Education Insights Center
Advancing Research and Policy for K-12 and Postsecondary Education

POLICY BRIEF
AUGUST 2017

California's Maze of Student Information: Education Data Systems Leave Critical Questions Unanswered

Part two in the series:
California Education Policy, Student Data, and the Quest to Improve Student Progress

By Colleen Moore, Kathy Reeves Bracco, and Thad Nodine

California collects expansive sets of data about students in its public K-12 and higher education systems—data that, collectively, have great potential to meet the information needs of state policymakers, local educators, and other stakeholders. But the data are collected and maintained in systems that are not connected, were designed for different purposes, are subject to different regulations, and often use different data definitions. As a result of these disconnects, important information about student progress is often impossible to access, share, and use—whether at the state, regional, or local level. While there may be a few benefits to the current structure, they are outweighed by major disadvantages, including inefficient use of taxpayer dollars. A significant weakness is that California's current approach leaves the state and its institutions unable to answer important questions about student progress and outcomes, such as:

- With California's public universities straining to meet student demand for enrollment, how many students meet current eligibility requirements, but do not apply? What are their educational outcomes?
- How many students attend multiple colleges or universities in various regions of the state during their college careers? How does this affect their educational trajectories and chances of college completion?
- California is investing millions of dollars in strengthening educational pathways from high school into college. Which programs are most effective, and for which students?

About This Series

California Education Policy, Student Data, and the Quest to Improve Student Progress

This brief is the second in a series that is examining California's approach to gathering and sharing longitudinal data about students' progress through the state's education systems. The series includes four briefs:

- An analysis of the perspectives of state and local leaders on who should be responsible for gathering and sharing data about students' progress.[1]
- An overview of student-level data collected and maintained in California, a summary of past efforts to develop a more comprehensive system, and an exploration of the strengths and weaknesses of the state's approach to education data.
- An analysis of local and regional efforts to share data across institutions and systems that includes the benefits and challenges of participating in these efforts.
- An exploration of lessons for California from some other states' efforts to improve their education data systems, a summary of our findings across the series, and some conclusions concerning a path forward to improve California's data systems for use in understanding and improving education policy and practice.

One effect of having disjointed data systems is that they contribute to fragmented education programming and policy making. As one person interviewed for this report said, "We don't think through the movement of students through the various sectors. We just start with, 'They've arrived, now we can start tracking. Oh, they left. Oh, darn it, that's too bad.'" Another interviewee said, "We can't really construct good state policy when everything is so localized that you can't understand what's actually happening to large swaths of students."

These observations are based on a review of research reports, websites and other documents,[2] as well as in-depth interviews with 28 people familiar with education data issues in California, undertaken by the Education Insights Center as part of a four-part series of policy briefs examining the state's student data systems (see About This Series). The interviewees included state policy staff, state agency staff, officials in the central offices of each public education system, staff in organizations involved in cross-sector data sharing efforts, education researchers, and policy advocates.[3] In our first brief, *Gaps in Perspective: Who Should Be Responsible for Tracking Student Progress Across Education Institutions?*, we found a disconnect in the views of state policy staff and local education leaders on the need for a statewide data system. Local educators expressed a need for—and a readiness to participate in—a state-level system to gather and track student progress across schools and colleges, but state policy officials were not convinced the state should take a leading role in gathering and sharing such information. At least six efforts in recent years have failed to create a new entity that can oversee data collection and use in California. This brief seeks to make clear the missed opportunities the state has had with regard to understanding its investments in public education.

California Collects Robust Sets of Data About Its Students

California's four systems of public education collect fairly comprehensive data about students' characteristics, their attendance and enrollment patterns, and their progress and outcomes (see Table 1). Collectively, these data systems contain many of the elements considered by national data-focused initiatives, such as the Common Education Data Standards (CEDS)[4] and the Data Quality Campaign (DQC),[5] to be essential when measuring students' educational progress and outcomes.

- **The California Department of Education** (CDE) began developing the California Longitudinal Pupil Achievement Data System (CALPADS) in 2002 in response to federal requirements under No Child Left Behind.[6] The system became fully operational in 2009. Public school districts report information about students from kindergarten through 12th grade. The primary purpose of CALPADS is to facilitate compliance with federal and state reporting requirements, but the CDE is beginning to explore more strategic use of the data to understand student progress and to support improvement of school processes.

> "I think the biggest plus is the potential. I mean, look at the size of it, how much information is there…Data systems in California contain so much information. I think figuring out ways to open that up for research could move things in a big way."
>
> *– Education data organization staff member*

- **The California Community Colleges** (CCC) developed the Chancellor's Office Management Information System (COMIS) in the early 1990s; it houses student-level data from all 72 districts (114 colleges). The data were originally used primarily for the allocation of funding to the colleges and for federal reporting on their behalf (e.g., reporting required by the U.S. Department of Education for the Integrated Postsecondary Education Data System, or IPEDS).[7] The use of the COMIS data has evolved to focus more on measuring student performance and outcomes,

driven largely by state accountability and reporting requirements. Due, at least in part, to limited staff capacity, the Chancellor's Office contracts out some data functions to the CCC Technology Center, which is operated by the Butte Community College District. That center runs the admissions application process for the colleges, an online course exchange, an electronic transcript exchange system, and new initiatives to develop a common assessment process and tools to facilitate education planning.

- **The California State University** (CSU) Chancellor's Office has an Enrollment Reporting System (ERS) that collects student information from the 23 campuses—data that are primarily used to meet state and federal reporting requirements. The Chancellor's Office is currently working to develop a central repository for data, combining the ERS data with other data historically held in separate departments (e.g., financial data and staffing information) in order to make better use of the information for program planning and improvement.

- **The University of California** (UC) Office of the President collects student data from its 10 campuses and uses the data for state and federal reporting. The Office of the President is currently working to develop the UC Data Warehouse to better integrate student, faculty, staff, and budget data for business- and academic-related analyses.

Table 1
California's education systems collect fairly comprehensive student-level data.

System	Student Identifier	Data Elements	Source of Data	Data Tools (publicly accessible)
California Department of Education	Statewide Student Identifier (SSID)	• Demographics • Course enrollments • Attendance • Special program eligibility • Discipline incidents • Completion (diploma, GED)	K-12 school districts report data twice in the fall, and at the end of the school year	• DataQuest • California School Dashboard • Other summary reports on CDE website • Ed Data[8]
California Community Colleges	Social Security Number (SSN)	• Demographics • Course enrollments/grades • Financial aid • Special populations/ programs (e.g. disabled) • Educational goal • Participation in orientation, other matriculation services • Degrees/certificates	Colleges/districts submit some data each term and other data annually	• Datamart • Student Success Scorecard • Salary Surfer
California State University	SSN	• Application information • Demographics • Course enrollments/grades • Financial aid • Degrees	Campuses submit some data each term and other data annually	• Student Information Dashboard • Other summary reports on Analytic Studies Division website
University of California	SSN	• Application information • Demographics • Course enrollments/grades • Financial aid • Degrees	Campuses submit some data each term and other data annually	• UC Information Center • UC Accountability Report

Each of these four data systems is fairly robust, and together they have great potential to provide information relevant to meeting the needs of policymakers, educators, and other stakeholders. But there are some key limitations, including:

- The use of different student identifiers in the K-12 and higher education systems, which complicates efforts to share data across the systems;
- Variable data definitions across institutions and systems, which makes it difficult to analyze and interpret information consistently;
- Inconsistent use of data quality control processes across institutions and systems, leaving questions about the reliability of some information;
- Limited information about the costs of higher education and the corresponding returns on investment for students and taxpayers; and
- Missing information on the particular certificate or degree programs CCC students are pursuing, and on student participation in career pathway programs in high schools. This is a significant issue given the state's investment in such programs (e.g., the Career Pathways Trust and Strong Workforce programs).[9]

Each of the existing four data systems is used primarily for compliance-related purposes (such as administering programs and meeting federal and state reporting requirements), but each responsible entity is moving toward making its data more accessible to key stakeholders through data dashboards, scorecards, and the like, aiming to increase transparency and improve program planning. For example, the CSU Chancellor's Office has developed data dashboards for internal use by its campuses. One use of the dashboards has been to identify courses with high rates of failure, with the idea that these courses could be redesigned to increase student learning and success. As another example, the CCC Chancellor's Office has developed Student Success Scorecards that are available to the public and that provide information on key metrics of student progress. In light of data analyses by institution- and system-level researchers that demonstrate limited student progress through developmental English and math, many community college campuses have begun working to improve those programs.

> "With the evolution of the various tools that we have, we're able to mine the data a lot more creatively than we have in the past, address more sophisticated research questions, and combine [pieces of information] in a way that we might not have thought of before."
>
> *– Educational system representative*

Somewhat ironically, it is this movement toward more strategic uses of data that helps to highlight the limitations of California's approach to collecting and maintaining student data. As long as compliance with external agencies is the overriding purpose of educational data systems, local entities tend to perceive data collection as more of a burden than an asset, and they tend to overlook the limitations of these reporting systems. As data sharing and use improve within the institutions and their system offices, however, the burdens of not having a statewide data system that integrates student data across education sectors become increasingly clear. As we explore in the next sections, even though each education system has fairly rich information about the students it serves, the information covers only a piece of the student journey; understanding the entire journey requires connecting data across systems.

Data Systems are Disconnected, With Limited Information Across Institutions and Sectors

Each of California's public education data systems is extensive, but, taken together, they can best be described as a compilation of disparate systems operated by distinct entities, designed and built for different purposes, subject to different statutes and regulations, and using different data elements and definitions. In charting the mandatory and voluntary data sharing and reporting requirements for each of the four public systems (K-12, CCC, CSU, and UC), the full complexity of administrative relationships around data reporting becomes clear (see Figure 1).

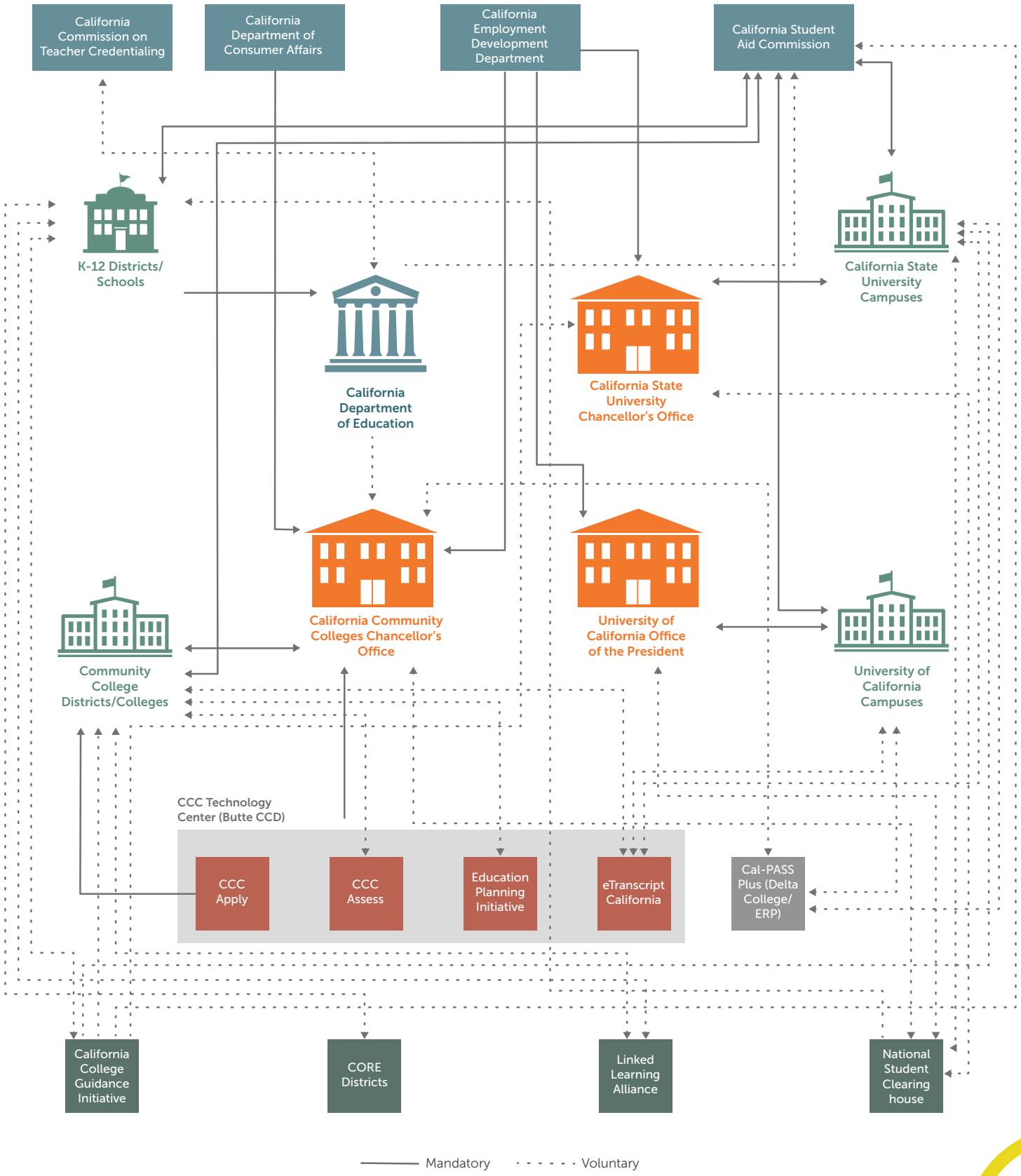
Mandatory Reporting

Educational institutions in the state are required to submit certain student-level information to the relevant systemwide office—the CDE, CCC Chancellor's Office, CSU Chancellor's Office, or UC Office of the President.[10] In addition to this reporting, the institutions or systemwide offices must also report data to or receive data from several state agencies (depicted along the top of Figure 1):

- **The California Student Aid Commission** (CSAC) collects information from students via the federal Free Application for Federal Student Aid (FAFSA) and the California Dream Act Application. Along with students' grade point averages (GPAs), collected from schools and colleges, the information is used to determine student eligibility and to administer state financial aid.
- **The Employment Development Department** (EDD) collects information from employers about their employees' wages on a quarterly basis for the purpose of administering the Unemployment Insurance Program. It supplies data to postsecondary systems about their students' employment and earnings—a task required by the state as part of postsecondary accountability reporting.
- **The California Department of Consumer Affairs** (DCA) administers state licensing for a variety of occupations, including those in cosmetology, private security, and health care. As required by the state, the DCA provides information to the CCC Chancellor's Office about students' success in achieving appropriate licensing following their education programs.
- **The California Commission on Teacher Credentialing** (CTC) collects information from the CDE about which teachers are employed and where, and the CTC provides the CDE with information about the credentials and authorizations for these teachers. In addition to its own Statewide Educator Identifier (SEID), the CTC has the Social Security number, which would allow for matches to CSU and UC student records to understand outcomes of students in teacher credential programs.

> "[The state's] view is local control, local control, local control. But you need to provide some unified infrastructure. You make [schools and colleges] do every single thing and build their own everything. That's not helpful. They want somebody to be responsible for providing some direction on what it takes to actually [use cross-sector data] effectively, because everybody is just scrambling."
>
> *– Education data organization staff member*

Figure 1
California's Maze of Student Data



Voluntary Data Sharing

The lack of a comprehensive education data system has led to a variety of efforts by partner organizations to collect student-level data from individual schools, colleges, and universities, and to match those data across institutions and sectors. Participation in these efforts is voluntary, and participating institutions sign a contract or a memorandum of understanding (MOU) governing the use of the data, they invest resources in putting together data files to submit to the organization(s), and they often pay a fee.

- **Cal-PASS Plus**, created and operated with state funding through the CCC Chancellor's Office, is managed currently through a partnership between San Joaquin Delta College and the nonprofit Educational Results Partnership (ERP).[11] The data are used to support several Regional Learning Collaboratives, and initiatives by the community colleges that include the Multiple Measures Assessment Project and the CTE LaunchBoard.[12]
- **The California College Guidance Initiative** (CCGI), an effort that receives both state and philanthropic funding, facilitates the electronic sharing of students' academic transcripts and their college and career plans between high schools and postsecondary institutions.[13] The data are used to support K-12 counselors in helping students prepare for college, to allow students to create profiles of their college and career plans, to inform community college placement processes and CSU admissions processes, and to facilitate CSAC's ability to match FAFSA data (which uses SSNs) with Cal Grant GPA verifications (which use SSIDs) for K-12 districts that partner with CCGI.
- **The CORE Data Collaborative** evolved from the nonprofit CORE Districts partnership among eight large school districts in the state.[14] The collaborative is open to all K-12 districts in California for a fee, and gives them access to additional measures of school performance and improvement not available through the state accountability system. These can include both academic and social-emotional measures as well as indicators of school culture and climate, depending on the instruments districts choose to administer.
- **Linked Learning Analytics**, currently under development by the Linked Learning Alliance (LLA), focuses on measures of college and career readiness and outcomes for students in career pathways programs.[15] The effort is open to K-12 districts for a fee, with a focus on students' experiences in grades 9-12 and into postsecondary education.
- **The National Student Clearinghouse** (NSC), based in Virginia, collects student-level enrollment and degree information from nearly all postsecondary institutions in the country.[16] By participating in the NSC, K-12 schools and postsecondary institutions can track whether and where their students enroll in college and if they complete a degree. The CCC, CSU, and UC system offices contract with the NSC. Currently, school districts wanting access to information about the college enrollment and outcomes of their students have to pay fees to obtain NSC data, either individually or through a regional collaborative.[17]

> "You end up with a situation where you've spent hundreds of millions of dollars on state data systems [at the various segments], and yet every college and K-12 partnership has to buy data from the National Student Clearinghouse, which is ridiculous, and then it only gives them a small picture of what is happening."
>
> *– Education data organization staff member*

Because the efforts identified above are voluntary, these data systems are not comprehensive (that is, they cover only a subset of K-12 and postsecondary institutions), and participation and actual submission of data can change over time. For example, while all community colleges have MOUs in place with Cal-PASS Plus and the CCC Chancellor's Office does the work to submit the data on their behalf, only one UC campus and nine CSU campuses have a current MOU. In addition, about half of those campuses have not actually submitted data for the past several years.

Additional Reporting

The data collection and sharing relationships in California are even more complex than depicted in Figure 1. A number of the entities shown also report data to the federal Department of Education for financial aid applications, IPEDS reporting, and other purposes. The graphic does not include the new longitudinal data system for adult education and Career and Technical Education (CTE) programs under development by the California Workforce Investment Board that will involve the CDE and the CCC Chancellor's Office, as well as workforce training programs offered by a variety of organizations through local workforce boards (see A Longitudinal Data System for Workforce Education). Private educational institutions at both the K-12 and postsecondary levels share student information with CSAC, and some participate in Cal-PASS Plus and other voluntary data sharing efforts. In addition, the education systems share information with the California Department of Social Services to ensure eligible students have access to appropriate social services.

A Longitudinal Data System for Workforce Education

As with student-level data for K-12 and postsecondary education generally, information on the individuals who participate in any of the state's workforce education and training programs is similarly disaggregated across multiple programs and sectors. In response to federal mandates, however, the California Workforce Development Board (CWDB) is currently working to develop a comprehensive longitudinal data system for workforce education and training programs.[18] The intent is to integrate workforce and education performance reporting across the different funding streams, providing a status report on credential attainment, training completion, degree attainment, and participant earnings related to the workforce education and training programs of all providers. The data system will include students enrolled in adult education (through K-12 districts or community colleges), CTE programs at community colleges, and workforce training programs offered by various organizations through local workforce boards. A pilot project is underway, and a warehouse of performance data is being created to meet reporting requirements and to allow evaluation and assessment of workforce training programs.

Efforts to Coordinate Education Data Statewide

Since the California Postsecondary Education Commission (CPEC) was discontinued in 2011,[19] several state legislators have attempted to either develop an alternative entity to serve as a coordinating and planning body for postsecondary education or to establish some other means of coordinating data across the education sectors (see Table 2). Most efforts failed to make it through the legislative process, and the two bills that did pass were vetoed by the governor. The failure of these efforts was largely due to concerns about the make-up, roles, and responsibilities of the board for the new state entity; how data collected by the entity would be used and by whom; and the costs needed to establish and maintain a new agency.

In addition, institutional research and planning officials from the CDE and the CCC, CSU, and UC system offices sought on their own to establish a "federated" model of sharing student data. In this arrangement, each system would maintain control of its own data, but agree to a set of processes and procedures for matching data for use by one or more of the systems as needed. The systems signed an interagency agreement that allows the exchange of student records for specific purposes, with approval of all parties required on the use of the data and how analyses are reported for each exchange. The effort did not involve creating the kind of central repository for student records that would streamline each system's use of cross-

sector data for assessment and planning purposes. According to a participant in this years-long effort, the attempt was stymied by staffing turnover, a lack of funding to support the work, and a lack of political will to provide the appropriate policy and governance support. As this education system official said, "We had very, very high hopes, and everyone was willing and able to participate, and we had all the right people in the room to do all the stuff about standardizing definitions and all the technical stuff. We actually had a technical plan ready to go. But what we lacked was the governance and the legal authority for any one segment to manage all of the segments' data. And we couldn't get whoever it was in Sacramento to help us move that along."

While various efforts to develop a more comprehensive student data system or other means of facilitating routine matches of data across California's education institutions and systems have failed, there have been some functional improvements to data sharing around specific processes. For example, among institutions that participate, the eTranscript California initiative run by the CCC Technology Center facilitates the electronic transmission of community college transcripts to UC and CSU for students who wish to transfer, easing the burdensome process of delivering hard copies of transcripts during the admissions process.[20] A recent MOU between the CDE and the CCC Chancellor's Office will allow for an automatic look-up in the CALPADS system to pull CDE's student identifier (SSID) into the CCC application. The CSU Chancellor's Office is pursuing a similar MOU. These agreements could facilitate the matching of K-12 and postsecondary student records in the future—if the political will is there to do so.

Table 2
Recent legislative efforts to improve education data have failed.

Purpose of Legislation	Bill Number	Status
Establish the Office of Higher Education Performance and Accountability • Office would serve as the statewide postsecondary education coordination and planning entity • Governing boards and institutions of postsecondary education would be required to submit data needed to support this entity's planning role	AB 217 (2017, Low) AB 1837 (2016, Low) SB 42 (2014, Liu)	Held in committee Held in committee Vetoed by governor
Establish the California Higher Education Authority • Agency would act as a clearinghouse for postsecondary education information and as a primary source of information for the legislature, the governor, and state agencies • Would be tasked to develop and maintain a comprehensive database	AB 1348 (2013, John A. Pérez) AB 2190 (2012, John A. Pérez)	Held in committee Held in committee
Establish goals for postsecondary attainment • Would require the CCC, CSU, and UC to set specific goals to meet statewide goals set by SB 195 (Liu, Chapter 367, Statutes of 2013) • Would require the systems to collaborate to define metrics and coordinate data	SB 1196 (2014, Liu)	Held in committee
Require the CDE to succeed data management responsibilities previously held by the CPEC	SB 1138 (2012, Liu)	Held in committee
Authorize the CDE, CCC, CSU, UC, and EDD to enter into a joint powers agreement to implement a comprehensive P-20 data system	SB 885 (2011, Simitian)	Vetoed by governor

Stakeholders Identify a Few Benefits, Many Weaknesses, in the State's Current Approach

Given the efforts to develop a more coordinated data system, we asked interviewees to identify both the strengths and weaknesses of the state's current decentralized approach to collecting, maintaining, and sharing student data.

Strengths

Most interviewees had difficulty coming up with any strengths with the current data structures. "There aren't any" was a common initial response. But several people, after some thought, were able to identify a few advantages, saying the current system:

- **Creates space for innovation.** Some noted that efforts by local and regional entities to facilitate voluntary data sharing, like Cal-PASS Plus and the CCGI, have resulted in more innovation in the development of data tools and processes than would occur within a state agency or other centralized data repository. They argued that such efforts have been more responsive to the needs of educators and students, using data in ways that provide valuable tools, such as the CTE LaunchBoard and CCGI's student profiles. Some suggested that voluntary participation in data sharing might lead to greater desire to use data for improvement purposes. One noted that, "you think differently about things you choose to be a part of versus what you're mandated to be a part of."
- **Avoids misuse and misinterpretation.** Some interviewees suggested that each education system is in the best position to manage its own data because its staff best know the data (including data limitations) and can provide context and clarity for appropriate use and interpretation. These interviewees primarily worked within the education systems, and they noted that these segments can make data sharing agreements as needs arise. Several interviewees from outside the systems, when asked about this point of view, acknowledged that familiarity with the data is important, but that unwillingness to cede control over information availability and use is a common symptom of systems that do not use and share data for improvement purposes.
- **Protects privacy, possibly.** A few interviewees suggested that the current structure allows for greater privacy and security because the data are dispersed, with no entity having all of a student's information. However, we also found considerable uncertainty and disagreement on this point, as some interviewees said that the disaggregated structure, with its multiple points of potential exposure, is actually less secure.

Weaknesses

While most interviewees had difficulty identifying strengths in California's education data structure, they had no trouble identifying its weaknesses, saying that it:

- **Supports a compliance approach to data.** There has been some evolution in using data to track and improve student outcomes. Interviewees said, however, that having only separate data systems in each of the segments, with no cross-sector data system, tends to reaffirm their use for compliance with state and federal reporting requirements. As institutions and systems focus on reporting what is required of them, cross-system issues tend to fall through the cracks.

> "We reinforce a myopic view of an institution's responsibility for student success by focusing only on the portions of that student success that the institution contributes to directly."
>
> *– State policy staff member*

- **Reinforces a siloed approach to education planning.** The current decentralized structure does not account for student enrollment and shared outcomes across multiple institutions and systems over the course of students' educational journeys. As one educational segment representative noted, "Nobody has the big picture." Interviewees said these divides reduce the capacity for deeper analysis at the institutional and system levels, particularly about issues that cross systems, and result in no one taking responsibility for focusing on the challenges facing California as a whole.

> "[The systems don't] see this as their role, like what to do around supporting successful student transitions. That's not part of what they do. There isn't anybody in the state that does that. There is no intersegmental entity; it's a vacuum."
>
> *– State policy staff member*

- **Reinforces fragmentation in data structure, quality, and definition.** Compliance-related requirements, such as those mandated by IPEDS, result in some commonalities in student data across institutions and segments. Nonetheless, there is considerable variation in the way data are collected and maintained across institutions, and even within a single system. The historical autonomy of California's institutions of higher education—along with shared governance processes and faculty control of the curriculum—has led to very different curricula, institutional policies, data definitions, quality control procedures, and data reporting schedules. Those issues create substantial complexity in analyzing and interpreting data across institutions and systems.
- **Creates confusion among stakeholders.** The variety of data sources and the many different public-use tools summarizing the data create confusion among policy staff and educators about where to find needed information, and about how to interpret or draw comparisons across the sources. The array of voluntary efforts to address stakeholders' needs for cross-sector information can lead to uncertainty among institutions about which data-sharing efforts are worth the investment of their limited time and resources.

> "This segment says this, and this other segment says that, but are they following the same metrics or definitions? It's confusing."
>
> *– State policy staff member*

- **Leaves stakeholders reliant on the education systems for data and information.** Under the current decentralized structure, policymakers, researchers, parents, students, and other stakeholders are reliant on the education systems or institutions for access to data. Policy staffers noted that the systems sometimes do not respond to requests for information or provide data in ways that suit the systems' own interests, not necessarily those of taxpayers or students. Education researchers reported even greater difficulty in gaining access to data, noting significant problems even with gaining access to de-identified student records in one system, much less across four. Stakeholders noted that this further limits capacity for the deeper analyses of data needed to improve education policy and practice.
- **Leads to duplication of effort and inefficiencies.** Institutions report similar data to several different agencies (see Figure 1), particularly institutions that participate in Cal-PASS Plus or other efforts aimed at giving them access to cross-sector information. Multiple public entities sometimes contract individually with the same vendor, which some interviewees noted as a failure to leverage the state's purchasing power. For example, the CSU Chancellor's Office contracts with the NSC, but some individual CSU campuses also pay for

> "The stakeholders who don't want this [kind of data sharing] to happen would be those that own the data and don't see any advantage in giving the data to other people who might criticize them for bad student results."
>
> *– Researcher*

matches to NSC data to track their students. As another example, an interviewee said that one vendor was providing technology platforms for several efforts—a college and career planning portal for students, the eTranscript initiative, and the CSU system's admissions application process—all on separate contracts with different state-funded entities. Other interviewees noted that the various partners trying to build out intersegmental data systems are in some cases funded by state dollars both directly (for example, the state funding provided to support Cal-PASS Plus via the CCC Chancellor's Office) and indirectly (for example, state funding provided to schools that invest staff time in assembling and submitting data files). While there would be state costs for investing more directly in a cross-sector data system or data-sharing effort, there is already considerable state investment in efforts to match and provide access to student records that cross institutions and systems. The state's current disjointed data systems exhibit considerable duplication and inefficiency, interviewees said, both in terms of money and effort.

> "All over the state, people are trying to build out intersegmental data systems because there isn't a state-level system. In most cases, they don't have the know-how, bandwidth, or capacity to actually make good on that and execute. It's wildly expensive to do it that way. It's just a huge lift for any one group, so why not do it well once and make it available to everybody?"
>
> *– Education data organization staff member*

The Biggest Weakness? Inability to Answer Critical Questions

Many interviewees said that the biggest weakness of all in California's approach to student data is that it leaves the state and its policymakers, educators, and taxpayers unable to answer important questions. They also indicated that one effect of having decentralized data systems that reinforce compliance, silos, confusion, fragmentation, and duplication of effort is that many stakeholders do not even think to ask the bigger, more important, questions that are vital to student success across institutional boundaries.

> "I understand the concern about [the] cost [of developing a statewide data system], but at some point, a decision is going to have to be made about what the state does value and if it is willing to put the infrastructure in place. Because, in the absence of leadership from the state, we've got numerous duplicative efforts that are rebuilding the same system."
>
> *–Education data organization staff member*

Interviewees provided numerous examples of important questions that either cannot be answered at all, or would require data matches that are currently difficult and costly to accomplish, including:

- **How many high school graduates are prepared to succeed in college?** The state's new accountability system for K-12 schools requires the CDE to report on students' college and career readiness. But understanding readiness depends on assessment of what happens to high school graduates once they enter higher education or the workforce, and this cannot be addressed using internal metrics available to the K-12 system (e.g., completion of the "a-g" courses required for admission to UC and CSU). Some school districts obtain data from the NSC to observe their students' college enrollment and outcomes, but NSC data do not include information about students' need for remediation.

- **How many high school graduates are eligible for admission to CSU/UC but do not apply?** The CSU and UC systems have noted that funding and capacity constraints are making it difficult for them to serve all eligible applicants, at both the freshman and transfer levels. But there are likely students who meet the requirements for admission who are discouraged from even applying because of their concerns about gaining admission to a campus or program of their choice, the cost to attend, or other issues. Without connecting K-12 data to application data for the universities, we have no idea about the scale of this issue or its varying impact across regions and student populations.

> "I'm sure every day we shut ourselves down from even asking some questions aloud because we know we can't get the data. We just accept that as the world in which we live, and it might suppress our likelihood or ability to raise some important questions in the first place."
>
> *–State policy staff member*

- **How many students attend multiple postsecondary institutions simultaneously or over the course of their college careers?** Understanding patterns of student movement across educational institutions and segments could shed light on the need for better alignment in curriculum, additional capacity in particular regions of the state, improved supports for students, or changes to state or system policies governing program approval, registration priority status, and other issues.
- **How many high school graduates earn a bachelor's degree within a certain time period, regardless of which education system they initially enroll in?** Each higher education system reports the graduation rate for students that enroll in it, but the overall rate of bachelor's degree completion for a particular class of high school graduates is unclear; it is impossible for high schools to learn, from data they provide to the state, about the overall success rates of their students. While this question could be answered if the CDE were to obtain data from the NSC, many related questions would remain about the factors associated with successful degree completion, as the NSC does not collect information about students' course-taking or other patterns. This information is only available through the data systems managed by the state's colleges and universities.
- **What is the cost of educating an undergraduate in the different higher education systems?** Many interviewees noted that we have limited information to understand the comparative costs to students and taxpayers and across systems, institutions, and programs. While rough calculations of the overall "cost" (that is, expenditures of dollars received from the state and students) per student enrolled or per degree awarded are possible with current data, better information about the allocation of money across divisions, departments, and levels of instruction would be needed to understand the return on investment of different options for serving undergraduate and graduate students, lower- and upper-division students, and students in different kinds of programs.

> "I think we tend to take a more siloed view toward the data questions we ask. We ask questions about community college students, and maybe turn around and ask, 'Well, what about CSU?' and then 'What about K-12?' and 'What about UC?'"
>
> *–State policy staff member*

While some technical challenges remain in joining student data systems, interviewees said that the issues preventing California from aligning or coordinating its disjointed data systems are mostly political and cultural. Many of those from within the education systems voiced concerns about who would hold and control the data, who would get access to which data and for what purposes, and how the data would be analyzed and interpreted. Policy staff shared their concerns about the cost of developing a more comprehensive or coordinated data system, the difficulty of deciding on the capabilities and limitations of the system, and the complexity of undertaking the effort. Interviewees also said there is some inertia and resignation in the face of the challenges involved, a lack of leadership on this issue at the state level, and some uncertainty about where such leadership should be expected to arise.

> "The technology is the easy part. It's the politics, it's the proprietary treatment of the data, it's the personalities, it's the organizational interests that get in the way of getting there. It's the making sure that nobody's going to have to read their name in the newspaper and assuring them, that's the hard part."
>
> *–State agency staff member*

In the meantime, policymakers appear to recognize the need to improve student success and increase educational attainment to maintain California's economic standing and social health. For example, the state is investing millions of dollars to strengthen educational pathways and facilitate student transitions from high school into college, but what are the impacts of these programs on student outcomes? Evaluating the return on investment from initiatives like the Strong Workforce Program and the Guided Pathways effort is impossible, because the data to track student progress across education systems and into the workforce are not complete and are not publicly available.[21] Meanwhile, educators at the local level need access to information about students after they enroll in other institutions to assess and improve their programs and curriculum. Our research indicates that, while California has some fairly robust sets of data, connecting those data across disjointed systems is so challenging and difficult that it constrains the kinds of questions that policymakers, educators, taxpayers, parents, and students can ask of the state's public education systems.

> "The problem is that we keep setting policy, and there's no infrastructure to make good on it, so then nothing happens. It's a vicious cycle, because you put money in—and they're putting millions and millions of dollars in this state into initiatives—and none of [those initiatives] can succeed. It is impossible for any of these intersegmental efforts to succeed, absent a data system, period."
>
> *–Researcher*

In our next brief in this series, we will examine some local and regional data-sharing initiatives in California to understand better why educators engage in these partnerships, how these initiatives are functioning, and whether expanding such efforts could be a good alternative to developing a comprehensive statewide student data system. The fourth and final brief will describe efforts by a few other states to build comprehensive student data systems, with the aim of identifying potential implications for California. The fourth brief will conclude with recommendations to ensure that Californians have sufficient public data and analytical capacity to support the success of students throughout their educational trajectories.

Acknowledgments

We thank the policy staff, education officials, researchers, and others for their generosity in sharing their time and expertise. Their willingness to share candid assessments of California's current approach to student data was essential to this research. We are grateful for the research support provided by Brock Grubb and for the inspiration from Alice VanOmmeren about displaying data relationships in the CCC system. The review and comments provided by Andrea Venezia, Kathy Booth, Cindy Kazanis, Camille Esch, Jeanice Warden-Washington, Brock Grubb, Jacob Jackson, Patrick Perry, and Tessa Carmen de Roy helped us improve an earlier draft of the report. We appreciate Gretchen Kell's editorial assistance, and the graphic design work of the Foundation for California Community Colleges and the Williams Group. The authors take responsibility for all content.

Endnotes

1 Moore, C., Grubb, B., & Esch, C. (2016). *Gaps in perspective: Who should be responsible for tracking student progress across education institutions?* Sacramento, CA: Education Insights Center.

2 For other discussions of issues related to education data systems in California, see the following reports and articles: Friedmann, E. (2017). *Building intersegmental partnerships.* Stanford, CA: Policy Analysis for California Education; Barnum, M. (2016, June 29). The purge: California leaves researchers (and policymakers) in the dark by gutting education data. *LA School Report*, retrieved from http://laschoolreport.com/the-purge-california-leaves-researchers-and-policymakers-in-the-dark-by-gutting-education-data; California EDGE Coalition (2016). *Data to serve policies, programs and people: Reinventing California's education and workforce data systems.* Sacramento, CA: Author; Taylor, M. (2016). *Improving workforce education and training data in California. Sacramento*, CA: Legislative Analyst's Office; Warren, P. & Hough, H. (2013). *Increasing the usefulness of California's education data.* San Francisco, CA: Public Policy Institute of California; Vernez, G., Krop, C., Vuollo, M. & Hansen, J. S. (2008). *Toward a K-20 Student Unit Record Data System for California*. Santa Monica, CA: RAND Corporation; McKinsey & Company (2008). *Framework for a comprehensive education data system in California: Unlocking the power of data to continually improve public education.* New York: Author; Barondess, H. (2008). *California's emerging education data system: A status report.* Mountain View, CA: EdSource.

3 To encourage participation in our study and frank discussions, we offered anonymity to the people we interviewed and, therefore, do not identify the interviewees' names, the specific organizations they work for, or their roles/positions within their organizations.

4 CEDS is an initiative of the U.S. Department of Education to facilitate better education data collection and use across education sectors, from pre-K to postsecondary. More information about the initiative is available at https://ceds.ed.gov/Default.aspx.

5 DQC is an organization that advocates for better collection and use of student-level data to support effective policy and practice. More information about the organization and its efforts is available at https://dataqualitycampaign.org.

6 The No Child Left Behind Act, passed in 2002, authorized several federal education programs that are administered by the states. The law required states to test students in reading and math at particular points, and included reporting requirements that necessitated a state-level data system like CALPADS. The NCLB was replaced by the Every Student Succeeds Act (ESSA), passed in December 2015.

7 All colleges and universities that participate in federal financial aid programs are required to participate in IPEDS. More information about the system is available at https://nces.ed.gov/ipeds.

8 Ed-Data is a partnership of the CDE, EdSource, and the Fiscal Crisis Management and Assistance Team/California School Information Services (FCMAT/CSIS). See http://www.ed-data.org.

9 For a discussion of data challenges related to the CCPT, see McLaughlin, M., Lundy-Wagner, V. & Groves, B. (2017). *Two years into CCPT: Many challenges & great promise.* Boston, MA: Jobs for the Future.

10 Institutions do not report all the information they collect to their systemwide offices. For example, colleges may collect data related to student participation in Summer Bridge or other specific programs to use for evaluation of their academic and student support efforts, but not send that information to the system office. Each system office has a defined set of information that institutions are required to report.

11 For more information about Cal-PASS Plus, see https://www.calpassplus.org.

12 The Multiple Measures Assessment Project is a collaborative effort involving The RP Group and Educational Results Partnership/Cal-PASS Plus to develop, test, and assess the implementation of a common assessment instrument for the CCC (see http://rpgroup.org/All-Projects/ctl/ArticleView/mid/1686/articleId/118/Multiple-Measures-Assessment-Project-MMAP). The CTE LaunchBoard is a data system and set of tools built from Cal-PASS Plus data, intended to help community colleges assess student progress through college to transfer and careers. Access to the LaunchBoard is restricted to schools, community colleges, and universities that are members of Cal-PASS Plus (see https://www.calpassplus.org/Launchboard/Home.aspx).

13 For more information about the CCGI, see http://foundationccc.org/What-We-Do/Student-Success/California-College-Guidance-Initiative.

14 For more information about CORE Districts and the CORE Data Collaborative, see http://coredistricts.org.

15 For information about the Linked Learning Alliance's Analytics effort, see http://www.linkedlearning.org/certification-analytics/analytics.

16 NSC data cover more than 97 percent of postsecondary enrollment in the country, and over 90 percent of degrees awarded. For more information, see http://www.studentclearinghouse.org.

17 The CDE is currently exploring options for obtaining data from the NSC to match to student records in CALPADS.

Endnotes

18 As part of the Workforce Innovation and Opportunity Act (WIOA), the federal government required states to develop accountability dashboards covering adult education, Career and Technical Education in community colleges, and other workforce training programs offered through Workforce Investment Boards. In response, California Assembly Bill 2148 (Mullin, Chapter 385, Statutes of 2014) called for the development of an annual workforce metrics dashboard that "measures the state's human capital investments in workforce development and that provides, among other things, a status report on credential attainment, training completion, degree attainment, and participant earnings from workforce education and training programs." The text of AB 2148 is available at http://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140AB2148

19 A short description of the CPEC was included on page 6 of the first brief in this series, available at http://edinsightscenter.org/Portals/0/ReportPDFs/gaps-in-perspective-brief.pdf.

20 Assembly Bill 1056 (Fong, Chapter 620, Statutes of 2011) required the CCC Chancellor's Office to implement a system for the electronic receipt and distribution of student transcripts. While the eTranscript system was in use prior to the legislation, the bill led to the development of the Electronic Transcript Standard for California, and a process for delivering CCC district-wide transcripts to trading partners. The Chancellor's Office now pays the membership fees for all community colleges to participate. For a list of participating institutions, see https://www.etranscriptca.org/colleges-online.

21 The CCC Chancellor's Office has or is working on (via its partnership with Cal-PASS Plus) dashboards to capture metrics related to the Strong Workforce and Guided Pathways programs, but those dashboards are only available to the colleges and the Chancellor's Office.