

The 2019-20 Budget: Creating an Integrated Education Data System

GABRIEL PETEK
LEGISLATIVE ANALYST
MAY 6, 2019



Summary

Creating an Integrated Education Data System Has Been a Key Legislative Priority. California's education system is made up of numerous segments. Currently, each segment collects and maintains data on its students but the data generally are not linked across the segments. This limits the ability of policymakers, educators, researchers, parents, and others to get answers to many basic questions about student progression from early education through K-12 education, through higher education, and into the workforce. In recent years, the Legislature has commissioned intersegmental work groups to study and make recommendations on developing an integrated data system. Legislation resulting from the work groups' efforts was vetoed by the Brown Administration, citing cost and other concerns.

Recommend Legislature Build on Governor's Proposal and Prior Efforts. The Governor proposes \$10 million one time for the development of an integrated data system. Of that amount, \$3.1 million would fund a new work group to study and make recommendations to the administration on the governance, data structure, and other elements of an integrated system. The remaining \$6.9 million would be for data matching and implementation of the project ultimately approved by the administration. We believe the Governor's proposal has some positive components, but also some shortcomings. We recommend the Legislature adopt an alternative approach. Compared with the Governor's proposal, our recommended approach leverages planning that has already been done, uses available funding more effectively, and strengthens oversight of the project throughout the development and implementation process.

In this brief, we provide (1) background on the state's education data systems and past efforts to connect them, (2) describe the Governor's proposal to develop an integrated education data system, (3) assess the proposal, and (4) make associated recommendations.

Background

In this section, we provide an overview of California's education structure and education data systems; discuss past legislative efforts to create an integrated education data system; and describe required processes

for implementing state technology projects, including data systems.

California's Public Education System Is Comprised of Many Entities. Preschool in California is offered by many local entities, including schools and nonprofit organizations. The California Department of Education (CDE) administers annual contracts for these providers to operate. Kindergarten through grade 12 (K-12) is run by roughly 1,000 school districts, with elected boards governing at the local level and CDE providing support at the state level. CDE reports to the independently elected State Superintendent of Public Instruction. The

State Board of Education (SBE), with members appointed by the Governor, interacts regularly with CDE in guiding implementation of state education laws. Public higher education consists of three segments—the California Community Colleges (CCC), California State University (CSU), and University of California (UC). UC and CSU have system governing boards that oversee their 10 and 23 campuses, respectively. The CCC also has a system governing board, but its autonomy is more limited, with each of the system's 73 community college districts having its own local governing board. In addition to all these entities, California's public education system includes other agencies—including the Commission on Teacher Credentialing (CTC), which is responsible for the credentialing of K-12 teachers in California, and the California Student Aid Commission (CSAC), a state agency responsible for administering financial aid programs.

State's Education Data Systems Are Siloed.

Education data in California currently are siloed, being maintained and managed in several separate data systems. Most notably, CDE collects K-12 data (including student demographics, courses taken, grades earned, and high school graduation). For tracking and reporting purposes, CDE assigns to each K-12 student a unique, nonpersonally identifiable number (known as a "statewide student identifier"). Each of the public higher education segments collects and maintains its own student-level data and uses different student identifiers from CDE. State-subsidized early education programs (including most preschool programs) are part of a separate data system that is not connected with CDE's student data system and generally does not assign statewide student identifiers. Given that existing data systems are siloed, tracking and assessing students' progress across segments is difficult.

California Is One of Only a Handful of States Without Some Type of Integrated Education Data System. According to the Education Commission of the States, as of November 2016, California was one of only eight states neither possessing nor in the process of creating an integrated education data system. In about

two-thirds of states with integrated data systems, education and workforce agencies send a copy of student-level and wage records to a designated central repository ("data warehouse") on a regular schedule (such as at the end of each academic term). Staff then match the data records and assign a nonpersonally identifiable P-20 number to each student. The integrated data are then available to perform analyses and run reports. By comparison, other states have developed "federated" data systems, whereby each education and workforce entity collects and keeps its own respective data (as opposed to copying and sending it to another entity). Under the federated model, separate data systems connect to each other upon request. Each time a query is required to be performed for a report or other purpose, each segment "opens up" its database for a short period of time—long enough for the data matching and analysis to occur—at which point the segments lock up their data systems again. Federated systems typically have developed in states in which entities contributing data were unable to agree on the location and control of a centralized data system. Both types of data systems report aggregated information (averages and group statistics), which, for privacy purposes, is stripped of any personally identifiable information.

Some Cross-Segmental Data Sharing Occurs in California. Although California has no comprehensive P-20 data system, some cross-segmental data efforts exist. Among the most notable of these efforts are:

- Most school districts, every community college, and some CSU, UC, and private four-year institutions voluntarily participate in Cal-PASS Plus, a project overseen by the CCC Chancellor's Office. Cal-PASS Plus collects and links data from participating institutions, then analyzes the data. It provides ongoing dashboards and reports to participating institutions—confidentially sharing outcomes with them so they may learn more about their students and how to improve their outcomes.
- Thirty-nine school districts participate in the California College Guidance Initiative, which

is housed within the Foundation for California Community Colleges. School districts upload verified academic transcript data into students' accounts on CaliforniaColleges.edu. When students from participating districts apply to CCC or CSU, certain high school data is shared, with colleges being able to use the transcript data to inform decisions about admissions, course placement, and student services.

- Each public higher education segment has an agreement with the Employment Development Department (EDD) that allows it to identify the quarterly earnings of its graduates. The data are matched using social security numbers, which most students provide when they apply to college.

A Former State Agency Previously Housed Higher Education Data. While it is now defunct, the California Postsecondary Education Commission (CPEC) formerly collected and maintained data from the three public higher education segments to support its statutory research and advisory responsibilities. CPEC's data system collected information such as transfers between CCC and universities, degrees awarded across the three segments, and statewide enrollment projections. When then-Governor Brown vetoed funding for CPEC in 2011, the agency's data system was shuttered.

Statute Created Two Work Groups to Advise Legislature on Creating Integrated Education Data System. For years, the Legislature has expressed a strong desire to enhance its oversight role of California's education system and facilitate continuous improvement in education by developing a P-20 data system. In an effort to spur the development of such a system, the Legislature enacted Chapter 561 of 2008 (SB 1298, Simitian). Chapter 561 created two work groups (referred to as "SB 1298 work groups"). The first work group, led by the state's Chief Information Officer, was charged with identifying the technical steps toward implementing an integrated education data system. The second work group was charged with making recommendations on the governance

of such a data system. The two work groups included many education representatives, including from the Superintendent of Public Instruction, the CCC Chancellor's Office, the CSU Chancellor's Office, and the UC Office of the President. The governance work group also included legislative staff from both houses and both parties as well as the Department of Finance. In addition, governance work group members invited the participation of an advisory group, which included teacher, school administrator, and school board associations; CTC; and other organizations, associations, and government entities.

SB 1298 Technical Work Group Identified Roadmap to Creating Integrated Education Data System. The technical work group met throughout 2009 and completed its report in early 2010. The report identified a series of recommended steps to undertake over a two-year period to build what could become either a centralized or federated data system. These identified steps included (1) identifying the key questions to be answered by the integrated data system; (2) creating standard data definitions across the segments; (3) assigning a unique identifier to each student in the P-20 system; (4) establishing policies and procedures to ensure security and privacy of student data; and (5) creating a single portal and data tools (such as dashboards and reports) for policymakers, educators, researchers, and the public to access the data. The work group also recommended that the data system be built with the capability to eventually link education records with records in health, social services, criminal justice, and other state data systems.

SB 1298 Governance Work Group Presented Three Options. The governance work group submitted its report in late 2009. The work group envisioned creating a centralized data warehouse and recommended that it be governed by a single entity whose mission crosscuts K-12 and higher education. The group recommended that this entity include representation from each of the contributing education segments as well as a "significant majority" of state-level, non-segmental members (including bipartisan representation from the Assembly and Senate and staff from

the administration). Work group members recommended arranging governance this way so as to mitigate potentially negative incentives for the segments to control data and use it primarily for segment-specific interests. The group did not settle on a particular structure for the new entity but put forward three options—creating a joint powers authority (JPA), creating a new state agency, or housing the data system within an existing state agency.

Legislature Authorized Creation of JPA but Governor Vetoed Bill. In response to the governance work group’s report, the Legislature passed SB 885 (Simitian) in 2012. SB 885 authorized the creation of an intersegmental JPA to plan, implement, and manage a P-20 data system. In addition, SB 885 called on the JPA to plan the new data system, with the intention of eventually incorporating data from noneducation sources, including employment, health, and corrections agencies. Though SB 885 received nearly unanimous support from legislators, then-Governor Brown vetoed the legislation, citing, among other reasons, concerns about cost and the state’s “current fiscal constraints.” Since that time, the Legislature has introduced various bills to create some type of integrated data system, with those bills either failing to make it out of committee or being vetoed by Governor Brown.

California Has Process for Approving Information Technology (IT) Projects. Led by the state’s Chief Information Officer, the California Department of Technology (CDT) is responsible for reviewing and approving IT project proposals developed by most state departments. CDT has multiple stages to its project approval process. Each stage requires departments to conduct specific planning-related analyses and submit an associated planning document to CDT. Collectively, the planning documents from these stages create a comprehensive plan for implementing a proposed IT project. Once CDT approves a department’s project proposal, CDT’s role typically changes to providing project oversight. Specifically, CDT provides independent review of the project—monitoring whether it remains within budget, on schedule, and on track to achieve its established objectives. In its

project monitoring reports, CDT identifies issues of concern, shares lessons learned from other projects, and recommends strategies to reduce project risks and fix identified issues. Current law generally requires CDE and the CCC Chancellor’s Office to use the CDT planning process for IT projects costing more than \$1 million. CSU and UC are not required to go through CDT’s process.

Governor’s Proposal

In this section, we describe the Governor’s proposal to study and fund an integrated data system.

Declares Intent to Create Integrated Data System. In proposed trailer bill language, the administration indicates the integrated system would be intended to benefit a number of groups, including policymakers, teachers, student advisors, parents, students, and researchers, as well as health and human services providers. The proposed trailer bill language gives significant discretion to a work group to determine what such a data system might look like. For example, the data system could be centralized or federated. According to the Department of Finance, the administration’s long-term goal would be to expand the data system so that educators and other staff across segments and agencies could access student data. Under such a vision, the data system could be used as a tool to provide academic advice and connect students to health and other services.

Authorizes State Board of Education to Lead Data System Work Group. As **Figure 1** shows, the proposed work group would consist of ten education and other departmental agencies, with a total of at least 11 representatives. (CDE would have at least two representatives—one from its early learning and care division and the other from its analytical and reporting division.) The Governor would select the members of the work group based on nominations provided by each entity. The executive director of the SBE would convene and lead the work group. With the approval of SBE, CDE would contract with third party entities (referred to as “planning facilitators”) to staff the work group.

Work Group Charged With Making Recommendations to Administration on Data System.

Under the proposed trailer bill language, the work group would be charged with studying and making a number of recommendations. The planning facilitators would compile these findings and recommendations into two required reports:

- The first report, due by March 2020 (or later if approved by the Department of Finance), is to include the work group's advice on matters such as (1) the type of data system to create, (2) the data elements that would receive the highest priority to be included in the data system, (3) the entity that would manage and secure the data system, (4) the protocols for protecting student privacy and addressing data security risks, and (5) the entities or persons that would have access to the data.
- The second report, due by September 2020 (or later if approved by the Department of Finance), is to summarize the work group's recommendations for "expanded and enhanced data system functionality," including (1) plans to expand the data system to "incorporate workforce, financial aid, and health and human services data," (2) steps to increase data quality provided by each entity contributing to the data system, and (3) a proposed timeline and budget for implementing its recommendations.

The work group would be required to submit both reports to the Department of Finance. According to the Department of Finance, the administration would then decide how best to respond to the work group's recommendations.

Requires Public Education Segments to Develop Way to Match and Share Student Data. Regardless of the work group's ultimate recommendations, the proposed trailer bill language requires CDE as well as the CCC and

Figure 1

Governor's Proposal Would Create Work Group Consisting of Ten Entities

- State Board of Education (leader)
- California Department of Education
- California Community Colleges Chancellor's Office
- California State University Chancellor's Office
- University of California Office of the President
- California Commission on Teacher Credentialing
- California Student Aid Commission
- California Labor and Workforce Agency
- Employment Development Department
- California Health and Human Services Agency

CSU Chancellor's Offices (and requests the UC Office of the President) to perform two activities within certain time frames. By December 2019, these segments must "develop a means" of connecting student records for K-12 students at the time of their enrollment in CCC, CSU, or UC. By December 2020, the planning facilitators are to ensure that the same segments begin implementing this data linkage.

Provides \$10 Million One-Time Non-Proposition 98 General Fund for Data Efforts. These funds would be appropriated to CDE for allocation purposes. Of the total amount, \$2 million would be available for CDE to contract with work group planning facilitators. In addition, CDE would receive \$200,000 to cover any costs associated with its planning activities. The other nine entities would receive \$100,000 each for work-group planning activities. In addition, each of the four public education segments would receive \$50,000 to develop a means of matching student data. The remaining \$6.7 million would be available to implement the initial phases of the data system or for the subsequent "expansion and enhancements" phase, contingent on the Department of Finance's approval. The proposed trailer bill language does not require the project to go through any part of CDT's review or approval process.

Assessment

In this section, we provide our assessment of the Governor's proposal. While we agree with the

Governor that having an integrated data system would be an asset for the state, we have a number of concerns with his proposed approach to developing one.

An Integrated Education Data System Could Have Significant State- and Local-Level Benefits.

In our view, California could benefit significantly from an integrated education data system. Each year, the state provides billions of dollars for education. Over the years, it also has made major policy changes. For example, over the past decade, the state has implemented the Local Control Funding Formula with an aim toward improving students’ college and career readiness, funded several new career technical education programs intended to improve the alignment of high school and college courses, and streamlined transfer pathways for CCC students to enroll in upper-division coursework at the universities. Lacking an integrated data system, the Legislature has had only a limited ability to assess the impact of those changes. Having an integrated data

system in place could help the Legislature better exercise its oversight role and better assess the effectiveness of its policies. **Figure 2** provides examples of the kinds of crosscutting policy questions the Legislature could answer with an integrated data system. Such a data system could also help local educators evaluate and improve their practices and provide useful, timely information to students and families.

Governor’s Proposed Approach Minimizes Legislature’s Role . . .

Though an integrated data system could provide notable benefits, we have serious concerns with the Governor’s proposed role for the Legislature in developing the system. Under the proposed trailer bill language, the only responsibility for the Legislature would be to appropriate the \$10 million in one-time funds. All other core responsibilities—including choosing the work group representatives, reviewing work group reports, deciding on the governance structure, selecting the data system structure, and approving \$6.7 million in funds for project implementation—

Figure 2

Questions That Could Be Answered With an Integrated Education Data System

1. Which early education programs and services have the greatest effect on reading and comprehension in elementary school?
2. What are the demographic, program, and course-taking profiles of K-12 students who enroll or do not enroll in postsecondary education?
3. What are the characteristics and educational paths of students who drop out of high school but eventually enroll in a postsecondary institution?
4. Is how districts use their supplemental grants under the Local Control Funding Formula affecting the proportion of their low-income students who enroll in and graduate from college?
5. What are the postsecondary enrollment and completion patterns of students in high school career technical education (CTE) pathway programs compared with similar students not in a CTE pathway?
6. Does dual (concurrent) enrollment by high school students in college courses promote more timely and efficient completion of associate and bachelor’s degree programs?
7. Do students who earn an associate degree for transfer (ADT) at a community college end up taking fewer total units to earn a bachelor’s degree than students who transfer without an ADT?
8. Are students receiving Cal Grant competitive awards more likely to enroll and graduate from college than those eligible students who just missed the cut-off for getting awards?
9. What are the employment outcomes of graduates from CSU and UC teacher preparation programs?
10. Which health and social service programs are most closely associated with improved educational outcomes of K-12 and college students?

would fall exclusively to the Governor or Department of Finance. In effect, the Legislature would be appropriating funds for a project without knowing what it would get in return and without any assurance that the final product would be consistent with its priorities.

. . . As Well as Sidelines CDT From Planning, Development, and Implementation Process.

Equally as troubling is the lack of any planning and oversight role for CDT in the proposed project. According to the Department of Finance, the project would not include CDT's involvement due to the administration's desire to move quickly. Given the state's historically mixed track record with IT projects, we question the prudence of bypassing CDT's planning and oversight functions. Completing CDT's project approval process could improve the quality of the project and provide the Legislature with a more complete plan before funding the project. CDT's independent oversight also could highlight risks and alert stakeholders (including the Legislature) to any significant changes in the cost, schedule, and scope of the project.

State Has Already Given Much Thought to Creating an Integrated Data System. Over the years, the Legislature, education segments, and related entities have deliberated at length on creating an integrated data system. As a result of these efforts, the segments already have produced a technical roadmap for creating such a system. Moreover, the Legislature even passed legislation authorizing the creation of a JPA for purposes of governing a new data system. Given this prior work, we are concerned with the administration's approach of effectively starting anew and authorizing another work group. We also are concerned with providing \$3.1 million for planning facilitators and the new work group to study and make recommendations on many of the same fundamental issues as former work groups.

JPA Governing Model Remains a Promising Approach. As noted by the SB 1298 governance work group, a JPA has several advantages that could help ensure the successful implementation of an integrated data system. With a JPA approach, the entities that know their data best (the data contributors) have a direct role in developing and

managing the new system. By having a majority of JPA representation come from outside the education segments—as also recommended by the SB 1298 work group—the state could achieve the benefit of ensuring the segments do not limit access to data or use the data for largely self-serving purposes. A JPA has the added potential benefit of leveraging some administrative and technology infrastructure already in place at the segments—resulting in possible efficiencies and the ability to organize a governing entity relatively quickly.

A Centralized Data Warehouse Has Advantages Over Federated System. Centralized data systems are the most common among states and have distinct advantages over federated systems. A federated system requires the segments to perform a data match every time a query is made—requiring them to respond to every individual data request. With a data warehouse, once the data are uploaded and matched, the data are available for staff to respond to any number of queries simultaneously. As a result, centralized systems tend to be more efficient—enabling access to more quickly produced data analyses.

Governor's Incremental Approach to Adding Data Sources Has Merit. We believe the Governor's proposed order of onboarding the segments and other data contributors generally is reasonable. Recognizing that all IT projects entail risks, first connecting K-12 data with data from the state's three public higher education segments, as proposed by the Governor, likely would be the most straightforward technical process, with considerable added value. The proposed trailer bill language is not explicit on whether early education data and EDD data should be included as part of the first phase. Given the generally fragmented state of current early education data, adding child care and preschool data in the first phase seems premature. Because the higher education segments already match student records with EDD, including EDD wage data in the initial roll out, however, likely would be feasible. We think adding student financial aid data to the system also would have considerable value, but CSAC is in the midst of replacing its existing data system. Waiting until

CSAC's IT upgrade has been completed could help make its inclusion in the integrated system more successful. After completing and testing the initial phases of the integrated system, the Legislature could consider adding data from other agencies, such as health and human services agencies.

Recommendations

In this section, we lay out a recommended series of steps for the Legislature to authorize and fund the development of an integrated education data system.

Recommend Enacting Legislation Creating a JPA. Similar to SB 885 (Simitian), we recommend the Legislature authorize the creation of a JPA to develop and maintain an integrated data system. We recommend the legislation specify that each education segment contributing data be a member but state-level, non-segmental members form a majority on the JPA. We recommend specifying that the non-segmental members include representatives from the administration as well as bipartisan representation from the Legislature.

Direct JPA to Focus Initially on Data From Core Education Segments and EDD. We recommend the legislation also define the initial scope and long-term vision of the project. We recommend the initial goals be to connect data from CDE (K-12), the three public higher education segments, and EDD. The legislation could state the longer-term intent that, once those connections are successfully made, adding data from other entities (such as from early education providers, CSAC, and health and human services agencies) be considered. Eventually, the state might consider adding even more data sources, such as from corrections and private schools. Each phase of

expansion, however, is likely to come with new challenges and risks.

Specify Expectation for JPA to Create Data Warehouse and Adopt Certain Policies. In addition, we recommend the legislation specify that the JPA develop a centralized data warehouse, update the SB 1298 technical work group's plan based on this expectation, and follow CDT's project approval process (including for the university components of the project).

Approve One-Time Funding, With Reporting Requirements. We recommend the Legislature approve \$10 million in initial funding for the JPA and provide associated expenditure authority through June 30, 2023. To ensure the Legislature is kept apprised of the project's progress, we recommend the Legislature appropriate the funding in the annual budget act and require the administration use the section letter process to notify the Joint Legislative Budget Committee prior to allocating any round of funding. We recommend making approval of initial rounds of funding conditioned on the JPA meeting certain milestones, including:

- Submitting a multi-year work plan to the Legislature by March 1, 2020.
- Submitting a budget to the Legislature by May 1, 2020.
- Progressing in a timely manner through CDT's planning phases.
- Developing protocols for maintaining data privacy and security.
- Developing protocols for making the data available for research purposes.

CONCLUSION

The administration's goal to create an integrated education data system for California is in line with a longstanding priority of the Legislature. Such a data system could give the Legislature a more holistic view of the state's education system and allow policymakers to make more informed budget and policy decisions. An integrated data system also could provide more information to educators about what happens to their students after leaving their particular education segment, thereby providing greater insight into the effectiveness of current

practices. Our review of the Governor's specific proposal to create such a system has identified some promising components, but also a number of shortcomings. Compared with the Governor's proposal, our alternative set of recommendations leverages planning that has already been done, uses available funding more effectively, reduces the risk of the project failing by requiring it to go through established state review processes, and strengthens legislative oversight throughout the development and implementation process.

LAO PUBLICATIONS

This report was prepared by Paul Steenhausen and reviewed by Jennifer Kuhn Pacella. The Legislative Analyst's Office (LAO) is a nonpartisan office that provides fiscal and policy information and advice to the Legislature.

To request publications call (916) 445-4656. This report and others, as well as an e-mail subscription service, are available on the LAO's website at www.lao.ca.gov. The LAO is located at 925 L Street, Suite 1000, Sacramento, CA 95814.