THE AVAILABILITY AND TRANSPARENCY OF EDUCATION DATA IN CALIFORNIA

Janet S. Hansen
Committee for Economic Development
2000 L Street NW, Suite 700
Washington, DC 20036
janet.hansen@ced.org

Abstract
This article describes the potential for using K–12 education data to support school improvement efforts and the effective and efficient use of education resources. It examines the availability and transparency of education data in California as part of the Getting Down to Facts effort to improve education decision making in that state. The study addresses three broad questions related to the state’s education data system: (1) What is the current condition of California’s education data system? (2) What is being done to strengthen it? and (3) What are the obstacles and challenges to improving the system so that it can be an effective tool for helping policy makers and education leaders understand how schools are performing and how resources can most effectively be allocated?
1. INTRODUCTION

Good data are key to school improvement efforts. Californians share with their fellow Americans a lofty ambition—educational excellence for all students—and annually invest over $45 billion in public schools to achieve this worthy objective. Investing these funds well depends on answering such questions as how schools are doing, whether investments are accomplishing their goals, whether teachers have information allowing them to tailor instruction to specific student needs, and what policies and practices show evidence of improving student achievement. The ability to answer these questions in turn depends on strong data systems that collect the relevant information and make it available to various stakeholders in the education enterprise in accessible and understandable ways. States that have led the way in education reform have typically developed a “culture of data” that encourages the use of information to guide policy decisions.

As part of the Getting Down to Facts project, this article was commissioned to examine the availability and transparency of K–12 education data in California and their capacity to support school improvement efforts and effective and efficient use of education resources.

Based on reviews of public documents and Web sites and interviews with twelve individuals knowledgeable about California and other state data systems and data system building, the article concludes that California is lagging most other states in developing education data systems capable of helping policy makers and others understand how schools are doing and which programs are effective for improving student attainment and achievement. The state has only in the last several years begun to move beyond the traditional approach to data collection: emphasizing discrete, disconnected data “silos” that address reporting and monitoring requirements but do not lend themselves to analyses that can guide policy and program improvement. Recent steps to begin implementing longitudinal student and teacher data systems and to improve education management are promising, but the success of these undertakings is not yet assured. While efforts to build strong data systems undoubtedly encounter technological challenges, the primary barrier in California currently seems to be the lack of strong state advocates for the development of a first-class education data program.

The article begins by briefly describing state and federal school reform efforts, federal data initiatives, and foundation-sponsored data improvement.
activities that provide a context for considering data system development in California. It then provides an overview of the condition of education data in California today and a discussion of the long way California still has to go to be capable of providing the information that policy makers and citizens need to address key policy questions. Next, the article examines the factors that appear to be holding California back in developing statewide K–12 data systems. Finally, it poses challenges that California policy makers will need to address if the state is to successfully implement the new data initiatives it currently has under way.

2. THE CONTEXT FOR CONSIDERING DATA SYSTEM DEVELOPMENT IN CALIFORNIA

The standards and accountability movement of the 1980s and 1990s, which helped shift the focus of school reform from processes to educational outcomes, caused many states to develop new or strengthened information systems capable of tracking student and school performance. The federal No Child Left Behind (NCLB) Act of 2001 upped the ante by specifying numerous kinds of information that would have to be reported about students, teachers, schools, districts, and states, often in more detail (for example, subgroup test scores, highly qualified teachers) than existing state and local information systems supported. Thus states like California that had not previously created integrated or longitudinal statewide information systems faced new pressures to do so in order to meet new requirements efficiently, while states with existing statewide databases often found that their infrastructures needed upgrading. In 2005 the federal government launched the Statewide Longitudinal Data Systems Grant Program to support the development of state data systems. The California Department of Education (CDE) received a three-year, $3.255 million grant in the initial round of 2005 awards to develop the California Longitudinal Pupil Achievement Data System (CALPADS).

Meanwhile, the federal government has also been moving to streamline its own data collections and move them to Web-based reporting formats, a development with consequences for state data systems. In 2002 the U.S. Department of Education (ED) began the Performance-Based Data Management Initiative (PBDMI), an effort to consolidate sixteen separate data collections (administered by seven program offices) into a single data collection system and to eliminate duplications, conflicting definitions, and information not needed to evaluate departmental programs. State education agencies are collaborating in this effort by helping to develop common data definitions and by piloting the new collection system while continuing to meet their ongoing reporting requirements (U.S. General Accountability Office 2005). In addition to modifying data collection, the initiative involves the development
of a Web-based education data exchange network (EDEN) to provide state education agencies and the federal government the capacity to transfer and analyze information about education programs.  

The federal government is not the only outside influence on the development of state data systems. The National Education Data Partnership, funded by the Bill and Melinda Gates Foundation and the Broad Foundation, has sponsored the development of a Web-based analytical service for education data (SchoolMatters) and the provision of consulting and technical services to assist states in building their data infrastructures. The Gates Foundation has also funded the Data Quality Campaign (DQC), created in 2005 with the goal of seeing longitudinal student data systems established in all fifty states by 2009. Operated by the National Center for Educational Accountability, key DQC activities include promoting data systems through research studies and policy briefs, surveying the status of data system development in each of the fifty states, and providing snapshots of every state’s progress in developing longitudinal student information systems.

3. THE CURRENT CONDITION OF EDUCATION DATA IN CALIFORNIA

If California has a data “problem,” it is not a scarcity of data being collected and reported on students, teachers, and schools. CDE currently has 130 active data collections. Key ones include the California Basic Education Data System (CBEDS), which gathers student and staff demographic information using three separate forms; fiscal data, collected since 2003–4 using a standardized account code structure (SACS); and student achievement data, provided to CDE by test vendors.

Additional data on teachers is gathered by the California Commission on Teacher Credentialing (CCTC, which is independent and not part of CDE) and the California State Teachers Retirement System (CALSTRS). Universities and the state’s Employment Development Department (EDD) also have information on teachers, such as their preparation programs and the wages paid to individual employees.

California makes much of the information it collects on students, schools, and teachers publicly available in various ways. CDE has developed a Data Resource Guide, an online catalog of the department’s data products. It maintains an online system called DataQuest that provides information on accountability scores test data, enrollment, graduates, dropouts, course enrollments,
staffing, and English learners at (depending on the data element) the state, county, district, and school levels. The DataQuest site includes QuickQuest (a faster way of accessing data), a parents’ page to facilitate obtaining information about a specific district or school, methods for creating individualized reports, and links to sources of directly downloadable data from CBEDS.5

CDE is a partner in Ed-Data, a Web site that provides fiscal, demographic, and performance data on California’s K–12 schools.6 Pull-down menus provide access to data at the state, county, district, and school levels. The site provides a glossary of terms relevant to understanding California’s education data and articles analyzing California education issues. Ed-Data is the product of a decade-old partnership among CDE (which makes data available), the Alameda County Office of Education (which designs, operates, and maintains the Ed-Data Web site), EdSource (an independent, not-for-profit organization that prepares graphic displays for the Ed-Data Web site, sets the context for the various reports and profiles, and monitors information on users and suggestions for future improvements), and the Fiscal Crisis and Management Assistance Team (an independent, state-funded entity that administers the partnership). CDE also provides data for the SchoolMatters project.

Finally, California has a School Accountability Report Card (SARC), which was mandated by Proposition 98 (the Classroom Instructional Improvement and Accountability Act) in 1988. SARC provides parents and the community with information about the condition and performance of each public school in the state.

CDE, which is responsible for most of the K–12 education data collected in California, has been working to manage its data programs more efficiently and effectively in response to both internal evaluations and a 2002 consultant’s report commissioned by the California Department of Finance. That report, conducted by MGT of America (2002), criticized CDE data management on a number of dimensions:

- Data collection within CDE was highly decentralized and varied within and between offices.
- There had been minimal efforts to coordinate data collection among CDE program offices or to obtain a holistic picture of the data collected by the department.
- Program offices managed data in a decentralized fashion without rigorous and aligned data management standards or a common vision about data management.

• There was no system for naming and defining data consistently throughout the department.
• There was limited and inconsistent data validation.
• The department relied heavily on paper submission of data.
• Data collections involved inconsistent units of analysis or time periods.
• There was limited documentation of existing data.
• Data were stored in many different ways and locations.

The lack of common data definitions, the limited and inconsistent data validation, and the reliance on paper submissions compromised the accuracy and reliability of CDE data and inhibited data sharing across the department. Data sharing was also constrained by inconsistencies in units of analysis and time period coverage, by limited documentation, and by multiple approaches to data storage. Limited data sharing at CDE increased the reporting burden on districts, which had to respond to multiple data requests from various CDE offices.

CDE has taken a number of steps to respond to the problems identified in the MGT report (and in its own earlier internal evaluation). It established a Data Management Division with responsibility for providing data management leadership throughout the department. As of 2006, the CDE staff had created “preferred data definitions” for forty-eight data elements with many more planned, thus beginning the important process of standardizing data definitions throughout various departmental data collections. Staff also developed the Data Resource Guide, a major contribution to making California’s data collection instruments, databases, and reports more transparent to users.

CDE reduced data collections from 158 to the current 130, with hopes of further reductions. Staff visited Florida and Ohio, states with reputations for strong data systems and plans, and began working on longitudinal data systems for both student and teacher information (more on these below).

CDE also supports the work of the California School Information Services (CSIS) Program, a legislative initiative inaugurated in 1997 to develop and implement an electronic statewide school information system to facilitate the exchange and reporting of student information by school districts to CDE. CSIS provides a vehicle for collecting data extracted from the student information systems in each participating district. Instead of making aggregate reports to the state, districts report student-level data to CSIS that are then aggregated into required state reports, enabling the electronic transmission of student records among districts and to postsecondary institutions. CSIS requires participating districts to use standardized data elements and student information systems that meet CSIS standards for interoperability. Thus CSIS was originally envisioned as the vehicle through which California would develop comparable information systems in all districts, but participation was made voluntary, not
mandatory. As of 30 June 2006, 263 districts (about a quarter of the state’s districts, representing about half of student enrollments) used CSIS to electronically transmit data required for three state collections: CBEDS, Language Census (R-30), and Student National Origin Reports (SNORs).7

Despite all the efforts to make education data available and understandable, the multiplicity of data sources in California can be confusing. The same CDE-collected data feed the various sites described above, but the distinctions between the sites are not clear to a casual user, and no road map exists to guide the user in choosing which source to consult for what kinds of data. Furthermore, the SARC has over time become so unwieldy and opaque that the state superintendent of education himself quoted a UCLA study comparing SARCs unfavorably to IRS forms and Windows XP driver installation instructions (O’Connell 2006). CDE redesigned the SARC template for the 2006–7 school year to make it more user friendly. In September 2006, however, Governor Schwarzenegger vetoed a bill (SB 1510) supported by Superintendent O’Connell that would have further clarified and simplified the SARC.8

4. STILL A LONG WAY TO GO

Despite the plethora of education data available in California and efforts to improve the management and usability of those data, the state still has a long way to go to provide the information its policy makers and citizens need to understand how students and educators are performing and how resources might most effectively be allocated. Among other things, California currently lacks the ability to:

- Track students individually and over time and link them to performance scores and outcomes such as graduation;
- Track teachers individually and link them to students taught, to their preparation programs, and to the professional development they receive;
- Provide teachers with student histories and performance indicators; and
- Link resource use and performance.

Moreover, the state has only fragmentary efforts underway to link K–12, post-secondary, and post-school records, severely hampering efforts to understand what happens to students and why as they move through the educational system and into the workforce.

---

7. Data are from the CSIS Web site: www.csis.k12.ca.us/project-management/local-projects/default.asp.
To illustrate how the absence of such capabilities inhibits decision making, the Data Quality Campaign has identified six “priority questions” facing policy makers and educators and has indicated how many states have data systems capable of answering them. California is not capable of answering any of the questions with its currently available data. Of the ten essential elements the DQC has identified for student longitudinal data systems, California in 2006 ranked below average among the states, having only four elements. For its Technology Counts 2006 report, _Education Week_ (2006) also surveyed states, among other things inquiring whether they had trackable identification systems for teachers. Forty-three states reported having such systems; California does not. Some states not only have such systems but have developed the capacity to link student and teacher data.

California has several data initiatives under way that promise to fill some of the gaps in its current education data capabilities; but as presently envisioned, even when (or if) completed, they will leave some significant holes. The previously mentioned CALPADS will enable California to comply with the requirements of NCLB. Almost all students have now been assigned unique identifiers that will allow their records to be linked over time, and a vendor to develop the system should be chosen in late 2007. The CALPADS system is scheduled to be rolled out in school years 2008–9 and 2009–10.

At the insistence of the California Department of Finance (perhaps among others, though not the legislation that authorized the data system), CALPADS as currently envisioned will include only those data elements required by NCLB. This will leave data of potential interest to education policy makers and researchers, including some connected to other federal programs such as vocation education and special education, out of CALPADS and will limit the ability of users to assess the value of educational investments and programs. Furthermore, there are as of now no plans to use CALPADS as a data warehouse. “That is, the system will not maintain the detail and summary aggregations of student-level transactions in a manner specifically designed to support elaborate and complex data selection queries for decision support analysis and research” (California Department of Education 2004, p. 6.5).

First steps are also being taken on a longitudinal teacher data system to bring together data currently housed at CDE, CCTC, and EDD. Legislation in 2005 (SB 77, the 2005–6 budget act) authorized a feasibility study on a longitudinally linked teacher data system with the single stated intention of

---

supporting high-quality program evaluations, especially for state investments in teacher professional development and retention programs aimed at improving teacher quality. CDE and CCTC worked on developing a system of unique teacher identifiers during 2006–7 (Stickel 2006).

In August 2006 the legislature passed SB 1614, a bill calling for a “comprehensive state education data information system in the department [CDE] that includes information regarding the teacher workforce.” The bill includes within the California Education Information System both CALPADS and a California Longitudinal Teacher Integrated Data Education System (CALTIDES) based on the feasibility study for the teacher data system. The legislation appears to open the door to the possibility at some future point of linking student and teacher data, a linkage opposed (according to several interviewees) by the Department of Finance and the powerful California Teachers Association (CTA). Despite its historical opposition to a teacher database and to linking student and teacher data, SB 1614 garnered CTA support, according to interviewees, because it contains what the union apparently views as a “firewall” against using CALPADS and CALTIDES data for teacher evaluation or pay.

California has two initiatives underway that begin to bridge the gap between its K–12 and postsecondary data systems: Cal-PASS (California Partnership for Achieving Student Success), a voluntary data-sharing arrangement among consortia of California elementary schools, secondary schools, and colleges and universities; and CCCTran, an effort of the California Community Colleges system to permit requesting, viewing, and transmitting academic transcripts among authorized educational institutions and their trading partners. Both initiatives are limited and largely involve (at the postsecondary level) the community college system rather than the state colleges or universities. Linkages are hampered by the fact that there is no unitary data system for California's public higher education institutions, with each of the three sectors (California Community Colleges, California State University, University of California) having its own policies and practices for information collection and dissemination.

5. WHAT SEEMS TO BE HOLDING CALIFORNIA BACK IN DEVELOPING STATEWIDE K–12 DATA SYSTEMS?

Despite recent progress, individuals interviewed for this article expressed skepticism about whether the current initiatives to develop statewide longitudinal data systems would be carried out as envisioned or on time. What seems to be holding the state back?

One factor appears to be the absence of a “culture of data” in California that emphasizes the necessary connection between good data and school improvement efforts. Policy makers in some other states recognized and acted
on this link many years ago. Thus Florida, for example, began developing common data definitions for education data as early as 1976 and was a pioneer in developing statewide databases when it began planning for its current information system in 1984. Today its education data system (now enhanced by a data warehouse that includes so-called P-20 longitudinal coverage of students and teachers) is widely recognized as the most extensive in the nation. The development of high-quality statewide education data in Florida was fostered by a longtime legislative emphasis on making data-driven decisions and by legislative mandates requiring (and funding) the creation of robust data systems. Similarly, lawmakers in Texas included information systems in the major 1984 school reform legislation that resulted from the recommendations of the Perot Committee. They mandated creation of a Public Education Information Management System (PEIMS) to collect and make available information on schools and students. PEIMS (which collected data for the first time in school year 1987–88) promoted data quality by defining what data were to be reported by school districts and in what formats. In 2001 the legislature began funding the Texas Prekindergarten–16 Public Education Information Resource (TPEIR), a collaboration among the state’s K–12 education agency, higher education coordinating board, and teacher certification agency to build an integrated data repository that would provide a comprehensive view of public education in Texas.

A culture of data was probably easier to foster in states such as Florida and Texas with comparatively strong traditions of state direction of education, in contrast with California’s greater emphasis on local control. Florida also has the advantage of having only sixty-seven school districts (though Texas has over one thousand). It is worth noting, however, that Iowa, a state with local control traditions as strong or stronger than California’s (and with a large number—371—of school districts to accommodate), has in recent years implemented Project EASIER (Electronic Access System for Iowa Education Records) and created a statewide student identifier system and a teacher identifier system. State and area education agency personnel are now discussing the desirability of augmenting the state’s “live” information systems with a data warehouse of cleaned, stable data to enhance reporting and enable analyses of education issues such as the impact of mobility on student learning and the effect of early intervention services.

A corollary to the absence in California of a state-level emphasis on the importance of education data has been the absence of incentives for school districts to care about the substance and quality of the data they provide to the state. In California, data flow up from districts through all the data collections that were mentioned earlier, but very little (with the notable exception of student test score data) flows back down to them except via the School Accountability
Janet S. Hansen

Report Card (and its shortcomings have already been noted). In Florida, by contrast, there has been a strong emphasis on building local support for statewide data systems by taking seriously a quid pro quo: those who supply data need to get something back. Therefore Florida officials have used statewide data collections to reduce the burden on local officials, to foster the exchange of data among districts and between districts and higher education institutions, to give teachers data they can use to improve classroom instruction, and to provide feedback to high schools and community colleges on how their students have fared as they moved to the next educational level. The state also emphasizes good communication with districts and helps them develop reliable data by providing financial assistance to support their information systems.

A number of individuals interviewed for this article also indicated that the development of statewide data systems in California was being held back by reservations at the Department of Finance (DOF), which is responsible for the budgeting and control of information technology expenditures throughout state government.

Some of this hesitancy is understandable, given plenty of evidence from the business sector as well as from public agencies at the state and local levels that big data projects not infrequently go seriously and expensively awry.11 DOF has long had concerns about CDE’s data management capabilities, as illustrated by its commissioning of the 2002 outside review by MGT of America. Moreover, provisions of the California state constitution barring unfunded mandates have made DOF and other state officials wary of imposing data requirements that could cause school districts to insist on the need for state funds to modify their local information systems.12

Despite these legitimate concerns, however, some supporters of an improved education data structure for California view the DOF as excessively cautious and a stumbling block to improving the system. The Legislative Analyst’s Office (LAO) has publicly commented in the past on the delays at DOF (and at the Department of Information Technology) in approving the scope of work for a project related to CSIS (California LAO 2001). It took DOF fourteen months (from August 2004 to October 2005) to approve the feasibility study

---

11. An Organisation for Economic Co-operation and Development (OECD) policy brief cites estimates that fewer than a third of government and industry IT projects in 2000 were successful with regard to budget, functionality, and timeliness; almost a quarter were cancelled (OECD 2001). The Washington Post (Eggen and Witte 2006) recently reported on “The FBI’s Upgrade That Wasn’t: $170 Million Bought an Unusable Computer System.” Closer to home, North Carolina recently encountered a number of problems in replacing a twenty-year-old student information system with a new data system called NCWise, and Idaho shut down its effort to build the Idaho Student Information System after encountering large unexpected cost projections. Idaho’s experience has been analyzed in a new report from the Idaho Legislature Office of Performance Evaluations (2006).

12. California Constitution, Article 13B, Government Spending Limitation Sec. 6(a).
EDUCATION DATA IN CALIFORNIA

report for CALPADS, without which CDE could not move ahead to preparing a request for proposal for the system. (In a perhaps hopeful sign, DOF took only two months to approve the feasibility study report for the teacher data system in 2006.) Individuals interviewed for this article (not CDE employees, it should be noted) commented on the long-standing “bad blood” between DOF and CDE and on the “hundred ways of saying no” that DOF has used to slow down the CALPADS development.

Most recently, DOF and CDE have been at odds on the most effective way to fund local districts to upgrade and maintain their capacity to provide reliable student-level data (including student identifiers). CDE and LAO wanted funds included in the 2006–7 budget to provide “CALPADS incentive grants” to districts to support local data activities and compensate for the work involved in maintaining the new student identifier system, the quality of which will be essential to the successful implementation of CALPADS. In exchange for funding, CDE would have required districts to meet several quality measures related to the identifiers. The department wanted incentives of $5 per pupil; it and LAO supported provisions in the senate and assembly budget bills that would have provided $2.50 per pupil, costing $15 million. These funds were removed from the budget act right before it was finally approved. Apparently DOF believed that funding for maintaining student identifiers and for other data quality enhancements related to CALPADS need not be provided to districts until CALPADS is operational. In the 2007–8 budget cycle the governor included $65 million in his final budget request to help school districts with CALPADS pre-implementation and training, but the legislature failed to approve this funding.

These disagreements, finally, seem reflective of a more general issue: as one interviewee described California’s efforts to improve its education data systems, “The stakeholders assert they are supportive, but they’re really not invested in making it happen.” Despite recent progress on the student and teacher longitudinal data systems, if California continues to approach the development of its education data infrastructure in such a half-hearted fashion, it will miss out on the enormous potential that modern technology offers for using data to enhance teaching and learning.

6. CHALLENGES FOR CALIFORNIA

California, as with all states, will have to develop an education data infrastructure compatible with its unique circumstances and needs. While a visionary system like Florida’s can provide inspiration for discussions about what kind of data system California might ultimately want, the immediate task is to effectively pursue the constructive initiatives already underway: CALPADS, CALTIDES, better linkages between the K–12 and higher education sectors,
and improved data management through continuing efforts at CDE to develop common data definitions, reduce data collections where possible, and make it easier for localities to provide more accessible and transparent data for users. Though California’s late start on some of these efforts will keep it behind more pioneering states, one clear lesson from past failures of major information technology projects is, as a veteran of North Carolina’s difficulties implementing a new statewide information system put it, “Don’t bite off more than you can chew” (Dillon 2006). This is especially true given what still needs to be learned. One of the federal government’s goals for its longitudinal data grants program is the identification of best practices in “designing, developing, implementing, and utilizing statewide longitudinal data systems.”

California needs to address several interrelated challenges if its current data system “bites” are to be successfully swallowed and digested.

Leadership
California needs influential education data “champions.” It seems unlikely that the state can build data systems capable of supporting data-driven policy and funding decisions without strong and long-term support from state leaders who can reduce inter- and intradepartmental and intersectoral rivalries and ensure that funding and other necessary support are available. Decisions such as how extensive CALPADS should be (i.e., limited to NCLB-required elements or incorporating other data desired by California educators and policy makers) deserve wider discussion than they are now getting.

Funding
One way that lukewarm support for education data system development in California has manifested itself is in inadequate funding to carry out system implementation at both the state and local levels. CSIS was originally intended to be deployed to all school districts over a five- to six-year period, but inadequate funding prevented the full rollout. In January 2006, CSIS staff estimated that $42 million would be needed to fund all the remaining districts at the “legislated compensation rate” normally given to help them fully participate in the system (CSIS 2006, p. 14). Instead, $31 million was budgeted for what an assembly budget subcommittee dubbed “CSIS light”—a less intensive version to help nonparticipating districts with hardware and software improvements necessary to prepare for CALPADS but not including “records transfer or any of the other long-term functional CSIS goals, which are also the goals of

---

EDUCATION DATA IN CALIFORNIA

CALPADS. As noted earlier, funding for ongoing maintenance of student identifiers and other CALPADS pre-implementation activities was rejected. According to the feasibility study reports for CALPADS and CALTIDES, system development costs will largely be paid with federal funds. There is little in this recitation to suggest that California is becoming significantly more willing to invest its own money to get information to help policy makers in deciding how best to invest the $45 billion they spend annually on public education. This reluctance to invest raises questions in this writer’s mind about, among other things, whether the ongoing, long-term commitment to communication and training (for both state agency and local personnel) that is critical to the successful implementation and utilization of complex data systems will be made, and whether funding for new data initiatives will be driven by available federal dollars rather than a realistic assessment of the costs of effective development and implementation.

Access to the Data
Stakeholders need to be able to access and use education data if the new California initiatives are to fulfill their potential. While current resources like DataQuest, Ed-Data, and SchoolMatters do make a fair amount of information available, often in user-friendly fashion for those interested in descriptive data on particular schools, districts, or counties, the more far-reaching benefits for policy makers will come from studies (conducted by analysts both inside and outside government) aimed at determining which educational programs and what kinds of resource use really make a difference in improving student achievement and narrowing achievement gaps. These analysts will need access to student-level data.

States take different approaches to fostering the use of their education databases for research and evaluation. In Florida, for example, the state investment in the Education Data Warehouse has put student-level data in a form suitable for research use, and staff in the Department of Education and the legislature’s Office of Program Policy Analysis and Government Accountability undertake many studies themselves. Outside researchers can apply to the Florida Department of Education for access to data from the warehouse and must agree to abide by a number of conditions relating to security, training on the structure and content of the warehouse, and other matters. Agency staff must be allowed to review outside reports before they are released. In North Carolina, where cuts to the state department of education have severely limited the capacity of staff to undertake all the tasks necessary to make data usable

14. No longer available www.assembly.ca.gov/acs/committee/ca/hearing/may%2022%202006%20part%20ii%20public.doc.
for analysis, a North Carolina Education Data Research Center has been established at Duke University to create a portal to the immense amount of education data that the state collects. A Center Advisory Board of members from the sponsoring consortium (Duke, the University of North Carolina, and the NC Department of Public Instruction) is responsible for setting access policies and reviewing applications to use the data. In Texas, researchers have applied more informally for access to student-level data through the Texas Education Agency (TEA), but a law (HB 1, passed in June 2006) called for the creation of up to three Education Research Centers at TEA, the Texas Higher Education Coordinating Board, or colleges and universities “to conduct research for the benefit of education in this state.” The first center awards were made in mid-2007. In New York, school finance analyses of interest to state policy makers have been carried out for a number of years under agreements between the state Department of Education and a university-based Education Finance Research Consortium. Researchers in the consortium have undertaken the work of putting together the student and teacher files needed to conduct consortium studies.

The analytical potential of new data systems is most likely to be realized under conditions in which funding is available to develop research-friendly data files, and outside researchers as well as state employees can have reasonable access to them. California might consider how outside, disinterested stakeholders could be involved in setting and carrying out policies relating to data access, especially for outside researchers, to ensure that appropriate requests are granted. If California does not develop a data warehouse or other means to prepare student- and teacher-level data so that it can be used for research and analysis, the state might consider sponsoring (and perhaps contributing funding to) an outside agency to do so.

**Final Comments**

Because this article was written as part of a project on school finance and resource allocation, a few final comments about finance data are in order. It was noted earlier that as of 2003–4, all school districts are required to report financial data to the state using a common accounting structure. The SACS includes subcodes for school-level financial reporting, although such reporting is voluntary, and common data definitions for school-level elements have not been developed. It is reasonable to ask whether California needs school-level financial data to answer important questions about how resources spent relate to student performance and to identify resource gaps that may be related to achievement gaps.

The effort required to develop school-level data does not at this time seem justified, despite well-known limitations with finance data that can be
disaggregated only to the district level. A host of problems has been identified in developing school-level finance data reporting systems, the most significant of which is the immense burden such systems place on local administrators (Picus 2001). Even if California did not already have enough new data initiatives on its plate, serious questions would exist about whether the benefits of attempting to obtain school-level data come anywhere near their costs.

A few years ago, the Joint Legislative Accountability and Review Committee (JLARC) in Washington State considered but rejected the idea of collecting school-level expenditure data, concluding that expenditures on staff (which account for a very large percentage of school and district spending) were far easier to obtain and could serve as a useful proxy for school-level expenditure data (Washington State 1999). A 2005 JLARC noted the potential advantages of school-level financial data in improving accountability, transparency, linkages with performance, and equity and observed that districts might find it advantageous to track expenditures to the school level in their own reporting systems (Washington State 2005). The costs and administrative complexities of attempting to develop consistent, comparable school-level financial reports on a statewide basis, however, remain daunting. Instead, California should concentrate on developing a reliable and detailed employee data system, including information on salaries and benefits, that can be linked to schools, students, and programs.15 Such a system could support analyses of many of the central questions about resource allocation without requiring whole new school-based financial reports.

REFERENCES


15. CALTIDES will include all certificated staff (teachers, administrators, and pupil services staff). California policy makers may want to consider the benefits and costs of including other education personnel in their data collection system, as Florida does.


