
California Policy Brief

Topic: State Data Systems

Main contact: Russlynn Ali, Executive Director, rali@edtrustwest.org

Statement of the Problem

California collects and reports a good amount of useful education data and its data collection efforts are improving. We know now, more than ever before, which schools and groups of students are meeting state standards and which are furthest behind. We know which groups of students and schools have access to certified and experienced teachers, and rigorous college and work preparatory courses, and which ones don't.

These kinds of data have been critical to *motivating* reform, as the State Chief Jack O'Connell's recent commitment to close California's achievement gap makes clear. Yet *achieving* big changes and sustainable reform requires data that is simultaneously more detailed and broader than is currently available. Effective change requires a deep and comprehensive understanding of what works in educating students and what doesn't, and why.

Today though we treat education as a black box: what goes on inside the schoolhouse or the classroom door is not subject to measurement and comparison, and all that we can do is assess outcomes *post hoc*. This is not fair to anyone in the system. It is unfair to teachers, who could use much more detailed information about their students and their own effective practices. It is unfair to administrators, who lack the tools to support continuous improvement in their schools. It is unfair to state and local policymakers, who are forced to allocate funds without information about which programs work and why. And it is most certainly unfair to students who are the victims of seemingly endless experiments intended to transform their schooling experience with little assurance or evidence of what actually works.

Currently, California's education data system barely merits the name: it is a confusing assembly of collection vehicles, aggregated at different levels, reported at different times, housed in a multitude of different databases and only linked manually according to the ever-increasing demands of federal and state reporting—work that is complicated by the absence of a central repository for student-level information. In addition, sharing data between school districts, like data on transferring students, is inefficient and sometimes doesn't happen at all.

Fortunately, some of this will soon change. Within the next three years, California should have new longitudinal systems to house student data (CALPADS) and teacher data (CALTIDES). But a wide gulf lies between what the new data sets should and could tell us, and what they will actually have the capacity to do. Even with the building of CALPADS and CALTIDES, answering some key fundamental questions -- for example, linking achievement levels in middle school to success in particular high school courses, or connecting high school performance to workplace success -- will remain out of reach.

For all the talk of increasing data-driven decision-making and conducting rigorous evaluations of educational programs and investments, California's data system as currently planned is likely to remain merely a means of producing required federal and state accountability reports. As CALPADS and CALTIDES are developed, the biggest risk is that the focus on cleaning up the current situation will trump the more important aim: incorporating data into decisions at every level of the system and building a world class education system that continuously learns and

improves. CALPADS and CALTIDES represent a significant improvement from current data collection and reporting, indeed. But while they will make some data use in decision-making *possible*, they cannot make it standard practice in and of themselves, CALPADS and CALTIDES don't reveal enough.

The biggest problem of the past seems to be that no real plan led the design of the current system – and if there was one, it was merely to comply with federal reporting and accountability requirements. A succession of tacked-on requirements added a series of data collection vehicles to the mix, and so we built a data system. Now California has an opportunity to design an integrated data system that is designed and used for continuous improvement at state and local levels. The state should seize this chance to consider how to best integrate data into education policy and practice—to evaluate trends and programs, predict the outcomes of future investments, and explore the interaction between various factors in student, teacher, and school success—rather than simply use it to describe the way things are. For brevity, we'll refer to this comprehensive information system as a “smart system” —one that bridges the education information gap and provides information to make smart education choices in real time.

Policy Issues, Options, and Recommendations

If conducting analyses and employing data in decisions are truly the goals of the state, California's longitudinal systems must be more thoughtfully designed into a deep K-24 data system and must sit at the center of a broader movement toward information use and continuously improving education systems. California could take some immediate steps to help make this happen.

Step 1: Build the Political Will to Get Good Data and Information: It's been said by researchers and policymakers alike that there is no constituency loudly calling for a good data system. That's true. In part that's because many people assume policymakers rely on good information as they make policy. Indeed, parents are shocked when we explain how little the state really knows about its public education system. And although policymakers and educators say they want good information, because decisions are being made everyday as if they already have it, there is little political will to develop a worthwhile system.

Stakeholders absolutely want good data, but the process to get from here (decent data, but not enough) to there (the smart system) needs some explaining to a lay audience. For starters, the word “data” should be replaced with the word “information.” “Data” emphasizes the work of collecting and storing unitary records. Instead what we should be talking about is “information” because it focuses on the purpose – helping people to inform themselves about what is happening inside and outside of schools. There might not be a constituency calling for good data systems, but surely we can build one calling for much needed information. Indeed, communities throughout the state organized and pushed for the reporting of restricted and unrestricted funds by source on the School Accountability Report Cards (SARC)—and they won in SB 687. As an immediate next step the state should centralize all of the SARC data and create a web based application to allow easy access and analysis.

Equally important, there must be a sea-change in how data and information are perceived by educators. There is a climate of fear that somehow data will be used against them, for punishment rather than as an aid for continuous improvement. Some incentives described below could go far to show how longitudinal information can be used to help—not hurt—the teaching profession.

Step 2: Think Big: The Smart System Should Link to a Multitude of Data Sets: A smart data system must house much more than data about K-12 education. When developed, CALTIDES and CALPADS need to link to data sets about higher education, including 2- and 4- year university data and postsecondary vocational participation data. The system should also link to employment data; military service; incarceration; and health and human services to get a clear portrait of what happens to students as they journey from our schools into adulthood.

Linking together data systems that were built as separate silos will require coordinating the efforts of various state agencies outside of the California Department of Education. As an immediate step to move past the climate of distrust between state agencies the state should create an independent data oversight commission, we describe this more fully below.

Step 3: Collect more data elements: Done right, California's smart system can drive critical decisions from the classroom to the Capitol building. But this will take more than tracking annual test scores and meeting reporting requirements. A few additions to the data items schools and districts collect could dramatically increase the number and scope of questions the system is able to answer.

Under the current plan, though a few data fields will be added, almost all of the data elements CALPADS will include are already collected in the systems it will replace. More collection is necessary. However, the California Department of Finance rigorously enforces the constitutional prohibition on unfunded mandates—which includes asking districts to collect more data without funding them. It's unlikely the CDE could ask districts to gather more data elements without statutory backup. Additionally, CDE will need extra staff to meet the new demands.

As an immediate step, and to build in the ability to conduct program evaluations with CALPADS data, legislators should specify a few additional elements for collection and allocate funds for districts to do so. These elements are currently collected by some agency but will not otherwise be available in CALPADS. The must-have elements include: student attendance records; student end-of-course grades; student scores on college ready assessments (AP, SAT and ACT), and teacher scores on each administration of exams required to receive a credential (the CBEST, CSET, and RICA).

Next, we should go further. For example, if we are to evaluate teacher and administrator training programs or curricula and intervention programs in which students are enrolled, we must know which students, teachers, schools, and districts are participating in them. In truth, any program that the state would wish to evaluate on an on-going basis should be included in the system at the level of the institution, personnel, and student.

We understand the CDE is developing a web-based application where schools will be able to maintain their school-level characteristics online, potentially including things like instructional material inventory, curriculum choices, and reform packages. Either such a system should be linked to CALPADS or data fields should be added in order to evaluate the effects of those school-level characteristics on student achievement.

Of course, these gross measures alone cannot tell us the whole story of a particular program. Duration and implementation can vary widely. And though this list is by no means exhaustive, these new data could hint at effectiveness and trigger more rigorous investigation.

Time is of the essence. According to the CALPADS RFP, it takes about two years to implement a new data requirement. Better, then, to implement changes during the design and phase-in of the new system so California analysts can begin to conduct the kind of rigorous evaluation necessary to inform decisions.

Step 4: Provide Better Access: Data cannot become evidence to support decision-making unless it's accessible to the researchers who can conduct important and provocative analyses. Currently, California is generous with its collected data. Because it includes no student-level information, the CDE can make much of it publicly available with no qualms about violating privacy laws.

A longitudinal system though, while it represents a much higher-quality and more powerful tool for data analysis, could perversely limit access to and use of that data. Without clear standards and policies—which do not currently exist—many stakeholders could fail to realize any benefit of the new, higher-capacity system. If California is to reap the full benefits of a more powerful education data system, it is vital that the CDE lawyers don't interpret privacy laws in an overly restrictive way, and that access procedures are in place in advance, lest access protocols become a bottleneck in data's translation into information for decision-making.

Additionally, CDE can build capacity to satisfy researchers' needs internally. They'll need to staff up fast though. With the implementation of unique student identifiers, data requests have already been growing. This demand will only increase once CALPADS is in place. The more strict the department's interpretation of privacy laws, namely FERPA (The Family Educational Rights Privacy Act), the greater the burden will be on its own staff to produce data for researchers that has already been analyzed and aggregated to mask student-level information.

Step 5: Fund and Provide Incentives for Data Accuracy and Strengthen Reporting: Increasing the quality of data in CALPADS requires additional funding and incentives for districts to report accurate information. While planned changes in the new system—like state verification of district data and the capacity to continually update submissions—will help, many districts will still lack the resources and the incentives to check data accuracy in a number of areas. To improve the quality of data, districts need both money and motive.

The question of money to districts has been recognized in budget process and proposed legislation, and attempts to address it have failed. Advocates will need to try again. On-going money will help districts build the new capacity, help aid clean and standardized input of new data elements, and maintain local systems.

The question of motive is more complex. As one district data administrator recently told us, districts do an excellent job of accurately reporting the enrollment and attendance information that determines annual per pupil funding, but there is no such incentive to ensure the accuracy of other data elements.

One way to offer districts a worthwhile exchange for accurate data would entail assisting them in making high-value use of the data they submit. Data-rich states, like Florida, have found that providing reports and detailed analyses to the entities that feed their comprehensive system is a good way of ensuring that the data input is accurate and valid. In California however, schools and districts must submit a good deal of data but get little back themselves other than student test scores and a general look at their comparative performance.

Moving forward future reports could contain descriptive information that might be otherwise complicated for schools and districts to obtain. For example, these reports could show trends over time and trace teacher and student mobility. The state could help districts obtain and fairly use value added analyses of programs, teachers and interventions. Another example might include offering teacher preparation programs reports on their teachers' students outcomes and a web-tool that allows them to conduct their own investigations.

Lastly, when implemented, CALPADS will include a robust ad-hoc query system for at least CDE staff. This should be expanded into a query tool that would allow agencies that link to the new system to pose their own questions and explore interactions themselves. District and school officials themselves should have access to a query system that allows them to learn about their own students, teachers, and schools using the data they feed into to the system. Indeed, local educators are far more likely to trust conclusions that they draw themselves rather than something provided by external agents.

The point is, although California has done an admirable job of data reporting with the current system, failing to expand that reporting and evaluation to exploit the power of an integrated, longitudinal system would represent a huge missed opportunity. Seizing this opportunity would garner much needed educator and public buy-in for the further development of the comprehensive smart system itself.

Step 6: Build Local Capacity for Data Analysis and Use: California may be designing a new state-level system, but it should not ignore the potential power of increased data use at the local level. Most decisions around education practice are still made by schools and districts, and promoting data use must involve teachers, administrators, and district staff as more than mere collectors of data. While the state system itself may not be sufficient to guide micro-level decisions, the state can do much to build districts' capacity as part of a broader commitment to data use in education.

For example, California could follow the example of other states and develop a web interface for teachers. Such a system could allow teachers to submit student test scores and receive feedback from the state system on how their students' performance measures up to state benchmarks and school, district and state averages. It could also become a repository for lesson plans and college prep high school course syllabi, instructional and classroom management tools, assessment items aligned with individual standards, and professional development opportunities—all tailored to teachers' individual needs.

Tools for local practitioners to access data are important, but what makes them truly useful is high-quality professional development. As an immediate step, the state should provide comprehensive and meaningful training in use of CALPADS and its integration with local systems where they exist. Knowledge of and confidence in the state system will help integrate the data it contains into local decision-making.

Step 7 (And Do This First): Establish an Education Data Oversight Commission: California is stuck. We've come a long way with CALPADS and CALTIDES, but we've not gone far enough. Getting any of the above recommendations in place is going to require a new player. Because the truth is, the primary obstacles to a comprehensive data system in California are about organizational territory and politics. Therefore, the first step in developing a smart system is to create an oversight commission to guide the development of the smart system.

An Education Data Oversight Commission should bring together the puzzle pieces for a more complete picture of California's data landscape. For example, linking existing data systems together—to answer questions about the relation between high school performance and workforce success—requires in practical terms coordinated effort on the part of several different state agencies. Mandating and managing this coordinated effort is a challenge that no existing agency is well-suited to taking on, precisely because they are already occupied in the details of managing their own data systems.

An Education Data Oversight Commission could determine how data should be linked, the terms under which access should be granted, and the priorities for building a comprehensive system. Indeed, these questions should be handled outside of the particular interests of any particular component of the system. Such a commission would include representatives of the various stakeholders in the system, including state agencies, researchers, and practitioners. It would be empowered to draw on resources within existing agencies to implement linkages between existing data systems, and would manage the legal and privacy issues involved in granting access to linked data.

Additionally and perhaps most importantly, the Education Data Oversight Commission should have an arm that acts as a data warehouse – a repository of different data sets and the producers of state issued annual reports. The Commission could also determine what the proper supports are that need to be in place to make sure the data collected make the transition from numbers in a warehouse to contextualized information, and from data to information that supports high quality decisions about education policy and practice. The Commission could be and perhaps should be a public/private partnership – endowed with promised support to ensure maintenance and continuity.

Summary of Research

Berry, B., Fuller, E., Laird, E., and Reeves, C. Linking Teacher and Student Data to Improve Teacher and Teaching Quality. Data Quality Campaign, March 2007 <http://www.dataqualitycampaign.org>

California Department of Education. *California Longitudinal Pupil Achievement Data System: Request for Proposal*. Sacramento, CA. 8 Sept. 2006.

CSRQ Center Report on Elementary School CSR Models and CSRQ Center Report on Middle and High School CSR Models. Comprehensive School Reform Quality Center. 2006.
<http://www.csrq.org/reports.asp>>

Haertel, Edward. Using a Longitudinal Student Tracking System to Improve the Design for Public School Accountability in California. August 2005. <<http://ed.stanford.edu/suse/faculty/haertel/Haertel-Value-Added.pdf>>

Hansen, Janet. *Education Data in California: Availability and Transparency*. RAND Corporation, 2006.

Ingram, D., Louis, K.S., & Schroeder, R.G. "Accountability policies and teacher decision making: Barriers to the use of data to improve practice." *Teachers College Record* 106 (6), 1258-1287.

Khanna, R., Trousdale, D., Penuel, W.R. & Kell, J. *Supporting data use among administrators: Results from a data planning model*. Annual Meeting of the American Educational Research Association, April 1999.

Mills, Jack. *State Data Systems and Privacy Concerns: Strategies for Balancing Public Interests*. Achieving the Dream, February 2005. <<http://www.jff.org/Documents/StateDataSystems.pdf>>

National Center for Educational Accountability. *State Data Collection Survey Results*. Data Quality Campaign, 2007. <<http://www.dataqualitycampaign.org>>

O'Connell, Jack. *State of Education Address*. Sacramento, CA. 6 Feb. 2007 and 7 Feb. 2006.

Springboard Schools. *Bringing the State and Locals Together: Developing Effective Data Systems in California School Districts*. Institute for Research on Education Policy & Practice at Stanford University, March 2007.

Sunshine Connections Homepage. Florida Department of Education. Retrieved 8 May 2007 from <<http://www.sunshineconnections.org/>>

Wayman, J.C. & Stringfield, S. *Teachers using data to improve instruction: Exemplary practices in using data warehouse and reporting systems*. Paper presented at the 2005 Annual Meeting of the American Educational Research Association, Montreal, Canada.