Strengthening California's Teacher Information System

Camille E. Esch
Patrick M. Shields
Viki M. Young

The Center for the Future of Teaching and Learning

Research conducted by SRI International

Teaching and California's Future

Teaching and California's Future is sponsored by The Center for the Future of Teaching and Learning. The Center is made up of education professionals, scholars and public policy experts who care deeply about improving the schooling of California's children. The Center was founded in 1995 as a public, nonprofit organization dedicated to increasing the capacity of California's teacher workforce to provide a rigorous and balanced curriculum and ensure every child's continuing intellectual, ethical and social development. Margaret Gaston and Harvey Hunt, co-directors of The Center for the Future of Teaching and Learning, organized and directed the work.

Co-sponsors include: The California State University Institute for Education Reform; Policy Analysis for California Education; University of California, Office of the President; and WestEd.

Research was conducted and reported by SRI International, an independent, nonprofit corporation that performs a broad spectrum of problem-oriented research and consulting for government and industry.

Design by KSA-Plus Communications, Inc.

The Center for the Future of Teaching and Learning
133 Mission Street, Suite 220
Santa Cruz, CA 95060
www.cftl.org

Strengthening California's Teacher Information System

Camille E. Esch
Patrick M. Shields
Viki M. Young

The Center for the Future of Teaching and Learning

Research conducted by SRI International

Table of Contents

Preface 1
Executive Summary 2
A Greater Demand than Ever for Good Information 4
What Types of Data Are Needed To Inform Decisionmaking? 5
Why Can't Policymakers Get the Information They Need? 7
What Can Be Done To Improve California's Teacher Information System? 9
Special Considerations 11
Conclusion and Recommendations 14

Preface

This policy brief is based on the experience and insight gained by The Center for the Future of Teaching and Learning (CFTL) and SRI International (SRI) in their joint efforts to document the status of the teaching profession in California and related public policy issues. Over the last four years, these efforts have led to a series of reports on the status of the teaching profession in California. This work has been carried out with cooperation and guidance from a key group of co-sponsors that include The California State University Institute for Education Reform; Policy Analysis for California Education; University of California, Office of the President; and WestEd.

These efforts have provided critical information to policymakers and the general public on the status of the teaching profession in California. However, they also have revealed significant gaps in the teacher workforce data that are collected and reported at the state level. In our interactions with policymakers over the years, we frequently have been asked important questions about the teacher workforce that simply cannot be answered due to the inadequacy of state-level data. Despite extensive efforts to secure, link and analyze special data from key state agencies, few answers have emerged.

In our work with available state data, we have developed a thorough understanding of their shortcomings and what changes could be made to increase their usefulness. The intent of this policy brief is to call attention to simple, straightforward ways in which our current system of data collection can be improved to provide answers to policymakers' most pressing questions about the teacher workforce.

Executive Summary

California has enacted a set of initiatives designed to produce more qualified teachers and to draw them to the schools with the greatest needs. However, serious problems exist in the availability of information needed to plan and monitor these efforts. Specifically, policymakers need more reliable information in the areas of teacher attrition (teachers leaving the workforce before retirement), teacher workforce participation (job-taking), teacher movement between schools and districts, the "reserve pool" of teachers, trends in different credential routes, and the effect of state-sponsored programs for teachers.

While a great deal of data on teachers are collected by several different agencies — including the California Department of Education (CDE), the California Commission on Teacher Credentialing (CCTC), the California State Teachers' Retirement System (STRS), and universities that prepare teachers — these data cannot be used to answer many of policymakers' most important questions about the teacher workforce. This is due to two primary, related problems:

1. **Fragmented responsibility for collecting and reporting teacher data.** Because these agencies were established to perform specific, independent functions that are not linked by a common plan for data use, they act in isolation and make decisions that often don't allow their data to be used in analyses of the bigger picture.
2. **The lack of a commonly used unique teacher identifier to allow linkage across data systems.** Though other key agencies collect Social Security Numbers (SSNs) for use as a unique identifier, the most important source of teacher data in the state, CDE's California Basic Education Data System (CBEDS), does not. Without such an identifier, CBEDS data cannot be linked with other agencies' data and cannot be linked over time, making the entire CBEDS data collection effort far less useful than it could be.

These issues can be resolved if the various agencies adopt a unique identifier for use across all teacher record systems and a common plan for data collection, linkage and analysis. Other states that have taken these steps, including Connecticut, Florida, Georgia and Texas, have systems that allow policymakers access to far more powerful information than California has on teacher placement, retention, retirement trends and key shortage areas.

In developing a comprehensive data system, policymakers will need to consider several additional issues. First, a new system will need to include strong safeguards to keep any unique identifier out of the public domain and protect the identity of individual teachers. Second, procedures should be established to ensure that the data are used appropriately and made available for legitimate research efforts. Third, a formal mechanism for coordinating the data collection and analysis must be established. Finally, measures are needed to check and improve the accuracy of data that feed into the system.

Guided by the principle of building on current efforts — and based on years of experience in workforce research — we make the following recommendations:

1. **An independent organizational structure should be adopted at the state level to oversee the teacher data system and ensure accuracy, validity and appropriate access over time.** This entity — be it a coordinating group or a new independent agency — would develop a time line and common vision for the system and oversee implementation of the following recommended steps.
2. **A common identifier, such as teacher SSNs (or alternately, another unique teacher identifier) should be used by all relevant agencies to enable longitudinal analysis and linkage across datasets.** Specifically, if SSNs are chosen, CBEDS teacher-level records should add teacher SSNs to their records; CCTC should continue to collect teacher SSNs; and state-supported teacher programs, such as Beginning Teacher Support and Assessment (BTSA) and California Professional Development Institutes (CPDIs), should begin or continue to collect participant SSNs.
3. **CCTC, CBEDS and statewide teacher program individual records should be merged on a regular, timely basis.** A dataset including the elements listed

in this paper (Exhibit 1, page 9) should be compiled annually and made available for analysis by approved agencies.

4. **Analyses of the merged dataset and longitudinal CBEDS data should be performed annually on a specified time line and made available to policymakers and the public.** In concert with the legislative session, accurate, reliable data should be made available to the policy community as a basis for decisionmaking. Exhibit 2 on page 10 lists recommended analyses.

5. **Steps should be taken toward including teacher preparation programs in analyses of the teacher supply pipeline.** Teacher preparation programs' data systems should be analyzed to determine how collected data could be coordinated across programs and with data from other sources to ensure a complete picture of the state's teacher development system.

6. **Measures to ensure access to the data for legitimate research should be established.** Raw and aggregate data (stripped of any identifying information) should be made available publicly, and/or procedures for researchers to request special access should be established to facilitate analysis for research purposes.

7. **A regular system of accounting for data accuracy should be established to ensure that data and subsequent analyses are reliable.** Inaccuracies within data systems stymie analysis and may lead to misunderstanding and poor policy choices. Regular and timely checks of the data should be routine in any database used for decisionmaking purposes.

8. **Standards should be developed and used across all involved agencies to protect teacher privacy and ensure appropriate uses of the data system.** In particular, these standards should safeguard against theft or inappropriate use of unique teacher identifiers, such as SSNs.

If California can improve coordination of separate agency efforts and make modest technical changes to link key datasets, it can provide policymakers with the data they need to continue their efforts to strengthen California's teacher workforce.

A Greater Demand than Ever for Good Information

Since the late 1990s, California policymakers increasingly have grown aware of a number of serious challenges facing the teaching profession:

- a severe overall shortage of credentialed teachers;
- a persistently inequitable distribution of qualified teachers among the schools of the state, resulting in students at poor, inner-city schools being most likely to have underprepared teachers; and
- a variety of shortcomings in the provision of professional development to current teachers.

In response to these issues, California's Governor has proposed and the state Legislature has enacted a set of initiatives designed to bring more prospective teachers into the education pipeline and draw qualified teachers to the schools with the greatest needs. These are important steps in the right direction. However, serious problems with the availability of and access to information needed to plan and monitor the state's major reforms may hamper these efforts to ensure that every child has a fully qualified and effective teacher.

Existing data sources in California cannot provide some of the most basic information about the teacher workforce on a regular, ongoing basis. Specifically, policymakers report that they do not have access to data needed to make reliable projections of the magnitude of the teacher shortage in coming years. They also are in need of data to better understand complex conditions, such as the dynamics of the teacher labor market that result in the maldistribution of underprepared teachers, to be able to design appropriate policy to address pressing problems. They need data to help them identify which parts of the system and which types of schools or districts are most in need. Last, they need data to provide a baseline against which the impact of existing and new policies and programs can be measured. Without such data, policymakers never can be confident about the overall success of the state's efforts and cannot gauge the progress of individual programs. In addition, important problems, such as the maldistribution of underprepared teachers or an impending drop in the supply of teachers, may remain hidden with little chance of redress. In short, without robust and reliable data, the state risks continuing to invest money in efforts that are not effective and potentially missing opportunities to maximize the state's investment.

Given what we know about the severity of the teacher shortage and the new initiatives in place to address it, there is now a greater demand than ever for good information. A comprehensive data system, capable of illuminating the specific causes of the teacher shortage problem and its characteristics in different schools, districts and regions, is the urgently needed next step. In this policy brief, we argue that a great deal of good data are collected currently, but because of a lack of a coordinated, systemwide plan and a few key technical issues, these data cannot be used to answer policymakers' most important questions. We illustrate how the existing data system can be made more efficient and effective. In addition, we propose how high-quality data required to answer relevant questions from policymakers and the public can be made available in a way that protects individual privacy. Our intent is to call attention to the critical need for a better and more reliable information system for the teacher workforce and set in motion efforts to address the need.

What Types of Data Are Needed To Inform Decisionmaking?

In recent years, education policymakers increasingly have focused on questions about the teacher workforce, such as:

- How do we attract teachers to hard-to-staff schools?
- How do we encourage teachers to stay in hard-to-staff schools?
- How do we get more teachers into and through the teacher preparation system?
- On what parts of the system (e.g., recruitment, job placement, retention in the first few years) should we focus resources?
- On which schools and districts should we focus resources?

These are broad policy questions that require reliable, current, statewide data and sound analysis in the following areas:

- **Workforce participation.** To monitor and project the supply of teachers and to better track the effects of recruitment and preparation efforts, data are needed to indicate how many newly credentialed teachers take teaching jobs, where they take these jobs and what their classroom assignments are (including "out-of-field" teaching). Also important is information regarding any variation in job-taking by preparation program, credential route or recruiting efforts.
- **Movement between schools and districts.** To monitor and predict teacher supply and demand at local levels, data are needed on the extent to which teachers move between schools or districts over the course of their careers. Also, data are needed on which types of schools or districts teachers tend to move away from or toward.
- **Teacher attrition.** To monitor and project the demand for teachers and to better track the effects of investments in recruitment efforts, data are needed to estimate how many teachers leave their particular school or district each year and how many new teachers leave the teacher workforce each year. To understand what factors contribute to or prevent teacher attrition, data are needed to reveal how attrition rates differ by variables such as the demographics or location of the school, the type of teaching assignment or teaching credential, and whether the individual has participated in the Beginning Teacher Support and Assessment (BTSA) program, an internship program or a preinternship program.
- **"Reserve pool" of teachers.** To better project teacher supply and to identify an untapped group for recruitment, data are needed on the individuals who are prepared and credentialed to teach but are working elsewhere. Data also are needed on how many such individuals exist, when they last taught and how many eventually re-enter the teacher workforce.
- **Trends in different credential and preparation routes.** Over the past 10 years, a number of alternative routes to the teaching profession have emerged, including intern and blended programs. In addition, the emergency permit has become, for many, the first step in becoming a teacher. To understand the effects of these different routes into the profession, policymakers need to monitor the progress of participants and determine how many successfully complete their preparation and enter and stay in the teaching profession. Also needed are data on how long it takes individuals to complete their preparation and on the relationship, if any, between teachers' routes into the profession and where they are assigned or choose to take jobs.
- **Program participation and impacts.** In addition to alternative certification routes, state policymakers have initiated numerous programs to strengthen the teacher workforce, including efforts to recruit more teachers into the profession, provide support for them in their first years of teaching to stem potential attrition, and assist them in developing new skills and strategies. Better data are needed on which teachers participate in these programs, what types of schools and districts they teach in, and whether these programs are effective in attaining such goals as retaining teachers at their schools or in the teaching profession.

All data listed here are important, though it should be noted that not all of the above analyses need to be addressed for every teacher every year. Some issues lend themselves to special one-time research projects, while others are addressed best by collecting data annually. The recommendations proposed in this brief would make possible the acquisition of the above data and also would facilitate primary research by a variety of organizations and institutions that seek to investigate issues related to the teaching profession and produce findings to strengthen policymaking. For example, primary data collection efforts (such as surveys of teachers; credential candidates; and the reserve pool of credentialed, nonteaching individuals) could address important questions, such as:

- Why do individuals choose particular preparation routes?
- Why do teachers take jobs in particular schools or districts?
- What might influence them to take jobs in high-need schools or districts?
- Why do teachers leave their school, their district or the teaching profession altogether?
- What, if any, incentives could prevent teacher attrition?
- Why do some who are credentialed choose not to teach?
- What, if any, incentives could draw credentialed, nonteaching individuals back into the profession?

Why Can't Policymakers Get the Information They Need?

Currently, there is no state-level data and analysis system to comprehensively address policymakers' most basic questions. But a substantial amount of data on the teacher workforce does exist across several different agencies and institutions. The California Department of Education (CDE), the California Commission on Teacher Credentialing (CCTC), the California State Teachers' Retirement System (STRS) and every university that prepares teachers collect information about the teacher workforce. Why then, despite the significant time and money spent on these data collection efforts, do policymakers still not have the kind of information needed for sound decisionmaking? There are two primary, related problems that hamper the state's current efforts:

1. Fragmented responsibility for collecting and reporting teacher data; and
2. The lack of a commonly used unique teacher identifier across data systems.

Both of these problems stem from the absence of a systemwide perspective that guides data collection and reporting efforts across the different agencies. These shortcomings are not the result of oversight but of an agency-specific, single-function vision of why the data are collected and how they should be used.

Fragmented Responsibility for Collecting and Reporting Teacher Data

While the state collects a great deal of data, no one agency, group or individual is charged with taking a systemwide perspective to ensure that these data are used to answer policymakers' critical questions. Instead, multiple agencies gather and hold various pieces of information. The databases within these agencies are very consistent with their basic missions, such as credentialing teachers or distributing retirement benefits, but they are far less useful when it comes to addressing the overarching issues of teacher supply, demand and distribution. For example:

- CCTC collects information on the credentials teachers hold and which university recommends their credentials, but that database stops short of being able to identify who actually goes on to teach.
- STRS has data on when individuals begin contributing to or drawing from the teacher retirement fund but cannot easily analyze if and when teachers leave the profession before retirement.
- CDE's California Basic Education Data System (CBEDS) collects information on what teachers teach, which schools they teach in and basic demographic information on teachers, but CDE has been stymied by the complications inherent in building capacity for longitudinal analysis. Consequently, the usefulness of CBEDS as an analytic tool is limited.
- Each teacher preparation program in the state collects data on prospective teachers, but there is no mechanism for aggregating these multiple data sources across the state.

Because these agencies are not linked by a common plan for data use, they often act in isolation, making it difficult for their data to be used in concert with those of other organizations. For example, our experience working with STRS data revealed that these data are unusable for analyses of teacher attrition (teachers leaving the workforce before retirement) because data are not collected and organized for this purpose. We found that isolating the necessary data elements to determine whether and when teachers stop teaching (prior to retiring) requires additional computer capacity and programming time, the costs of which are not included in agency budgets.

In addition, we found that individual STRS contributions are an inadequate proxy for employment as a teacher. STRS data do not distinguish between practicing teachers and other nonclassroom personnel, K–12 instructors and community college instructors, or full-time and part-time employees. This makes it impossible to isolate and analyze specific groups, such as K–12 classroom teachers. Because the data collection efforts are not driven by key policy questions, the data collected are not specific enough to answer such questions.

Also, when data are collected by different agencies that do not share a common purpose, there are barriers to linking these data. Attempts to link data from different agencies have revealed that there is no commonly used unique teacher identifier across all data systems.

Lack of a Commonly Used Unique Teacher Identifier Across Data Systems

This lack of a commonly used unique teacher identifier, more than any other problem, renders California's data collection efforts inefficient and ineffective. As a state, we currently collect virtually all of the information needed to perform an array of critical analyses, but without a commonly used unique identifier in all relevant databases, the data cannot be used to answer important policy questions.

Unique identifiers are commonly collected from adults in our society. For example, agencies such as CCTC, STRS and some university databases include Social Security Numbers (SSNs). However, the CBEDS system, run by CDE, does not consistently collect and keep records of teacher SSNs. Instead, each teacher record has a locally assigned identification number — in some cases the teacher's SSN, in other cases, a locally generated number. This creates two significant barriers to analyzing the extensive individual-level data they do collect. First, CBEDS data cannot be linked with other agencies' data to address policy-relevant questions. For example, CCTC credential data and CBEDS teacher-level data cannot be integrated to determine how many credential holders take jobs, the types of schools in which they take jobs and the types of schools they tend to leave. However, these analyses are crucial to unraveling issues associated with the maldistribution of qualified teachers.

Second, because the teacher identifiers are generated locally and often are not consistent from year to year, the teacher-level data collected by CBEDS for many years cannot be linked over time. In other words, data collected on an individual in 2001 cannot be linked to data collected for the same individual in 2002. Because of this shortcoming, the entire CBEDS data collection effort is far less useful than it could be.

So, despite collecting extensive information on individual teachers (including demographic data, years of experience, credentials held, subjects taught), CBEDS is virtually useless for analyzing what happens to teachers over time. Thus, while CBEDS can be used to count the number of teachers *in* the workforce each year, it cannot reveal how many *leave* the teacher workforce each year. This number, though critical to planning and monitoring investments in recruiting and retaining teachers, *is not knowable in California.* Further, we have no way to track patterns of teachers switching schools or districts over the course of their careers or re-entering the workforce after having left for a period of time. Last, there is no way to identify and survey those who have left teaching or re-entered the workforce, which would improve our understanding of their behavior.

Lessons from Other States

In other states, such as Connecticut, Florida, Georgia and Texas, policymakers have resolved the problems of data linkage and longitudinal analysis by using the same unique identifier in all relevant databases. In some states, credentialing data and data on the teacher workforce are linked easily because the state department of education is the credentialing agency and the data are housed in the same system. California, because it has a credentialing agency that is independent of the state department of education, must take special measures to overcome the data management problems created by this organizational structure. These states also have the advantage of being able to analyze teacher data longitudinally. This allows access to far more powerful information than California's policymakers have on teacher retention, retirement trends and key shortage areas.

In Connecticut, for example, the state has an accurate measure of teacher attrition and can analyze how attrition varies by subject area, school and age of the teacher. This is done by using a model that includes data for all participants in the preparation system and currently teaching in the schools and takes into consideration part-time and full-time hiring patterns by assignment; inter-assignment migration, as well as interdistrict migration; and enrollment growth by elementary, middle and high schools. This is useful because it allows the state to make specific, detailed projections of the number of new teachers needed in future years in different regions and subject areas. Connecticut's policymakers rely on the system's ability to analyze teacher data longitudinally, something that California's policy community cannot do.

In Florida, the state collects and analyzes longitudinal data on the number of vacant positions, positions filled by teachers without the appropriate disciplinary background, and the projected supply of teachers from out of state and candidates graduating from state preparation programs. Texas and Georgia collect similar data that can be used to track and project how many credentialed individuals take jobs, how many teach "out of field" and how many leave teaching.

What Can Be Done To Improve California's Teacher Information System?

California needs a teacher data system capable of producing the analyses needed to answer policymakers' questions about the teacher workforce. An effective teacher data system can be accomplished without the development of a substantial new infrastructure; existing data collected by the different agencies can be used if driven by a common, well-defined plan.

Adoption of a Common Identifier

Several steps are required to improve the usefulness of California's teacher data system. First, a common teacher identifier such as an SSN must be adopted by all appropriate agencies to enable longitudinal analysis and linkage among datasets. Specifically, if SSNs are selected as the common identifier, CBEDS teacher-level records would then begin collecting teacher SSNs; CCTC would continue to collect teacher SSNs; and state-supported teacher programs, such as Beginning Teacher Support and Assessment (BTSA) and the California Professional Development Institutes (CPDIs), also would collect participant SSNs so that important program data could be included in the statewide data system.

Data Linkage

Once a common identifier is adopted, individual records can be linked across agencies. As a first step, data from CCTC, CBEDS and statewide teacher programs should be merged on a regular, timely basis. Specifically, the data elements listed in Exhibit 1 should be merged, making many critical analyses possible, either directly or by facilitating original data collection efforts, such as surveys.*

Next, steps should be taken to investigate how best to include teacher preparation programs in analyses of the teacher supply pipeline. Data from institutions of higher education are needed to answer questions such as, "How many individuals who begin teacher preparation programs actually complete them, and how many then enter the workforce?" Teacher preparation programs' data systems should be analyzed to determine what data are collected currently and how they could be coordinated across programs and with data from other sources. Data now being compiled by the institutions of higher education under new federal requirements may contribute to increasing the availability of teacher preparation information.

Exhibit 1
Recommended Data Elements for a Linked Dataset

Source of Data	Data for Linked Dataset
CCTC	• Teacher credential history, including number and dates of issued credential(s), preparation program
CBEDS	• Teacher demographics • Teacher assignment history, including grade and/or subject area • Teacher employment status history (whether full time or part time) • Teacher school assignment (which school and district)
Statewide program data from CCTC or CDE	• Teacher program participation history (whether and when in BTSA, internship program, preinternship program, professional development program, etc.)

**Note: STRS data do not need to be merged into the linked dataset. Instead, they may be used in isolated analyses of teacher retirement trends.*

Data Analysis

Once linked, these data should be analyzed to address key policy questions in the areas of teacher attrition and retirement, workforce participation, movement between schools/districts, the "reserve pool" of teachers, and trends in different credential routes. Some analyses require data from multiple agencies, whereas others require data from multiple years. Exhibit 2 lists recommended analyses to be performed annually on a specified time line and made available to policymakers and the public.

Exhibit 2
Recommended Analyses of Linked Dataset and Longitudinal Data

Analysis	Description	Source
Workforce participation	• How many newly credentialed teachers take teaching jobs and where they take them. Disaggregation by preparation program and credential route.	Linked dataset
Movement between schools and districts	• How many teachers move between schools or districts each year and over the course of their careers. Disaggregation by type of school/district and years of teaching experience.	Longitudinal CBEDS data
Teacher attrition and retirement	• How many teachers leave their school/district each year and how many teachers leave the workforce each year. Disaggregation by demographic or location of the school; type of teaching assignment; teaching credential; years of teaching experience; and whether the individual has participated in BTSA, an internship program or a preinternship program.	Linked dataset and longitudinal CBEDS data
"Reserve pool" of teachers	• How many former teachers hold valid credentials but no longer teach, how many former teachers re-enter the profession each year and the average length of time they are out of the profession.	Linked dataset and longitudinal CBEDS data
Trends in different credential routes	• How many emergency permit holders and intern certificate holders convert to regular preliminary credentials and the average length of time to convert.	Longitudinal CCTC data
Program participation and impacts	• Which teachers are participating in state-funded programs to support teachers. Disaggregation by teacher characteristic (e.g., years of teaching experience) and school/district characteristic. What the impacts of participation are on teacher retention and other program goals.	Linked dataset and longitudinal CBEDS data

Special Considerations

While there is mounting frustration with the inability to secure the necessary information to ensure informed decisionmaking, there is an equal desire to ensure that any new state-supported data system is accurate, economical and aligned with the purpose intended. To this end, considerations most often noted are those regarding the protection of individual privacy; how to maintain appropriate access to and uses of the data system; how to best lead and organize data collection, merging, analysis and reporting activities; and how to enhance the overall quality of data feeding into the system. Here we briefly discuss each of these considerations and various options for resolving them.

Protecting Teachers' Privacy

While SSNs commonly are collected from adults in our society, some concern about using them remains because of the potential threat their use poses to teachers' privacy. Teachers may fear that SSNs may be used inappropriately to gather personal information or will make them vulnerable to identity theft. These are valid concerns that point to the need for strong safeguards to keep SSNs, or any unique identifiers, out of the public domain and protect the identity of the individual. It is important to note that unique identifiers such as SSNs are needed only to link data files; in and of themselves, they do not contain information needed for the analyses described here. Therefore, the most important aspect of any system that includes unique identifiers such as SSNs is that they be available only to data analysts or managers with clearance to use them to link data and that they be removed from any files made available to anyone else.

One approach to eliminating this concern is to use unique identifiers such as SSNs only to link files and then strip them out of the database altogether. Another option is to scramble SSNs or match them with another unique identifying number for use in public versions of the data, while retaining the match between real and scrambled SSNs or other identifying numbers in a protected file that is not made public.

Additional special measures can be taken to safeguard data. Departments of education in other states use teacher SSNs as unique identifiers and have developed secure processes with appropriate safeguards in place to ensure that their use is not abused. Consequently, data analysts and managers in Connecticut, Florida and Texas report never having had any controversy surrounding the collection of teacher SSNs. In one example, a Connecticut Department of Education representative stresses how seriously the state takes the responsibility of protecting individual teachers' identities: "We have very strict confidentiality practices for transfer and dissemination of data. State auditors monitor publicly available data. SSNs are available only to people who have authority [to work with them] and have been granted access through passwords and special procedures." This responsibility extends to contracted work outside the department as well. "When sending data to a contractor, we use sophisticated Web-based encryption. We use a highly reputable contractor who has lots of experience protecting confidential information," the representative says.

Why choose Social Security Numbers?

Because all teachers have SSNs and many current and historical databases already use them, they would be the most practical unique identifier. A possible alternative strategy for securing unique identifiers is to begin assigning teachers unique identification numbers when they receive their credentials. However, this option is less desirable because it would prevent the use of historical credential data, causing an information lag of many years before the credential histories of current teachers could be analyzed. It also would prevent analysis of patterns in teacher preparation programs where candidates have not yet received a unique teacher identifier. Additionally, using an identifier other than SSNs would necessitate additional and costly efforts for all agencies involved, including the tasks of generating and keeping track of an additional set of numbers. In Connecticut, the assignment of new unique identifiers was attempted but ultimately abandoned because there were so many errors during data entry. Because there were no "source data," the identifiers could never be checked against other data files or reliably remembered by individuals.

Access to and Appropriate Use of Data

Related to privacy considerations are the issues of access to and appropriate uses of the data system. To maximize the system's usefulness, a minimum set of analyses should be performed on a regular basis at the state level and made public. For example, the state could publish a report describing the number of teachers who entered or left the workforce in the previous academic year. Then, both the raw and aggregated data (after being stripped of any identifying information) could be made available. This type of access would follow existing models in other states, as well as those promoted by federal-level agencies exploring similar research and policy questions, such as the National Center for Education Statistics. Alternately or additionally, procedures should be established for organizations to request access to the raw data for legitimate research efforts.

This information system should be used only to provide teacher workforce information to policymakers. Data from the system should never be used for purposes other than valid research or evaluation. Moreover, the system should never be used to identify individual teachers or groups of teachers, and at no time should this database be linked to other databases not related to education (e.g., databases containing legal, financial or medical records).

Leadership and Organization

Implementing needed changes to the teacher data system will require coordination across several agencies. To accomplish this, a formal mechanism for coordinating the data collection and analysis needs to be established, raising the question of leadership and administration for the system.

A limited and informal approach to this issue would involve a legislative and/or budgetary directive to the various agencies involved to form a coordinating group or council to develop the data system and oversee its operation. Such an approach would depend, in the final analysis, on the willingness and enthusiasm of each of these independent agencies to work together and their capacity to supply data to and staff the coordinating function. Further, decisions would have to be made regarding the ways in which pragmatic functions, such as developing and distributing products, would be divided (for example, this function could rotate every five to seven years or permanently be assigned to a participating agency), all contingent upon budget allocations.

A second option includes the formal expression of administrative and legislative desire to create a coordinated information system, coupled with a directive, oversight authority and budgetary support to an existing agency or organization to bring the various agencies involved together (perhaps through the development of a memorandum of understanding or other written agreement) to develop and implement the system. Among the benefits of this option is the fact that the designated lead agency is likely to have some data system in place that could be expanded to accommodate the information from other participating organizations and staff familiar with the database functions already on hand and knowledgeable about the operations necessary to merge all data systems. However, there are concerns inherent in this option, including the risk that the desire for a collaborative, independent effort could fall prey to the day-to-day realities of the host agency's primary function or that responsibilities for the development of a product would remain while budgetary support falls away.

A third option is the establishment of an independent entity to set up the database system and oversee its operation. An independent entity with a legislative mandate to establish a data system would underscore the priority policymakers placed on the effort. This entity could operate in much the same way as other independent oversight groups with its own board drawn from representatives from state agencies involved in database coordination; a delegate each appointed by the Governor, the Senate Rules Committee, the Speaker of the Assembly, the California State University System and the University of California; and field representatives, including the public, classroom teachers, principals, superintendents and members of the research community. This option would have the advantages of a formal structure and mandate while directly involving consumers and the diverse agencies in the governance of the system. And while this agency is designed to feature independent analysis of data collected and unbiased reporting of information, this option could involve duplication and unnecessary additional expense if parameters for growth and development were not put into place.

Quality of Data

Finally, the usefulness of the entire enterprise ultimately rests on the quality of the data — that is, the reported information must reflect accurately the current status of the teacher workforce. Given our experience working with data gathered from various sources, we believe that new procedures will have to be put in place to check the accuracy of data. As more options arise that allow schools and districts to enter data electronically, such checks can be built into the appropriate software so, for example, a school could not report conflicting credential information. In the short term, regular reviews and troubleshooting of data collection activities are needed. Additionally, enhanced communications and technical assistance to local administrators may be required to ensure high data quality.

Conclusion and Recommendations

Thorough and reliable data on California's teacher workforce are critical to making sound decisions about addressing California's shortage of qualified teachers. Many key questions about the teacher workforce cannot be answered using current data systems because of the combined effects of lack of coordination and poor data linkage. This flawed system leaves policymakers without even basic information about the dynamics of the teacher workforce. California needs improved coordination of separate agency efforts coupled with modest technical changes to link these datasets through simple straightforward actions and leadership. The fuller data that will emerge carry with them the promise of better-informed policy decisions needed to strengthen California's teacher workforce and aid in the learning of the students it serves.

In summary, we make the following recommendations to integrate the diverse sources of teacher data into a comprehensive system:

1. **An independent organizational structure should be adopted at the state level to oversee the teacher data system and ensure accuracy, validity and appropriate access over time.** This entity — be it a coordinating group or a new independent agency — would develop a time line and common vision for the system and oversee implementation of the following recommended steps.
2. **A common identifier, such as teacher SSNs (or alternately, another unique teacher identifier) should be used by all relevant agencies to enable longitudinal analysis and linkage across datasets.** Specifically, if SSNs are chosen, CBEDS teacher-level records should add teacher SSNs to their records; CCTC should continue to collect teacher SSNs; and state-supported teacher programs, such as BTSA and CPDIs, should begin or continue to collect participant SSNs.
3. **CCTC, CBEDS and statewide teacher program individual records should be merged on a regular, timely basis.** A dataset including the elements listed in this paper (Exhibit 1, page 9) should be compiled annually and made available for analysis by approved agencies.
4. **Analyses of the merged dataset and longitudinal CBEDS data should be performed annually on a specified time line and made available to policymakers and the public.** In concert with the legislative session, accurate, reliable data should be made available to the policy community as a basis for decisionmaking. Exhibit 2 on page 10 lists recommended analyses.
5. **Steps should be taken toward including teacher preparation programs in analyses of the teacher supply pipeline.** Teacher preparation programs' data systems should be analyzed to determine how collected data could be coordinated across programs and with data from other sources to ensure a complete picture of the state's teacher development system.
6. **Measures to ensure access to the data for legitimate research should be established.** Raw and aggregate data (stripped of any identifying information) should be made available publicly, and/or procedures for researchers to request special access should be established to facilitate analysis for research purposes.
7. **A regular system of accounting for data accuracy should be established to ensure that data and subsequent analyses are reliable.** Inaccuracies within data systems stymie analysis and may lead to misunderstanding and poor policy choices. Regular and timely checks of the data should be routine in any database used for decisionmaking purposes.
8. **Standards should be developed and used across all involved agencies to protect teacher privacy and ensure appropriate uses of the data system.** In particular, these standards should safeguard against theft or inappropriate use of unique teacher identifiers, such as SSNs.

THE CENTER
FOR THE FUTURE OF TEACHING & LEARNING

133 Mission Street, Suite 220
Santa Cruz, CA 95060
831-427-3628
www.cftl.org