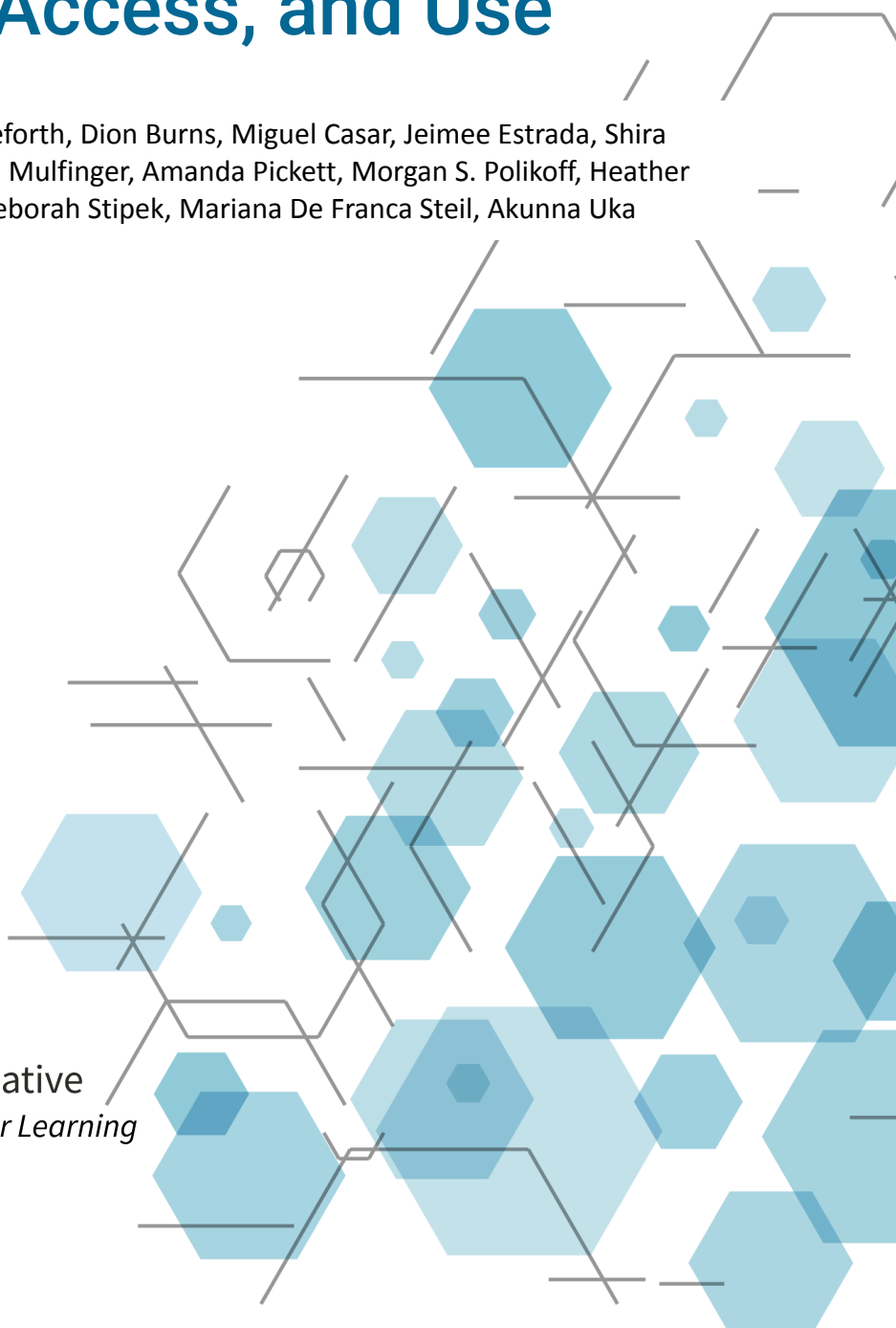
Getting Down to FACTS

Research Brief | May 2026

Education Data in California: Infrastructure, Access, and Use

Jon Fullerton, Jacob Alonso, James Bridgeforth, Dion Burns, Miguel Casar, Jeimee Estrada, Shira Haderlein, Julie Marsh, Beth Meloy, Laura Mulfinger, Amanda Pickett, Morgan S. Polikoff, Heather Price, Vandeka Rodgers, Beth Schueler, Deborah Stipek, Mariana De Franca Steil, Akunna Uka



Stanford | SCALE Initiative
Accelerator for Learning

Introduction

A strong data infrastructure can help district and school leaders understand student progress in school and after graduation, enable researchers to identify equity gaps and evaluate which programs and strategies help close them, provide policymakers with evidence about implementation and who benefits from policy, and support schools, students, and families in navigating complex systems and making informed decisions.

California has built a substantial education data infrastructure over the past two decades. The California Longitudinal Pupil Achievement Data System (CALPADS) provides a standardized statewide source of K–12 student, course, teacher, and program data. The California College Guidance Initiative (CCGI) provides tools that help students, families, schools, and counselors track college readiness and complete key college-going steps. The California Cradle-to-Career Data System links information across K–12, higher education, workforce, and other public systems in ways that were not previously possible. These systems are beginning to make data more actionable, helping students track college readiness, allowing districts to examine postsecondary outcomes, and giving the state new ways to study policy implementation and equity.

At the same time, California's education data ecosystem remains complex, uneven, and incomplete. Local education agencies (LEAs) rely on many different operational systems, with substantial variation in local data capacity, staffing, and access to analytical support. County Offices of Education, research partnerships, and third-party providers help fill some of these gaps, but access to those supports varies widely by geography and local resources. As a result, some districts can integrate and use data to guide decisions, while others only have capacity to focus primarily on compliance and reporting.

Important gaps also remain in the state's own data. Some data are not collected at the right levels or at all, including early education and program participation data for major initiatives. Other data are defined inconsistently across LEAs, including criteria for reclassification in multilingual programs, limiting comparability across districts. Some potentially important data, including information on immigration status and LGBTQ students, are not collected or linked at the state level, in part reflecting legitimate privacy and policy considerations. Data about multilingual programming, such as curriculum or multilingual program type, and associated student enrollment or teacher assignment, are missing and thus cannot inform analysis of what opportunities may work for various groups of students.

Still other data are collected and reported in ways that inhibit usability and learning, including key planning and accountability documents such as LCAPs and SARCs, which are only available in static formats that make comparison difficult. Public accountability tools that use data, such as the California School Dashboard and School Accountability Report Cards (SARCs), are difficult to use, though the state has recently made some efforts to provide stand-alone data downloads from the various SARC tables

("Downloadable SARC Data Files"). Even when data exist, they are not always shared in forms that are easy for educators, families, researchers, and policymakers to use (Stipek and Meloy; Haderlein and Polikoff; Fullerton).

This brief draws on the Getting Down to Facts III technical reports to describe California's current education data infrastructure, identify major strengths and persistent weaknesses, and outline policy choices that could make the state's data systems more useful. The core question is whether California's growing data infrastructure will reach educators, families, and policymakers in forms they can use to improve student outcomes.

Definitions

CALPADS
California Longitudinal Pupil Achievement Data System. California's statewide K–12 longitudinal data system. It standardizes student, course, teacher, and program data reported by districts and is the state's primary K–12 data source.

CCGI
The California College Guidance Initiative. An operational data system that helps students track A to G completion, apply to public colleges, send transcript data, and complete financial aid steps.

C2C
The California Cradle-to-Career Data System. It is the state's new cross-sector longitudinal system linking education, workforce, and other public data to support public dashboards, query tools, student guidance, and research.

CTC
The California Commission on Teacher Credentialing. It maintains data on teacher credentials, permits, preparation pathways, and related educator workforce information.

DataQuest
A CDE public data site providing aggregate historical information on enrollment, assessments, graduation, college attendance, staffing, and other topics.

Ed-Data
A public site that includes much of the same underlying data as DataQuest, along with finance data and a more graphical way of presenting the data.

California School Dashboard
The California School Dashboard, a public accountability tool that reports school and district performance on multiple indicators such as achievement, graduation, suspension, and college and career readiness.

SARC
The School Accountability Report Card. SARCs provide additional public information on school demographics, resources, safety, staff, and performance.

Key Findings

1 California's data landscape has improved substantially, with more timely information available to students and families and new cross-system linkages that expand what the state, researchers, and the public can learn from education data.

Systems such as CCGI and the California Cradle-to-Career Data System have strengthened the state's ability to connect information across education and postsecondary pathways in ways that were not previously possible.

2 Access to data, analytic support, and integrated systems varies substantially by LEA size and geographic region.

Differences in local capacity, data systems, and reporting practices mean that some LEAs can integrate and use data for improvement, while others have far less access to usable data and analytical support. Variations in data definitions and data entry also limit comparability across LEAs.

3 Significant data gaps remain, especially in early childhood and in the state collection of information needed to understand program implementation and student experience.

Data on early childhood are often fragmented and not linked across systems, and in other areas such as expanded learning the state collects too little information to support basic implementation monitoring or evaluation.

4	**California's public data and accountability tools provide substantial information, but they are often difficult to interpret, compare, and use.**

Multiple public data sources report similar information in different ways, and the lack of clear alignment across tools makes it difficult to understand trends or compare results across agencies.

The Evidence Behind These Findings

California's data landscape has improved substantially, with more timely information available to students and families and new cross-system linkages that expand what the state, researchers, and the public can learn from education data

Fullerton describes how California's education data landscape is much stronger than it was a decade ago. CALPADS continues to provide a solid statewide standardized K–12 data foundation. CCGI tools connect local LEA course-taking data to college and financial aid processes and give students and schools more actionable information about college readiness and application steps.

The new California Cradle-to-Career Data System (C2C) is an especially important step forward because it addresses a long-standing problem in California: the absence of a comprehensive cross-agency data infrastructure that can track students across systems. It is intended to provide linked data that will better allow agencies to serve their clients, public dashboards enabling a broad base of users to understand the successes and challenges of education in California, and a protected research environment for approved projects.

This linked data can serve different users in different ways. Students and families can use the data to navigate school and postsecondary pathways. Districts can use the data to evaluate the success of their students after graduation. The state can evaluate whether policies and programs are being implemented and whether they are producing the intended results. Researchers can study outcomes over time and across systems. California now has more of the infrastructure needed to support these uses.

Research practice partnerships such as the Los Angeles Education Research Institute and district improvement networks such as CORE have used linked administrative data to study issues including college access and middle school math policy. These examples show how better integrated data can

support policy evaluation and improvement when districts have the infrastructure and partnerships to use it. Indeed, many of the analyses in the GDTF III series of technical reports leverage this data infrastructure extensively. In this sense, the new data infrastructure is already changing what researchers and policymakers can see about California's education system.

Access to data, analytic support, and integrated systems varies substantially by LEA size and geographic region

Fullerton also notes that many of the data systems that matter most for local decision making remain local and fragmented. LEAs use different student information systems and many separate platforms for assessment, special education, finance, human resources, learning management, and school climate. Larger districts may build their own data warehouses to link data across systems or partner with external providers to connect data across systems. County Offices of Education also can provide strong data services, including dashboards, data warehouses, and access to postsecondary outcome data; but what services are provided differs across counties. In addition, key data systems such as those maintained by the California Commission on Teacher Credentialing (CTC), which track teacher credentials and preparation pathways, remain separate from K–12 data systems, limiting the ability to link information about teacher preparation, assignment, and outcomes.

This means that access to quality data and analytical support is uneven. Large urban districts and districts with strong research partnerships often have much greater capacity to link and interpret data while smaller districts may have only one staff member handling state reporting, with little time left for analysis or improvement work. Due to data suppression rules, smaller districts may also have less usable data in public reports, such as C2C's Pathways to College in California dashboard. This uneven capacity and ability to access data is a central feature of the state's data landscape.

Differences in data quality and standardization also contribute to these disparities, as LEAs vary in their data entry practices, validation processes, and ability to ensure consistent and accurate reporting. These disparities reflect variation in LEA and county office capacity, as well as uneven access to external partners such as research organizations and technical assistance providers.

Finally, Marsh et al. note that a large majority of school board members would like more training on evaluation and using data to improve decision making, highlighting the importance of developing user capacity to use the data that is available effectively.

Significant data gaps remain, especially in early childhood and in the state collection of information needed to understand program implementation and student experience

Stipek and Meloy highlight that data on children, services, workforce, and quality improvement are often disconnected at both the state and local levels. Because there is no shared unique identifier across many ECE systems, the state often cannot tell how many children are being served across programs, what combinations of services they receive, or how those experiences connect to later outcomes.

Important workforce and child outcome data are also limited. California lacks systematic statewide information on the ECE workforce, including compensation, qualifications, and movement across programs. Child level outcome data are inconsistent across settings, and the state still does not collect systematic achievement data before third grade in ways that allow strong statewide analysis of early learning trajectories. Program quality data are also incomplete.

Multilingual programming data are also missing. As Burns and Price point out, the type of program at a school is voluntarily divulged to CDE for a directory. It is incomplete at best and provides no information on the enrollments in these programs nor information about the curriculum. Basic information on the language of instruction also continues to be an unfulfilled data element.

Fullerton highlights that the state also does not collect some important data in other areas, and some data are not linked in useful ways. For instance, after spending billions of dollars on the Expanded Learning Opportunities Program, the state is only beginning this year to collect the most basic attendance data on who participates. In general, the state often neglects to capture sufficient data to allow for the evaluation of major programs it funds.

At the same time, some areas of reporting, such as special education, illustrate what more developed data systems can look like, with more standardized definitions, required reporting, and the ability to track services and outcomes across students and programs. These examples highlight what is possible when data collection is designed and implemented more consistently.

California’s public data and accountability tools provide substantial information, but they are often difficult to interpret, compare, and use

Haderlein and Polikoff examine California's public accountability data infrastructure, especially the California School Dashboard and the SARCs. These tools reflect a broader shift toward multiple measures, transparency, and continuous improvement and are intended to make school performance visible and usable for families, educators, and policymakers.

In practice, however, there are major usability and coherence problems. The relationship between the Dashboard and SARCs is unclear. The tools present overlapping information without a clear explanation of how they fit together. Users must navigate multiple portals, definitions are not always aligned, and

important historical trend and subgroup information can be hard to find. Haderlein and Polikoff also point out that California has not clearly defined primary audiences or theories of use for these tools, which makes it harder to use them to support action.

Fullerton illustrates that this is true of broader public reporting tools as well. DataQuest, Ed-Data, the Dashboard, and SARCs often rely on much of the same underlying data but present them in different ways. This can make the system confusing for users trying to understand school performance. Better consolidation and clearer crosswalks across measures would improve public understanding.

Implications for California

Better collection of high-value data

The evidence suggests the importance of building implementation and outcome tracking into major programs from the outset. California will remain limited in what it can evaluate if important data are missing on early childhood participation and program quality, multilingual program experiences, intervention dosage, program implementation, and other key services.

Public tools designed with users and use cases in mind

The evidence points to the importance of public-facing tools with clearer audiences, clearer explanations, better cross-linkages, and easier access to historical trends and comparisons across LEAs. Families, educators, and policymakers do not all need the same interface or the same level of detail. These findings suggest the value of more user-centered design focused on clearly defined users and use cases.

More support for local analytical capacity

The findings suggest that a strategy to maximize effective data use cannot rely only on large districts and a patchwork of partnerships to do analysis. Smaller districts need more stable access to data infrastructure, technical assistance, and analytic support. County offices can play an important role in this, but their staffing and analytic capacity would need to reflect the scale of local need. Marsh et al. show that school board members also want more support interpreting and using data, and that a nontrivial share overestimate student performance relative to actual performance data. The broader implication is that statewide policy needs to address unevenness not only in access to data, but also in the capacity to interpret and act on it.

Conclusion

California has made substantial progress in building a robust education data infrastructure. CALPADS, CCGI, the California Cradle-to-Career Data System, and related systems give the state a much stronger foundation than it had even a few years ago. These systems are already creating real opportunities to identify equity gaps, track student progress across levels of the system, understand postsecondary and workforce outcomes, and support more informed decisions by schools, districts, families, and policymakers.

However, important gaps remain in the data collected around early childhood and other high-value areas. Some of the public-facing data and accountability tools still make it too hard for users to understand and act on the information they provide. Local access to integrated data and analytical support remains uneven across districts.

As California has begun moving from building individual pieces of data infrastructure toward connecting them, the harder work now is ensuring that what has been built is highly usable by the students, families, educators, and policymakers it is meant to serve.

GDTFIII Technical Reports Referenced

Burns, D., & Price, H. (2026). *Multilingual learners of English: Progress of California's English learners and the resources that support their educational achievements*. *Getting Down to Facts III*, SCALE Initiative, Stanford University.

Fullerton, J. (2026). *Education data needs, availability, and access in California*. *Getting Down to Facts III*, SCALE Initiative, Stanford University.

Haderlein, S., & Polikoff, M. S. (2026). *Public accountability in California: Evaluating the SARCs and the California School Dashboard*. *Getting Down to Facts III*, SCALE Initiative, Stanford University.

Marsh, J., Schueler, B., Bridgeforth, J., Alonso, J., Estrada, J., Pickett, A., Uka, A., Steil, M., Rodgers, V., Mulfinger, L., Casar, M. (2026). *Local Control in a Time of Change: The Work of California School Board Members. Getting Down to Facts III*, SCALE Initiative, Stanford University.

Stipek, D., & Meloy, B. (2026). *Early childhood education – Section 6: Data & data systems*. *Getting Down to Facts III*, SCALE Initiative, Stanford University.