



Modernizing California's Education Data System

Jacob Jackson | Kevin Cook | Research support from Courtney Lee

Supported with funding from the Bill and Melinda Gates Foundation and the Sutton Family Fund

Summary

Higher education is a key driver of economic mobility and future growth in California. Yet there is much we do not know about how students advance from K–12 schools to postsecondary education and into the workforce—and where they falter. This lack of knowledge stems from the fragmented nature of California's education data and inhibits informed decision making among policymakers and educational leaders.

This report reviews existing research to examine the shortcomings of the status quo, identify the benefits of a statewide longitudinal data system, and outline steps to make a new system as effective as possible. Our recommendations:

- **California should develop a statewide longitudinal database that can track students across educational institutions and into the workforce.** California is one of only a handful of states without a student data system that can answer important questions about the educational pipeline and the impact of education on work and earnings. An integrated data system would also encourage stronger collaborations among institutions to improve student outcomes.
- **Multiple stakeholders should help determine the structure of the data system.** Policymakers and educational leaders should begin by establishing the key questions they want answered. Engaging outside of government—with the business community, regional education initiatives, researchers, and the broader public—will also be essential to shape participation in and governance of the data system.
- **A system designed for growth would benefit California in the future.** Including data from California's public K–12 and higher education systems as well as workforce agencies in a central repository would lay a strong foundation. If the state begins with a flexible centralized system, later it will be easier to add private institutions and other government agencies, which would increase the value of the data system.
- **Other states can serve as models on governance, privacy, and security issues.** California can learn from states that are already using comprehensive data systems to evaluate the impact of education policies. Many successful models exist, demonstrating that such systems can work without compromising data security and student privacy.

Promoting student success and institutional effectiveness in California ultimately requires a better understanding of how prior educational experiences affect students' subsequent academic achievement, work, and earnings. Given the gubernatorial transition in the state, the time is ripe to develop an integrated data system that will help California students thrive in their academic career and beyond.

Introduction

Unlike most other states, California does not have a data system to track students' pathways from K–12 schools to college and into the workplace—putting the state and its students at a distinct disadvantage. A statewide longitudinal data system is necessary for California to evaluate which policies and investments have been most effective in improving student outcomes. The lack of an integrated database also hinders coordination across educational systems and limits the state's ability to track progress on specific educational goals.

Over the past decade, the state has invested billions of dollars to improve its public education systems, including the creation of multiple new programs to streamline the educational pipeline so students can better navigate the path from preschool to the workforce. During this time, many institutions have seen steady improvements in educational outcomes, such as persistence and graduation rates. But insufficient information about how students fare during key transition points makes it difficult to assess which programs and interventions have been most effective, which are not worth the investment, and what we could do to improve them. Although some institutions share data across sectors for research or practical purposes, these connections are mostly infrequent, inefficient, incomplete, or ad hoc.

PPIC and other research organizations have long called for a statewide longitudinal data system (Moore and Bracco 2018a; Reed et al. 2018; California Competes 2018; Warren and Hough 2013; Phillips et al. 2018). With new policies that focus on students' transitions along the educational pipeline and a gubernatorial leadership change, the time is ripe to invest in a data system that will improve education in California. We highlight the benefits to students and the state offered by a statewide longitudinal data system and suggest how to get the most out of a statewide database.

California's Education Data Systems Are Fragmented

Using data and evidence in educational decision making is critical to student success. Recognizing this fact, educational institutions collect and analyze vast amounts of data on their own students to improve programs and services and to inform the public of their progress. However, it is currently difficult to study the critical transition points that occur as students advance in their academic career and move into the workforce.

In fact, there are many important questions that can only be answered if we take a holistic approach to students' educational experiences and connect student data across educational sectors (Moore and Bracco 2018a; California Competes 2018). These are a few overarching examples:

- **The educational pipeline.** Which students successfully transition from high school to college and from community colleges to four-year colleges? Which of those successfully earn degrees? Are we appropriately placing students in remediation? What are the equity implications of these transitions?
- **State investments in higher education.** How much does it cost the state to put students through the various pathways to a four-year degree (e.g., starting at a four-year college versus a community college)? How much does it cost a typical student to attain a four-year degree?
- **Work and earnings.** What are the impacts of different degrees or certificates on earnings? How can California schools and colleges produce the right workforce for the state?
- **Program evaluation.** Is the new K–12 funding formula producing better postsecondary and work outcomes, especially for students from schools that have received increased funding?

The ability to answer these questions is key to assessing the impact of California's recent educational reforms, many of which are critical to the state's mission of producing more college graduates and reducing equity gaps. For example, the Common Core State Standards explicitly promote college and career readiness for K–12 students. However, schools generally receive little feedback on which of their former students went to college or entered the workforce. Furthermore, schools often do not know whether their students were ready for college-level work if they did

enroll in college, or whether they eventually earned a degree. Schools need feedback if they are to improve, but there is currently no systematic way to connect and share this information.

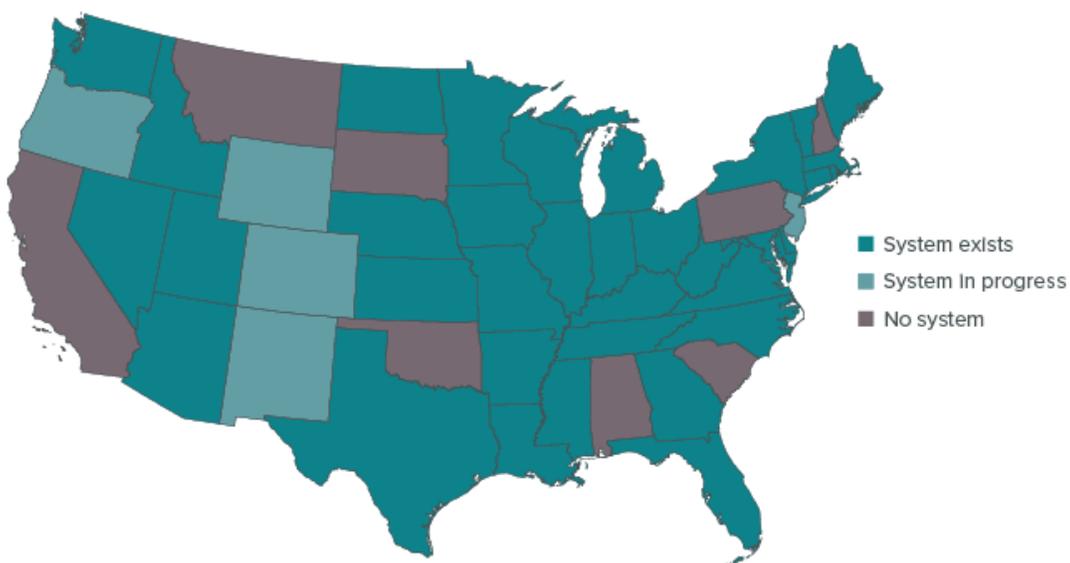
Within higher education, data sharing between the three public systems—California Community Colleges, California State University (CSU), and the University of California (UC)—would enhance the efficacy of the educational pipeline. Several new programs and policies focus on streamlining the transfer pathway, which can often be complicated and confusing for the many students who enter community college with the plan to transfer to a four-year institution and earn a bachelor’s degree. For example, the Associate Degree for Transfer between California’s community colleges and the California State University system aims to increase transfer rates and guarantees that students have to complete only two more years of coursework to graduate from a CSU campus. However, evaluating the success of the program is challenging because the community college system does not have information on the graduation rates and time to degree of students who transfer.

The need for cross-sector cooperation has led to a few partnerships, but current connections are insufficient. Some regional consortia (e.g., Long Beach College Promise) and statewide consortia (e.g., Cal-PASS Plus) share data across sectors, but these systems are usually incomplete (Moore and Bracco 2018a). Partnerships tend to have a limited number of participating institutions and provide at best a partial picture of who enrolls in postsecondary institutions, especially since students often travel outside of their home regions to attend college.

Benefits of a Statewide Longitudinal Data System

Educational leaders and policymakers need better data to identify problems and develop effective solutions. As shown in Figure 1, California is one of only eight states without a statewide longitudinal data system—linking student-level data across time and educational sector—or plans to create one (Education Commission of the States 2016). Some statewide data systems link certain segments (e.g., K–12 and postsecondary education) but not others. Altogether, sixteen states and Washington, DC, have systems that link early learning, K–12, postsecondary, and workforce data—providing a detailed portrait of students’ educational trajectories and work outcomes. Such a robust data system is instrumental to both providing feedback for institutions and evaluating state-level policies.

Figure 1. California lags behind other states in establishing an integrated education data system



SOURCE: Education Commission of the States, 2016.

NOTE: States in the figure with longitudinal data systems link data from at least two sectors of education and/or the workforce. Education sectors include early childhood, K–12, and postsecondary. Alaska and Hawaii are not pictured, but they do have longitudinal data systems.

FEEDBACK AND EFFICIENCY FOR INSTITUTIONS

Institutions benefit from knowing more about the educational trajectory of their students and responding to their needs. As California's community colleges implement Assembly Bill (AB) 705, which requires the use of high school records as the main criteria for placement into college-level coursework, they need access to high school grades and courses. Currently, most colleges rely on self-reports of high school courses and grades or require students to provide a high school transcript. An integrated data system would allow community colleges to more efficiently place incoming students into the appropriate courses, potentially avoiding costly remediation.

Universities can also benefit from knowing more about the trajectories of students who do not attend their institutions. For example, UC currently operates over a dozen separate programs intended to improve the academic preparation of high school students in underserved areas, increase college enrollment (especially at UC), and close achievement gaps. A system that connects K–12 and higher education data would shed light on the outcomes of students who participated in these programs but did not attend a UC, and could guide outreach efforts.

An integrated statewide data system could also free up local resources. Many regional partnerships recognize the value of analyzing the entire educational pipeline and already share data across sectors. But given the amount of resources that communities and institutions spend to construct and maintain local data partnerships, a statewide system could offer time and cost efficiencies. It would not preclude any institutions from working together or sharing data but would instead provide a firm foundation on which regional partners could base their work. In addition, since students often move across regions for school or work, a statewide system could serve to fill in gaps that regional efforts now face. Indeed, some regional entities see a statewide solution as helpful to their goals (Moore and Bracco 2018a).

EVALUATION AND COORDINATION FOR THE STATE

A more robust data system would also lead to better evaluation and coordination across K–12 and higher education systems. Without accurate data on issues like transitioning from high school to college, transferring from community college, time to degree, and job experiences after graduation, policymakers lack evidence about how well the education system overall addresses the needs of students and the state. By providing this “eagle’s eye” view, an integrated data system would help the systems collaborate in a way that aligns with the state’s goals.

A statewide data system could also help the state and local education officials determine if they are making the best program investments or if other policy interventions are warranted. For example, Texas’s education data system was instrumental in assessing the effectiveness of dual-credit coursework, college-level coursework that students complete in high school. Using the state’s data system, which was established in 2006 and houses preK–12 and higher education data, researchers found that dual-credit coursework is a promising strategy for increasing postsecondary access, persistence, and completion. Additionally, the authors compared the efficacy of dual-credit courses against advanced courses such as Advanced Placement (AP) or International Baccalaureate (IB), and their results suggest that dual-credit courses are associated with better college outcomes (Giani, Alexander, and Reyes 2014).

A statewide data system for California could help answer our own policy questions. Though it’s not the norm, occasionally researchers have been able to connect data across systems to gain new and valuable insights about California’s educational pipeline. For example, the research projects below would have been impossible without linking data across educational sectors:

- Linking records between community colleges and California’s Employment Development Department allowed PPIC researchers to show that health career education credentials generally have sizeable economic returns. Researchers also highlighted which certificates resulted in higher wages and found that “stacking” multiple, related credentials in health fields provides significant economic benefits. These findings can help inform program development to ensure students continue to benefit from wage increases (Bohn, McConville, and Gibson 2016).

- Cal-PASS Plus data, which includes high school and postsecondary records, allowed PPIC researchers to determine that 30 percent of students who had successfully completed college preparatory courses in high school were placed into remedial courses at community colleges. This research further identified potential problems in community colleges' placement procedures and could help high schools pinpoint the best timing for interventions to ensure students are ready for college (Gao and Johnson 2017).
- Linking community college records to high school records allowed UC Davis researchers to show that high school student characteristics play a big role in determining community college outcomes and rankings. This research showed how college rankings change after accounting for the academic and demographic characteristics of incoming students. This is especially important given the movement toward performance-based funding for California's community colleges (Kurlaender, Carrell, and Jackson 2016).
- Combining data from a California State University campus with data from the nearby California school districts allowed UC Davis researchers to investigate whether a statewide policy—the Early Assessment Program—reduced the need for remediation at CSU. They found early testing and notification regarding high school students' preparation for higher education reduced the need for remediation among students who eventually enrolled at the CSU campus. Many states, including California, now use a similar early warning system with their Common Core 11th-grade tests (Howell, Kurlaender, and Grodsky 2010).

Though linking data has been feasible in some instances, the lack of a comprehensive system is still a major obstacle. Researchers must come to data agreements with multiple institutions—one likely reason such cross-sector work is infrequent. Researchers must then navigate technical roadblocks that arise from combining data systems that were not designed to work well with each other, which can yield incomplete information when answering these and other questions.

Getting the Most out of California's Education Data

A statewide longitudinal data system could take many shapes, but its structure and governance should be aligned to the important questions that the state and its education systems want answered. These questions should determine who participates, who governs the system, and which data get collected. In order to get the most out of a data system, the data should cover the entire educational pipeline—from K–12 to workforce—and be governed in a way that can ensure cooperation between institutions, guarantee transparent collection and use, meet security and privacy concerns, and ultimately benefit students, educational institutions, taxpayers, and the state.

DATA COLLECTION AND STORAGE

The most beneficial data system would act as a statewide student-level data repository. Each participant in the data system would securely submit student or worker data at predetermined intervals to a central repository where the data could be linked by a unique identification number and stored. Compared to the alternative, where the systems themselves link data on a project-by-project basis, a repository would allow for faster, repeated, and more varied research reports and save the systems from needing to respond to every individual data request. Most states that have longitudinal data systems use a centralized data repository model (Education Commission of the States 2016). Research by Moore and Bracco (2018b) at Education Insights Center further suggests a centralized data repository would best meet California's needs.

California could start relatively small and add agencies to the data system along the way. The participation of the K–12 system, California Community Colleges, California State University, and the University of California would answer most questions about the educational pipeline and allow for better coordination among the state's public education systems. The addition of workforce data would be necessary to answer other key questions about employment outcomes. Institutions could begin by creating and using a student identification number that is kept constant across systems and reporting only institutional data that they already collect, such as demographic information, grades, and coursework. This would allow researchers to track student experiences across institutional boundaries and could be developed relatively quickly.

Scaling up to a more robust system would enable the state to answer a broader range of questions. The system could be built up over time to include more information on students' educational and life experiences:

- **Program participation.** Data on specific interventions could allow for evaluation of specific programs and policies. For example, student participation in CSU's new California Promise Program—in which students agree to take a full load of courses in exchange for priority registration and more academic advising—could be reported and analyzed at a statewide level.
- **Private colleges and universities.** At least 25 percent of California postsecondary students attend private institutions. These institutions are an essential part of the state's educational pipeline and would be a vital source of information on student outcomes. Private colleges would also benefit from knowing more about students who do and do not attend their institutions.
- **Other education sources.** Other California and nationwide organizations keep important data on students. Two examples are the California Student Aid Commission, which maintains information on financial aid distribution for California students, and the National Student Clearinghouse, which tracks students who enroll in private institutions and institutions in other states. More broadly, the state may also wish to include data on early learning and/or adult education and licensing institutions.
- **Other related governmental agencies.** Data from other state agencies could shed light on the impact of various services and programs. For example, including the Department of Social Services, which houses data on the state's food assistance and welfare programs, could allow the state to evaluate participants' educational and employment outcomes.

More data could be advantageous for California in the short and long term. But additional data sources would add to the system's technical and logistical complexity as well as its costs. If supplemental features serve as roadblocks to starting a helpful data system, a better option is to start simple with the flexibility to expand as necessary.

GOVERNANCE AND USE

A data governance body would likely be responsible for meeting many of the challenges associated with a longitudinal data system. The body would have to determine transparent and secure data acquisition and linking, make decisions about who can use the data and for what purpose, and protect student privacy. These are all critical challenges, and it is important to note that many other states have solved them when establishing and administering their own databases. California can benefit from their successes and missteps by engaging with experts from other states as California plans and implements its own system. For example, learning how other states navigate federal and state privacy laws can help California think through how best to protect student privacy. Research by Phillips, Reber, and Rothstein (2018) includes a detailed discussion of issues related to privacy and statewide longitudinal data systems.

The governing body would also need to determine a way to work with institutions, policymakers, regional consortia, researchers, and the general public to advance the agenda of the state. After all, the real value of a data system lies in how it is used. Regions will want to tap into the statewide data to study educational pathways within their region, as well as the educational trajectories of students that come into and out of their region. The general public also stands to benefit from tools that can help answer basic questions. For example, many states with data systems make some information—such as college attendance and college graduation rates by high school—available to the public.

After the data system is established, the governing body should partner with outside researchers in order to expand capacity. As shown in the previous section, outside researchers are interested in the same kinds of questions that the state is interested in and have the ability to independently and rigorously evaluate policies and uncover solutions to vexing problems. State education agencies from three states that adopted longitudinal data systems published an outline of the benefits and challenges of working with external partners, recognizing that researchers are essential to maximizing the promise of their student data systems (Conaway, Keesler, and Schwartz 2015).

Which organization is best suited to serve as the governing body? In some states, a higher education coordinating body runs the state longitudinal data system. In other states, the data system is run by a separate state or non-state agency, or it is a collective effort of the public education systems (Moore and Bracco 2018b). In the past, California had a higher education commission that managed some student data across higher education sectors. If reestablished, this commission's success could depend on having more detailed and robust education data. But this does not necessarily mean that such a commission should administer the data system. The text box below examines the pros and cons of having a higher education commission manage an integrated education database. Research published by the Education Insights Center suggests a separate state agency could best meet California's needs for data governance and includes a detailed analysis of different types of governance systems (Moore and Bracco 2018b).

Should a Higher Education Commission Manage the Data System?

California currently lacks regular, systematic coordination between its higher education entities. The former coordinating body, the California Postsecondary Education Commission (CPEC), was shuttered in 2011. As a part of its coordination efforts, CPEC collected, stored, and published aggregated data across sectors. Several bills to reinstate a higher education coordinating body have made their way to the governor's desk in the past few years only to be vetoed. The gubernatorial transition may open a window for reestablishing a higher education commission.

If a new commission were established, developing and maintaining a statewide longitudinal data system could be essential to its work. A coordinating body could use the data to assess the strengths and weaknesses of the educational pipeline, evaluate reforms, and determine how to improve coordination between institutions and systems on issues such as the transfer pathway.

However, keeping the data governance separate from a higher education commission has some advantages. While higher education is a critical component of the education-to-workforce pipeline, to be truly useful, the data system would extend into K–12 and the workforce—and possibly include other related governmental programs like social services. If these agencies are included, their input should be reflected in how the data is managed. Second, housing the data system elsewhere could perhaps offer greater stability. When CPEC was shut down in 2011, with it went a major data source in California. Lastly, while a robust data system could help the state better coordinate its higher education systems, it would have many users beyond higher education as well. This wide range of stakeholders might be better reflected in an alternative governance structure.

Conclusion

In recent years, California has transformed many aspects of its educational system in hopes of improving student outcomes and reducing achievement gaps, with a particular focus on issues of college readiness, transfer, and remediation. Yet California's current education data practices prevent the state from answering important questions, diagnosing problems in the educational pipeline, and developing policy solutions to improve students' progress through school and into the workforce.

Given new leadership in the state, the time is ripe to modernize California's education data. Establishing a statewide longitudinal data system would require a significant investment of resources, and decisions around governance, privacy, and security need to be carefully considered. But the returns for students, colleges, and the state as a whole could be substantial. California needs an integrated data system to ensure that the state's educational policies and programs are indeed improving outcomes for all students.

REFERENCES

- Bohn, Sarah, Shannon McConville, and Landon Gibson. 2016. *Career Technical Education in Health: An Overview of Student Success at California's Community Colleges*. Public Policy Institute of California.
- California Competes. 2018. *Out of the Dark: Bringing California's Education Data into the 21st Century*. California Competes.
- Conaway, Carrie, Keesler, Vanessa, and Nathaniel Schwartz. 2015. "What Research Do State Education Agencies Really Need? The Promise and Limitations of State Longitudinal Data System." *Educational Evaluation and Policy Analysis* 37: 16S-28S.
- Data Quality Campaign. 2018. "Roadmap for Cross-Agency Data Governance." Data Quality Campaign.
- Education Commission of the States. 2016. "State Longitudinal Data Systems - All State Profiles (October 2016)." Education Commission of the States.
- Gao, Niu, and Hans Johnson. 2017. *Improving College Pathways in California*. Public Policy Institute of California.
- Giani, Matthew, Celeste Alexander, and Pedro Reyes. 2014. "Exploring Variation in the Impact of Dual-Credit Coursework on Postsecondary Outcomes: A Quasi-Experimental Analysis of Texas Students." *The High School Journal* 97 (4): 200–218.
- Howell, Jessica S., Michal Kurlaender, and Eric Grodsky. 2010. "Postsecondary Preparation and Remediation: Examining the Effect of the Early Assessment Program at California State University." *Journal of Policy Analysis and Management* 29 (4): 726–48.
- Kurlaender, Michal, Scott Carrell, and Jacob Jackson. 2016. "The Promises and Pitfalls of Measuring Community College Quality." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2 (1): 174–90.
- Moore, Colleen, and Kathy Reeves Bracco. 2018a. "Scaling Goodwill: The Challenges of Implementing Robust Education Data Sharing Through Regional Partnerships." Policy Brief 3 in *California Education Policy, Student Data, and the Quest to Improve Student Progress*. Education Insights Center.
- Moore, Colleen, and Kathy Reeves Bracco. 2018b. "A Hunger for Information: California's Options to Meet Its Statewide Education Data Needs." Policy Brief 4 in *California Education Policy, Student Data, and the Quest to Improve Student Progress*. Education Insights Center.
- Moore, Colleen, Kathy Reeves Bracco, and Thad Nodine. 2017. "California's Maze of Student Information: Education Data Systems Leave Critical Questions Unanswered." Policy Brief 2 in *California Education Policy, Student Data, and the Quest to Improve Student Progress*. Education Insights Center.
- Moore, Colleen, Brock Grubb, and Camille Esch. 2016. "Gaps in Perspective: Who Should Be Responsible for Tracking Student Progress Across Education Institutions." Policy Brief 1 in *California Education Policy, Student Data, and the Quest to Improve Student Progress*. Education Insights Center.
- Phillips, Meredith, Sarah Reber, and Jesse Rothstein. 2018. "Making California Data More Useful for Educational Improvement." Policy Analysis for California Education.
- Reed, Sherrie, Patrick Lee, Michal Kurlaender, and Ambar Hernandez. 2018. "Intersegmental Partnerships and Data Sharing: Promising Practices From the Field." Continuous Improvement Brief. Policy Analysis for California Education.
- Warren, Paul, and Heather Hough. 2013. *Increasing the Usefulness of California's Education Data*. Public Policy Institute of California.

ACKNOWLEDGMENTS

Thanks to Bill and Melinda Gates Foundation and the Sutton Family Fund for supporting this work. We appreciate thoughtful reviews by Hans Johnson, Colleen Moore, Patrick Perry, and Paul Warren. We also thank Vicki Hsieh and Lynette Ubois for reviews and editorial support.

ABOUT THE AUTHORS

Jacob Jackson is a research fellow at the PPIC Higher Education Center. His research includes work on college costs, college readiness, community college participation, access to higher education, and college completion. Before joining PPIC, he was a postdoctoral scholar at the University of California, Davis, and a middle school and high school teacher in Sacramento, California, and Columbia, Missouri. He holds a PhD in education from the University of California, Davis, and a master's degree in education from the University of Missouri.

Kevin Cook is the associate center director of the PPIC Higher Education Center. His research interests include education finance and policy for both K–12 and higher education, and his current work focuses on performance-based funding, online learning, and public pensions. Before joining PPIC, he worked as a research assistant in the investments division of the California State Teachers' Retirement System, where he conducted sustainability risk analysis. He holds an MPPA from Sacramento State University and a BA in history from Occidental College.



PPIC

PUBLIC POLICY
INSTITUTE OF CALIFORNIA

© 2018 Public Policy Institute of California

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research. PPIC is a public charity. It does not take or support positions on any ballot measure or on any local, state, or federal legislation, nor does it endorse, support, or oppose any political parties or candidates for public office.

Short sections of text, not to exceed three paragraphs, may be quoted without written permission provided that full attribution is given to the source.

Research publications reflect the views of the authors and do not necessarily reflect the views of our funders or of the staff, officers, advisory councils, or board of directors of the Public Policy Institute of California.

PUBLIC POLICY INSTITUTE OF CALIFORNIA

500 Washington Street, Suite 600
San Francisco, CA 94111
T 415 291 4400 F 415 291 4401

PPIC SACRAMENTO CENTER

Senator Office Building
1121 L Street, Suite 801
Sacramento, CA 95814
T 916 440 1120 F 916 440 1121

ppic.org