student to support data quality.
- CSIS Best Practices Cohort participants and CSIS State Reporting districts receive one-time funds of \$8.51 per student or more, depending on enrollment, in preparation for CALPADS.
- The Legislative Analyst's Office recommends spending \$2.50 per student (\$44 million) in one-time federal funds to establish good data use and management practices.
- McKinsey and Co. estimates one-time costs of \$10 to \$30 per student to support local training, hardware, cleansing, and incremental data collection costs. Some of these funds could come from existing programs, such as the Best Practices Cohort.

Time is running out, and questions remain unanswered

As the fall 2009 rollout of CALPADS nears, many districts are unsure how they will meet expectations for timely and accurate data submission. With the state facing historic budget shortfalls, districts are already struggling to make ends meet. Anticipating the increased work and resources that CALPADS will require, some local officials have warned that they may not be able to afford to participate in CALPADS. Officials at CDE struggle to respond, but both mandating participation in CALPADS and providing supplemental resources to do so is beyond their purview.

It is likely that the first year of CALPADS will be one of trial and error. The ongoing challenge of local data quality—and its importance—should not be taken lightly. That said, the task should get easier as everyone involved gets used to the new system.

Long term, California's shift to CALPADS will almost certainly represent a major milestone in the development of a longitudinal data system that is current, accurate, and reliable. Such a system can help the educators who manage the state's public schools, the policymakers who oversee them, and the researchers who study them all better understand how the state's students and schools are progressing. 



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
Clinical Professor of Urban School Leadership,
Claremont Graduate 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 About CALPADS and Data Quality

- CSIS Best Practices Cohort Project: www.csis.k12.ca.us/shared/bp-cohort_menu.asp
- CDE's CALPADS website: www.cde.ca.gov/ds/sp/cl
- Data Quality Campaign: www.dataqualitycampaign.org/
- *A Byte at the Apple: Rethinking Education Data for the Post-NCLB Era*. Marcie Kanstoroom and Eric Osberg, eds. for the Thomas B. Fordham Institute, November 2008. www.edexcellence.net/doc/20081117_ByteAtTheApple.pdf
- *California's Emerging Data System: A Status Report*. EdSource, October 2008. www.edsource.org/pub_datasys10-08.html
- *Education Data in California: Availability and Transparency*. A "Getting Down to Facts" research study conducted by Janet Hansen, March 2007. www.stanford.edu/group/irepp/cgi-bin/joomla/index.php?option=com_content&task=view&id=33&Itemid=45
- *Framework for a Comprehensive Data System in California: Unlocking the Power of Data to Continually Improve Public Education*. Research and analysis conducted by McKinsey and Co., December 2008. www.cde.ca.gov/eo/in/pc/documents/yr08mckinsey1218.pdf

Acknowledgments

This report was researched and written by:
Heather Barondess

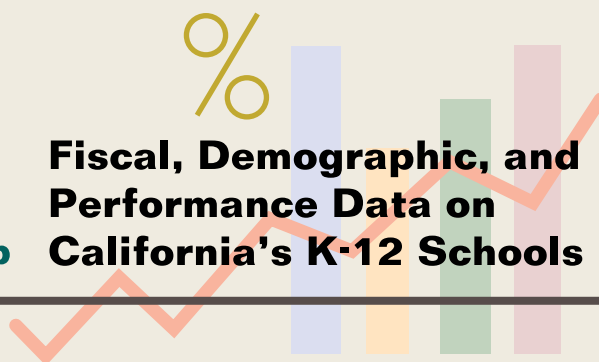
Edited by:
Mary Perry



Ed-Data

Education Data Partnership

Fiscal, Demographic, and Performance Data on California's K-12 Schools



Turning K–12 Data Into Information You Can Use

The Ed-Data Partnership website—www.ed-data.k12.ca.us—provides Californians with comprehensive, accessible education data for schools, districts, counties, and the state as a whole. You will find performance, staffing, and demographic data, as well as financial reports. Clear data explanations accompany each report, and powerful comparison and filtering tools make examining data easy. In addition, you will find extensive documentation, charts and graphs, and articles about education in California. The site offers multiple ways to approach and view information.

Ed-Data provides:

- School, district, county, and state profiles, including the Academic Performance Index and adequate yearly progress results, graduate/dropout rates, SAT results, enrollment, student characteristics, and staffing information that provide a comprehensive portrait of schools.
- Teacher salary and benefits data, including district and state averages that outline some of the most important expenditures of school districts.
- Bond and parcel tax election data for districts.
- Easy-to-use comparisons of schools and school districts based on the criteria you choose.
- Clear and comprehensive district and county office of education financial reports, with charts and comparisons.