Title
Connecting the Dots: Harnessing Integrated Data to Improve Education in California

Authors
Phillips, Meredith
Reber, Sarah
Rothstein, Jesse

Publication Date
2019-02-01

CALIFORNIA POLICY LAB

POLICY BRIEF

February 2019

Connecting the dots: Harnessing integrated data to improve education in California

The following brief is a summary of a September 2018 report co-authored by California Policy Lab faculty affiliates Meredith Phillips, Sarah Reber, and Jesse Rothstein titled "Making California Data More Useful for Educational Improvement." The full report was released by the "Getting Down to Facts II" project, which aims to bring evidence to bear on the conditions of California education and to guide future policy. We would like to thank the California Department of Education for their thoughtful feedback.

Executive Summary

Crafting successful education policy requires evidence. Generating this evidence requires data that do not currently exist, or are not used to their full potential, in California. An integrated educational data system would arm teachers, schools, districts, service providers, and policy makers with the data they need to support students from early childhood through higher education completion. Such data can enable continuous improvement efforts driven by teachers, school leaders, and districts, facilitate coordination among human services agencies and educators to care for children's in-school and out-of-school needs, and enable policy makers to create, identify, and scale programs proven to improve educational outcomes.

In the last decade, the California Department of Education created the CALPADS data system, which assembles student-level data from districts throughout the state into a single longitudinal K-12 data system. This key step forward has enabled numerous important analyses not previously possible. However, California still remains behind other states, many of which have benefited from integrated educational data systems that extend well beyond K-12. An absence of state-level coordination and strategy has led to duplication of effort, ineffective service delivery, and holes in the web of state-provided services.

California can continue to expand on its efforts by integrating CALPADS with data from other agencies that collect relevant information about students' school- and non-school experiences. With significant but achievable effort, California could create a linked longitudinal data system that would allow it to design and run the world-class education system that Californians need.

California can establish an integrated data system by:

1. **Expanding data linkages,** built on existing infrastructure
2. **Championing integration through strong political and agency leadership,** both by articulating the immense value of integration and convening stakeholders to build a collaborative system
3. **Establishing a governance structure** that involves leaders from appropriate agencies to develop rules on access, use, privacy, security, and IT infrastructure
4. **Building infrastructure** to support linked data systems, and the capacity to use them

The Power of Linked Data

Data are traditionally siloed – districts collect data about their students' in-school performance, colleges maintain data about the students who enroll there, and social service, criminal justice, child welfare, and other agencies have information about their own program participants. Integrated data systems link data from these and other sources, making it possible to measure the prior circumstances and subsequent outcomes of the children that each system serves. An integrated educational data system would link administrative data routinely collected about students by schools, districts, higher education systems, and service providers. This system would also be longitudinal, tracking students over time.

A robust system would link K-12 student demographic, attendance, special program participation, course enrollment, and state test score data with:

- Enrollment, financial aid, academic outcomes, and completion data from California's three higher education systems;
- Information on availability, enrollment, and outcomes of child care and pre-kindergarten programs;
- Program participation, service utilization, and arrest and conviction information from county and state human services, health, and criminal justice agencies; and
- Adult income and wage data from databases maintained by the Franchise Tax Board and the unemployment insurance program at the Employment Development Department.

Such a system would paint a complete picture of the whole student. These linked data could empower educators, administrators, service providers, and policy makers to serve California's students in a more personalized, efficient, and effective manner.

An integrated educational data system can:

- **Promote more coordinated service delivery**
- **Support individualized learning strategies** and continuous improvement efforts driven by teachers, school leaders, and district officials
- **Identify organizations that stand out** as effective and replicate their success
- **Provide a platform that supports evaluations of the effectiveness** of policies and programs

1) Facilitate more coordinated service delivery.

Districts could use data on social service program participation to coordinate efforts with local service providers to ensure, for example, that youth in foster care receive appropriate educational and non-educational services. Welfare and nutrition programs, criminal justice authorities, and health providers could use data on educational enrollment and outcomes to coordinate service delivery and to measure success of their programs in supporting student learning. A juvenile justice

diversion program, for example, might assess the progress of the young people it serves by measuring their attendance and progress in school.

Case study: Allegheny County, Pennsylvania connected academic and human-services-related data to help school staff understand students' mental health and child welfare involvement while helping social workers and caseworkers understand children's school performance, attendance, and disciplinary history.

2) Support individualized learning strategies and continuous improvement efforts driven by teachers, school leaders, and district officials.

Teachers could use data on individual students' past performance to customize instruction and interventions such as English language instruction or reading assistance. School and district leadership could use data to inform classroom groupings, staffing decisions, and other curricular and resource-allocation decisions. Access to longer-term student outcomes, such as college-going, college-completion, and career success rates, can help leaders measure the effectiveness of current programs and guide improvements. California's colleges could also use a linked data system to inform outreach campaigns aimed at identifying high-performing students who might not be contemplating college. They might also use data collected from students' high school transcripts to streamline admissions procedures, course placements, advising, and even financial aid awards. Institutions that accept transfer students might use the information in community college records for similar purposes.

Case study: The University of Chicago Consortium on Schools Research used individual-level, longitudinal data from the Chicago Public Schools to develop a 9th grade measure of course completion and course failure that was a strong predictor of whether students would fail to graduate from high school. The school district incorporated this "on-track" indicator into its data systems and promoted initiatives that helped schools monitor students' dropout risk and reduce students' course failures. Chicago students' 9th grade "on track" and high school graduation rates improved as schools used the new indicator to identify students in need of extra support and developed ways to help students improve their grades.

3) Identify organizations that stand out as effective and replicate their success.

Linking short- and long-term measures of success across students by school, district, or program allows policy makers to identify successful organizational practices. Agency staff or researchers can then investigate those organizations' practices, share what they learn more broadly, and pilot innovative approaches in other schools or districts to test their effectiveness. California schools are currently assessed primarily based on student test scores. Richer data about students' post-high-school outcomes could paint a fuller picture of which schools and districts are improving their students' lives.

Case study: In Tennessee, the National Center on Scaling Up Effective Schools at Vanderbilt University has been working with two urban school districts to use longitudinal data on high school students to identify more and less effective high schools and learn about the practices in those high schools that may be contributing to their effectiveness.

4) Evaluate the effectiveness of policies and programs.

The goals of many education, health, criminal justice, and social services programs reach across sectors--education programs may have positive impacts on criminal justice and health out-comes, while social service programs may improve academic performance. Furthermore, programs can have long-term and far-reaching impacts, including on higher education and wage outcomes. Linking data across sectors and over time allows researchers to track the proximate and wide-reaching effects of programs—giving a full picture to policy makers of what works, where to invest taxpayer dollars, and how to target resources to maximize impact on California's students.

Case study: In Florida, researchers have used statewide longitudinal data on teachers and their students to evaluate the effectiveness of state policies designed to reduce teacher shortages in math, science, and special education. Florida researchers also used longitudinal data from a large urban district to show that a universal screening program for identifying gifted students increased the representation of low income and minority students in that district's gifted programs.

California has made significant progress, but still lags behind

California has built and maintained several foundational datasets that facilitate some links between K-12, higher education, workforce, and social services data—including CALPADS and Cal-PASS Plus. However, these efforts are still isolated and not linked to each other. They fall short of providing a complete picture of how students move through California's education systems, and how various state systems interact with each other.

Many educational institutions maintain robust student datasets. Some stakeholders have recognized the importance of linking these across districts and with other institutions. The systems remain limited in value, however, due to less-than-complete scopes. Nevertheless, they establish a baseline infrastructure that can be built upon to support improved future data systems.

Foundational state-wide efforts

Two recent state-wide efforts provide a partial demonstration of what can be accomplished with more comprehensive data systems.

The California Longitudinal Pupil Achievement Data System (CALPADS), maintained by the California Department of Education, brings together some of the student- and staff-level data collected by school districts across California. This allows CDE and collaborating researchers to:

- Generate counts of high-need students used to determine Local Control Funding Formula (LCFF) supplemental and concentration grant amounts
- Match student data to information from the California Department of Social Services (CDSS) to provide information to school districts about whether students are foster youth and/or eligible for free meals
- Measure student progress, even as students move across schools and districts
- Compare outcomes among different schools, districts, and geographies more flexibly than with standardized accountability metrics

None of these analyses would have been possible prior to CALPADS.

Another foundational effort has pioneered linkages between K-12 and postsecondary education data. Cal-PASS Plus, a partnership created through leadership and funding by California Community College Chancellor's Office, contains linked, longitudinal education data provided by participating school districts, community colleges, and universities. It makes it possible for participating agencies, and collaborating researchers, to match students from K-12 education to community college (and some four-year college) records.

Room for improvement

These efforts generate real value, and represent major improvements since the release of the first Getting Down to Facts report in 2007. However, they still fall short of providing a complete picture. Linkages across systems remain limited— although CALPADS is statewide, Cal-PASS Plus is limited to participating school districts and higher education institutions.

Neither provides a complete picture of how students move through California's education systems, and how various state systems interact with each other. Moreover, these highly valuable datasets are underused. Experience in other states has shown that longitudinal data systems generate much more knowledge when governance systems and resources are in place to ensure access for a wide range of worthwhile projects.

In the absence of a strong, statewide coordinated effort, some local and regional agencies have invested significant resources in building local systems. Some of these efforts have led to innovative solutions that provide valuable information for local leaders and educators. However, this vacuum also promulgates duplicative systems— with local agencies linking data for isolated purposes and investing heavily in creating infrastructure locally that could be generated once at the state level and shared. Furthermore, such efforts are far more feasible for large districts; smaller and more rural districts are left out of the benefits they provide.

One major, and important, concern regarding the creation of a statewide, linked data system is that data breaches would threaten student privacy. Ironically, however, the current patchwork collection of educational data systems in California may generate *greater* privacy risks than would a single statewide system, both because wider groups of analysts must have access to sensitive information with multiple overlapping systems and because some of these systems may not have sufficient scale or expertise to support state-of-the-art measures for protecting data security.

Regional and District Efforts

The two statewide systems are not the only source of recent innovation. There have also been important efforts made at the district and regional levels, though by their nature they are unable to provide value throughout the state.

Regional efforts:

- *The Silicon Valley Regional Data Trust (SVRDT)* is working with three counties in Silicon Valley to build data systems that integrate data from the 27 school districts with data from child and family services, juvenile justice, mental health, and other county social service programs.
- *The Education Equals Partnership*, an effort to improve educational outcomes for students from foster care, has worked with four California counties to link individual-level child welfare and education data and make these data available in real time to support youth.
- *CORE*, a consortium of school districts led by eight large, primarily urban districts, have developed a data collaborative that combines student-level data, including measures (e.g., of socio-emotional development) not available in CALPADS.

District efforts:

- Several large districts have created their own data systems, partnering with outside researchers to support their efforts. These partnerships include the *Los Angeles Education Research Institute (LAERI) in Los Angeles, San Diego Education Research Alliance (SanDERA) in San Diego, and the Stanford/San Francisco Unified School District Partnership.* Some districts have linked their students with college enrollment and completion data purchased from the National Student Clearinghouse.

The path to improved data systems in California

While real effort will be required, building an improved data system in California would generate both short- and long-term payoffs. We outline a four-part strategy for moving the state forward, and provide further detail on each step in the full report.

1) Expand data linkages by building on existing infrastructure

Rather than trying to build a single, fully integrated, PK-20 data system that serves all potential uses, California should proceed incrementally, steadily improving statewide data systems and working gradually toward a more comprehensive system – or set of systems – that would meet many of the state's needs. In the short- and medium-term, California should incrementally build on existing foundations to integrate its data systems by:

- Coordinating efforts to link K-12 to higher education data
- Linking workforce data systematically to higher education data, and eventually to K-12 data
- Expanding linkages to state social services data
- Facilitating linkages between county, district, and statewide datasets

2) Champion integration through strong political and agency leadership.

An integrated data system that draws on information contained in many agencies' records will require close coordination among those agencies. Political leaders need to encourage this coordination: formally through the allocation of resources and informally by clearly articulating the importance of improved data systems. Agency leadership, in turn, needs to ensure that a range of voices are included in the design and governance of the data system. Because the California Department of Education (CDE) has already set up successful models that can inform future efforts, we recommend that CDE lead the conversation among state agencies and other stakeholders. CDE could both champion and drive discussions around privacy, access, and data governance, especially among other participating state agencies.

Beyond Accountability: Previous pushes for data integration focused heavily on using data for accountability purposes; many of these efforts were perceived as punitive by educators. Today's policy makers should emphasize that education data systems can be used for a wide range of non-accountability purposes as described throughout this report; for example, better data systems make it easier for teachers and administrators to tailor their approaches to students, to better coordinate service delivery, and to assess and improve program effectiveness.

3) Establish a governance structure that involves leaders from appropriate agencies to develop rules on access, use, privacy, security, and IT infrastructure

CDE, or any other agency that leads the process of developing a data system, will have a crucial role to play in designing the ground rules and infrastructure for a more integrated education data system in California. It will need to drive conversations between collaborators to:

- Decide where to centralize data storage
- Centralize data linking
- Build a data security plan
- Establish rules about data access and use
- Create a tiered access system with guidelines around use
- Establish and streamline use agreements

Each of these decisions requires managing competing considerations: for example, between the absolute importance of protecting student privacy and the goal of ensuring that the data can be used for a wide range of worthwhile purposes. Other states, such as Washington, Texas, Florida, Kentucky, Maryland, and North Carolina, have

established successful models of linked educational data systems with guidelines for access, use, privacy, and infrastructure. California can draw on their lessons learned to establish a governance model that serves California's needs.

Data on Immigration Status: In the current environment, the risk that data systems can be used for inappropriate purposes must be taken extremely seriously. Care should be taken to make sure that any new data systems do not increase risk to immigrant children and their families. The state might consider storing any information about immigration status separately from other data, without direct links to personally identifiable information, and restricting the set of uses for which such data can be made available.

4) Build infrastructure to support linked data systems, and the capacity to use them.

A linked data system in California will require resources and expertise. Policy makers should:

- Invest in secure IT infrastructure that adequately safeguards data and prevents private information from being inadvertently released.
- Establish partnerships that provide capacity to conduct the analyses that end users need to inform decisions. These could be with universities or external research groups, local districts and agencies, or regional data partnerships.
- Build staff resources in local and state agencies and state universities to help process, analyze, and ultimately use the data systems that will be developed.