EDFACT

May 2002

EdSource®

Clarifying Complex Education Issues

California's Student Data System:

In Need of Improvement

New federal education legislation, the No Child Left Behind act, is likely to have a significant impact on California education policy and more specifically on how the state collects and uses data. Two features of this act are especially noteworthy:

- ✔ Its emphasis on the use of test-score data to reward and sanction schools, as well as to provide information to parents and other stakeholders; and
- ✔ Its required use of "scientifically-based" educational programs.

The testing provisions require annual student assessments and public reporting of results for subgroups of students within each school. The mandate for scientifically-based programs requires showing that program participation improves student performance. Common to both of these provisions is the need for valid and reliable student achievement data.

The federal legislation reflects a more general trend among policymakers to want to use student performance data to evaluate the effectiveness of schools and of various educational programs. California currently uses test-score data from its Standardized Testing and Reporting (STAR) program to rank schools to determine which are subject to rewards or sanctions and to evaluate programs such as class size reduction. However, because the state's data system does not include a unique identifier for each student, educators, researchers, and policymakers cannot take advantage of the full scope of test-score information. An accurate identifier would make it possible to link student records over time and track the progress of individual students or selected groups of students as they move through the education system.

This report looks at how California would benefit from such a system, provides examples of how states and districts have used longitudinally linked data, and discusses the current status of an ongoing effort in California.

Longitudinal data would help California track its progress

As California's assessment and accountability systems evolve over the coming years, the ability to link individual student records over time would have several advantages.

First, it would enhance the accuracy of school-level achievement information by highlighting the impact of student mobility. Year-to-year changes in school-level test scores are often interpreted as indicators of the effectiveness of the school, but some of these changes result from the fact that different students take the test from one year to the next. A data system that links student records over time provides a way to compare scores for only those students who remain in the school during the academic year and from year to year, so that score gains or losses for schools with highly mobile populations are not influenced by students who were not present in the school long enough to have been affected by the instruction there.

Second, longitudinal data can help disentangle the impact of schools and teachers from the effects of factors that are not under the control of the school, such as family characteristics or outside-of-school enrichment opportunities. For example, do students with different background characteristics (e.g., poverty levels) show different rates of growth over time? Measuring test-score gains for individual students would provide a more precise estimate of the effectiveness of the education system.

Third, linking student records would allow researchers and educators to track the growth of individual students and answer such questions as: Do initially high-achieving students maintain their high levels of performance as they progress through school? Are some schools more effective than others with students whose initial scores are low?

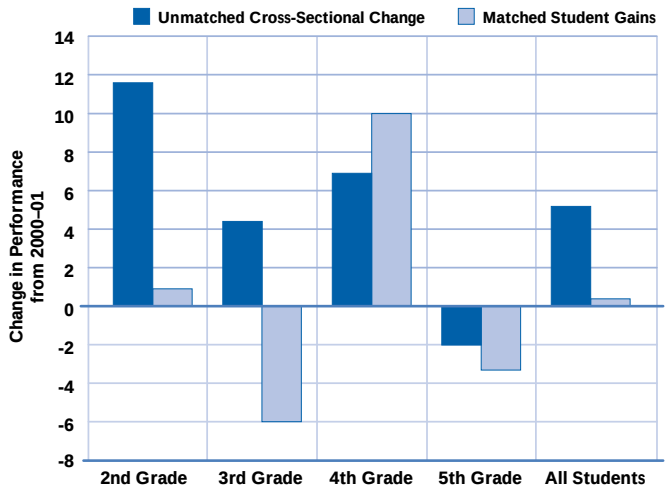
Finally, a longitudinal data system could provide more accurate evaluations of the effectiveness of educational programs by capturing how much time each student has spent in the program. Currently it is difficult to evaluate a program's effect on student learning because the data system does not allow researchers to determine each student's length of exposure to it.

The recent Class Size Reduction (CSR) Research Consortium's evaluation of California's K–3 program illustrates this problem. Much of the enthusiasm for class size reduction in California was based on the results of an experimental study in Tennessee. When researchers there followed K–3 students who were in reduced-size

EdSource thanks The Stuart Foundation for underwriting the research and development of this EdFact.

Figure 1

LAUSD's comparison of cross-sectional and matched student scores on the Stanford-9 reading test yield different pictures



At this elementary school in Los Angeles Unified School District (LAUSD), the unmatched cross-sectional data (e.g., comparing second graders in 2000 to second graders in 2001) show dramatic gains in reading. However, when the performance of the 2001 second graders was compared to their own individual results from first grade, the improvement almost disappears, suggesting that the results obtained when comparing cohorts may be due primarily to differences in student ability across cohorts. (LAUSD administers the Stanford-9 reading test to its first graders.) Variations are evident across the other grade levels as well. In other schools, the reverse results have been seen, with the matched longitudinal data showing greater gains than the unmatched, cross-sectional approach.

Data: LAUSD data for one elementary school EdSource 5/02

classes, they found significant positive effects on student achievement. Unfortunately, CSR researchers in California were unable to compare their results against Tennessee's because they could not track individual students to determine how many years they participated in reduced-size classes. Instead, they had to rely on a combination of aggregated district-level data and statewide data for cohorts of students. In part because of the inadequacies of California's statewide data system, the CSR evaluators' ability to assess the impact of the program on raising student achievement is weaker than it would be if longitudinally linked data were available.

Other statewide evaluations, including that of charter schools, face the same problem. Many researchers have addressed the issue by asking a sample of districts to provide the linked data, but this increases the cost of the evaluation, reduces the ability to generalize about results, and does not include students who moved from one district to another during the period being evaluated. Philip Kaufman of MPR Associates, in a forthcoming report from the University of California Linguistic Minority Research Institute (www.lmri.ucsb.edu), explores the potential benefits of longitudinal data for the purposes of research and evaluation. Kaufman also provides examples of the kinds of questions the state could answer if it had suitable data.

Other states and organizations use longitudinal data

A number of other states, as well as many California school districts, routinely collect data in a form that permits them to link individual student records over time. An organization called Just for the Kids (JFTK) illustrates what can be done with this type of data. JFTK operates a website that allows users in several states to access information about the performance of schools and to compare any school's performance with other schools that have similar student populations. One of the goals is to identify consistently high-performing schools that can serve as models of best practice. California also identifies "similar schools," but the comparisons are less useful because changes in scores over time may result from changes in the students who took the tests. JFTK uses student-level data to control for the effects of student mobility.

JFTK provides an analysis of what it would take for each state to replicate its approach. Along with the creation of a student identifier, JFTK recommends that California maintain records on students who are absent or exempt from the state test in order to ensure a fair comparison among schools. The group also says California needs to link its spring test-score data with fall enrollment data to determine which students remain enrolled in the same school (see www.just4kids.org/US/California.asp).

The Los Angeles Unified School District's Program Evaluation and Research Branch (PERB) has used longitudinal data to publish school-level test scores that include only those students who remained in the system from year to year. The scores follow the same group of students from one grade level to the next. According to Ted Bartell, the director of PERB, the district believes that matched student-level gains provide a more accurate indicator of school performance than does growth on the state's Academic Performance Index (API). The latter uses an unmatched, cross-sectional approach to compare this year's second graders to last year's second graders. A look at scores from one elementary school illustrates that performance gains or losses are different depending on the comparison used. (See Figure 1.) LAUSD officials have found these differences compelling and therefore include matched reading gains as one of four indicators in their new local accountability system.

Bartell notes that the matched gains help identify schools that need support and assistance or that should be recognized for exemplary performance. Some of these schools would be overlooked if only school-level API data were used.

Improvements to California's data system are in progress

As the example in Figure 1 illustrates, longitudinally linked student data provide opportunities for powerful and informative analyses of school and student achievement growth. If

California implemented such an education data system statewide, it could improve the quality of information about school and program effectiveness and about what works and does not work for individual students. This would help create a "value-added" analysis of the state's schools, enabling educators and policymakers to estimate the effectiveness of various educational experiences on student progress over time.

Some of the necessary components for this system are already in place. California currently administers two types of achievement tests to students in grades two through 11, as well as a high school exit exam that is required for graduation beginning in 2004. The standardized national Stanford-9 tests are "vertically equated"—that is, the "scaled" score means the same thing across grades. Changes in scores provide an indication of growth, or lack of it. However, the increasingly important standards-based tests that match California's curriculum are not vertically equated. As a result, some types of value-added analyses could not be conducted for the Standards Tests, although linking student scores on these tests over time could still be useful for other kinds of analyses. The state receives scores on all of these tests each year, but without a unique, permanent student identifier. The state can therefore examine achievement by school, but it cannot link individual student records to follow specific groups of students over time.

In addition to test scores, the California Department of Education (CDE) collects extensive demographic and financial information from schools, districts, and county offices of education (COEs) through the California Basic Education Data System (CBEDS). This includes demographic information about students and the Personnel Assignment Information Form (PAIF), which collects staffing information. Again, the school is the lowest level for reporting these data; i.e., individual students are not identified.

Foreseeing the need for an improved data collection system as well as streamlined reporting, state policymakers created CSIS (California School Information Services program) in Assembly Bills 107 (1997) and 1115 (1999). The stated purpose of CSIS is to enhance the ability of school districts to collect data, simplify the transmission of school or district information for the multiple state and federal reports, and enable the electronic transfer of individual student data from school to school and eventually from school to higher education. Once the CSIS office was established in Sacramento, staff quickly learned that even in the high-tech age, assembling basic education data and transmitting it safely are difficult challenges.

The CSIS system is voluntary (i.e., not state-mandated), with incentive funding for districts to support the formidable task of getting started. In 2002, its third year of operation, CSIS consists of 12 consortia with 219 school districts and COEs (20% of the state total) serving more than 2.2 million students (37% of the total). The CSIS estimate is that 3.3 million students in 319 local education agencies will have unique identifiers by the beginning of the 2003–04 school year.

Student privacy remains a serious concern

A crucial issue concerns the confidentiality of student information. The state must be able to address the privacy issues to the satisfaction of many stakeholders, from students and parents to the Legislature and State Board of Education. This is particularly important if external organizations, such as researchers conducting program evaluations, will use the data.

Other states have student identifiers and have found ways to ensure student privacy. For example, student identifiers can be encrypted or stripped from the data file before it is given to researchers or anyone else outside the state. The state can require anyone using the data to sign a confidentiality agreement and can develop criteria to determine who may have access to the data.

CSIS is already working on the transfer of student records, including test scores, demographic and health information, and enrollment history. The encrypted data will be sent from the student's prior school directly to the new one. Further, by fall 2002 more than 100 local education agencies will replace their CBEDS reports with electronic CSIS submissions. (Additional information about CSIS is available at: www.csis.k12.ca.us)

California needs to address data issues

The current lack of longitudinally linked data in California has important implications for policymakers seeking to understand which schools and which programs are effective; for parents, teachers, and administrators who want to know how well students are progressing; and for anyone who is interested in understanding the status of public education. And the new federal requirements add an extra sense of urgency. (For more information, see ESEA 2001 Policy Brief from the Education Commission of the States: www.ecs.org)

To be able to follow its students' progress over time, the state will have to take several important steps. The fundamental task is to create a data system that includes a student identifier, making it possible to match individual student test-score data with enrollment data as well as with the state's existing CBEDS databases. California has begun this process with CSIS. But it remains unclear whether CSIS, as currently configured, will adequately address the state's data needs, or what changes and additions policymakers might call for.

For example, the Office of the Secretary for Education (OSE) is studying data collection needs related to the API that ranks schools. The API is supposed to include teacher and student attendance, dropout rates, and graduation data in addition to STAR scores. Currently that information either is not collected uniformly at the school level or is not considered

Tennessee uses teacher identifiers to further extend its analysis

The state of Tennessee has received recognition for its Tennessee Value-Added Assessment System, or TVAAS. In addition to linking student records over time, Tennessee maintains a database that links each student to his or her teacher each year—a politically sensitive step. Statistician William Sanders and colleagues developed a statistical methodology that allows them to estimate the effects of teachers on student achievement growth. The work has led to a number of compelling findings, including the following:

- ✔ Teacher and school effects on student learning are quite large, and they tend to outweigh socioeconomic influences.
- ✔ The size of student gains was not related to initial achievement; that is, initially low-performing students gained at roughly the same rate as high-achieving students.
- ✔ The sizes of school effects were unrelated to the racial/ethnic composition of schools or to whether the school was in an urban, suburban, or rural neighborhood.

The TVAAS system has also provided useful data for teachers to use as they think about how to improve their own teaching. Replicating these analyses in California would require linking student and teacher records. Because understanding the effectiveness of teachers may be of greater interest to districts than to the state, it may be desirable for these links to be maintained at the district but not the state level. Even in the absence of teacher links, this type of value-added analysis could be done down to the school level if the state had longitudinally linked data on students.

See www.k-12.state.tn.us/assessment/scores.asp for additional discussion of TVAAS.

valid because of unreliability in the collection procedures. Further, the student performance ranking uses cross-sectional data that looks at growth between, for example, this year's second graders and last year's second graders. The OSE report could make a variety of recommendations, including a switch to a value-added measure of school performance. Meanwhile, the state Department of Finance is preparing a recommendation about the data management practices of the CDE, including merging several reports through CSIS.

When these various recommendations are presented, the Legislature could make considerable changes to California's system of education data. Legislation was also introduced in 2002 (Senate Bill 1453) to enable the CDE to develop and maintain a system for longitudinal achievement data, including a "unique pupil identification number," to be developed in consultation with CSIS.

To the extent that it can be adapted or expanded to meet these various data needs, CSIS certainly gives California a head start. The independent Legislative Analyst's Office (LAO) says CSIS is "the most capable entity" to address many of the data needs noted above. However, it is not yet fully operational, even within its current configuration and objectives. The Legislature's goal is for 90% of the districts and COEs to be participating by 2004–05. In its analysis of the 2002–03 proposed budget, the LAO points out that the CSIS operation is underfunded for reaching that goal: "The Legislature faces a trade-off between the speed at which the program is completed and the annual funding level that the Legislature must provide." The LAO recommends that the Legislature adopt an intended completion date and "align funding accordingly." Even if that happens, the value and strength of CSIS will depend on the level of participation statewide, which could be an important issue if the program remains a voluntary one.

Policymakers and researchers note additional uses for longitudinal data. An important one is linking K–12 educational data with data on students' postsecondary education participation and performance, perhaps continuing into the workforce. Another is the possibility of adding teacher identifiers to the system, such as was done in Tennessee. (See the box on this page.)

Better statewide information would help California improve its analysis of test scores and its educational evaluation at both local and state levels. Linking participants' test results with program data would help make clear the value added to student learning by specific programs.

Developing a sophisticated longitudinal data system raises technical, economic, and political challenges. Overcoming them will require commitment and cooperation from a number of stakeholder groups. If successfully achieved, the new system should prove worthwhile by providing reliable data, streamlining the data collection process, and improving the state's analytic capabilities.

EdSource is a not-for-profit 501(c)(3) organization established in California in 1977.

Independent and impartial, EdSource strives to advance the common good by developing and widely distributing trustworthy, useful information that clarifies complex K–12 education issues and promotes thoughtful decisions about California's public schools.

For more information about our organization, a catalog of our publications, or details on how to subscribe to our Information Service, please contact us at:
4151 Middlefield Rd., Suite 100
Palo Alto, CA 94303-4743
650/857-9604 Fax: 650/857-9618
E-mail: edsource@edsource.org
Or visit us on the web:
www.edsource.org

This EdFact has been prepared by:

Laura Hamilton,
Behavioral Scientist, RAND

With support from

Mary Perry, project manager/editor
Penny Howell, writer/editor