Improving California's Student Data Systems To Address The Dropout Crisis

California Dropout Research Project Report #10
May 2008

By

Georges Vernez
RAND

South Hall, Room 4722
University of California
Santa Barbara, CA 93106-3220

www.lmri.ucsb.edu/dropouts

Phone: 805-893-2683
Fax: 805-893-8673
Email: dropouts@lmri.ucsb.edu

ABSTRACT

The extent of California's dropout problem is unknown, placed at between 33 and 16 percent. California needs to build a robust student data system—called a student unit record (SUR) system because it contains information on every student from entry in kindergarten to exit from college and eventually into the labor force—in order to accurately measure the dropout rate across schools and colleges, hold educational institutions accountable, evaluate the effectiveness of programs, help identify students at-risk, and assess return on educational investments. The state can do so by building on its existing student data systems enhanced by adding and standardizing data elements (e.g. program participation and attendance), integrating the existing SUR from the four California education segments, linking the integrated SUR data file with other state and federal data files, and broadening access to the resulting data set

INTRODUCTION

The extent of California's dropout problem is yet unknown. In 2005-06, the state Department of Education reported two widely different estimates of the rate at which California students graduate from high school: 83.2 and 67.1 percent.[1] Others have placed the state's high school graduation rate at somewhere between 71 and 74 percent.[2] At the postsecondary level, the 5-year graduation rate at California state universities is placed at below 50 percent, although this rate may be too low because it does not account for students who may have completed their degrees on a campus other than the one they started with.[3] The accurate measurement of the high school and college graduation rates--and by implication, dropout rates--is not an issue unique to California. At the national level, for instance, the U.S. high school graduation rate has been estimated to be anywhere from 66 to 88 percent depending on sources, definitions, and methods used. The range of estimated minority graduation is even grater—from 50 to 85 percent.[4]

To address this issue, improve capabilities to assess the status of education and improve the quality of education, states are increasingly building robust student data systems. These systems are commonly termed "student unit record (SUR)" systems because they contain an individual electronic record corresponding to each student enrolled in an educational institution. SUR data systems permit the tracking of an individual student's progress over time—from entry into kindergarten or even-school to exit from college, and eventually into the labor market. These data are vital to measure graduation rates accurately and to formulate policy and answer questions at the core of studies of educational effectiveness.

Currently, 22 states have the ability to match individual student records between their K-12 and postsecondary education systems. Several other states are developing this capability.[5] California, however, lags behind other states. Despite the California legislature's clear intent that a longitudinal student data management system capable of tracking students

---

[1] http://cde.ca.gov/dataquest (As of November 13, 2007).
[2] Rumberger (2007a) . See Appendix A for a discussion of how these rates are computed and their limitations.
[3] http://www.calstate.edu/as/
[4] Heckman and Lafontaine (2007) and Mishel and Roy (2006).
[5] Data Quality Campaign (2008)

from elementary school through college be developed and maintained,[6] the state has only recently committed to developing a K-12 California Longitudinal Pupil Achievement Data System (CALPADS), and at this time it is not planning to link CALPADS with the existing state's postsecondary student data systems.[7]

A K-20 SUR system would enable the state, education administrators, faculty, and the public to address questions that currently cannot be addressed with current data systems—questions such as how California's students are progressing over time; who are the students that dropout and how can they be helped; how to improve course articulation between high school and college; how effective are programs to retain, prepare and help students succeed in high school and college; what classes of students may need special intervention and attention; and how well are students prepared to meet California's future labor demands.

In this paper, we address three questions:

- What are the main purposes SUR data systems serve and what student information is required to address them?

- What are California's current student data system's capabilities to address these purposes?

- What enhancements to the existing state student data systems are desirable and what are the challenges that need to be addressed in order to implement them?

The remaining of the paper addresses these questions in turn and ends with a concluding section.

PURPOSES OF SUR DATA SYSTEMS AND INFORMATION REQUIREMENTS

A K-20 SUR serves four main purposes:

---

[6] For instance SB 257 (Section 52052.5, 2003) states that the legislature's intent if to "promote good data management practices with respect to pupil data systems and issues including…linking pupil data with data from other agencies and users, including a mechanism to monitor pupil progress in postsecondary education."

[7] Hansen (2007). The state has three quasi-independent postsecondary education systems: Community Colleges; California State University and the University of California.

- To hold educational institution accountable

- To evaluate the effectiveness of education programs

- To help identify students at risk and design programmatic interventions

- To assess return on education investments.

Accountability

Information on graduation rates, dropout rates, and student achievement are most frequently used for accountability purposes. In the case of graduation and dropout rates, different measures may be used depending on different definitions of who is a graduate or a dropout, the time perspective taken, and the institution being held accountable.

The *on-time freshman class* high school graduation rate has been recommended by the National Governor's Association (NGA) Task Force on State High School Graduation Data to be used as a standardized measure by all states. It is calculated:

> "…by dividing the number of on-time graduates in a given year by the number of *first-time* entering ninth graders four years earlier. Graduates are those receiving a high school diploma. The denominator can be adjusted for transfers in and out of the system. Special education students and recent immigrants can be assigned to different cohorts to allow them more time to graduate.[8]"

This definition does not consider recipients of a GED or of an alternative certificate as a high school graduate and, thus, is particularly relevant for use in individual school, district and college accountability.

For a more complete understanding of the graduation problem, a longer period of time to graduation should also be used to account for the frequency with which 9th grade is repeated by students and other students who may take longer to graduate, such as a *five-or six-year* instead of a four-year time period. This rate is typically referred to as a "*completion rate.*" A variance to this high school "*completion rate*" is to also count as high school graduates recipients of a GED or alternative certificate.

The on-time graduation and the completion rates may be computed at the school, district, college, or university levels as well as at the district, education segments—K-12,

---

[8] National Governors' Association (no date).

community college, California State University, University of California—and state levels. Because of the high mobility of students, it may be "unfair" to hold a school, district or college, or even the state accountable for students who leave and complete their education in another school or college, district, or state, respectively. For instance, about one in every five college students attending a college in one state will attend a university in another state before graduating (Edelman, 2007). In this event, students who have transferred out may be excluded from the computation of the school or college, district, or state computation of the on-time graduation, but not necessarily from the completion graduation rates. Separate on-time graduation rates may also be computed for students that enroll in a school, college, district, or the state after the freshman year, and, hence, have not done all of their high school or college year in the same educational institution.[9]

Similarly, there is agreement on two different definitions of the dropout rate. A strict definition of a dropout is anyone who has not re-enrolled in another public or private school or college pursuing a high school or college degree and who has not graduated from high school or college.[10] As with the graduation rate, a broader definition would consider that a student that pursues a GED or alternative degree or attending a college offering a baccalaureate or associate degree is not a dropout.[11] Either using the strict or the broader definition of dropout, a dropout rate can be computed for each year (or any other period of time) using the class entering at the beginning of the year as the denominator and for specific cohorts of students provided that the information about where the student is and what she/he is doing is collected accurately and comprehensively.

For accountability purposes, these various measures are often also computed separately for specific subgroups of students at each institutional level, such as by race-ethnicity, gender, economic characteristics, immigration status and for special education students (IEP) and students with low English proficiency (LEP). All of these measures

---

[9] There is some evidence that schools may "push out" or transfer low-achieving students to alternative schools (Rotermund, 2007a) in order to enhance their measured performance. Computing separate on-time graduation and completion rates for students who do all of their schooling in the same school and for students who transferred out from that schools and transferred in from that school would allow to identify schools that engage in that practice.

[10] After accounting for students who have died, are temporarily out of school for illness, suspension, or other reasons.

[11] In the current California definition of dropouts these students are considered not a dropout. See http://dq.cde.ca.gov/dataquest/gls-drpcriteria.asp (As of November 30, 2007).

contribute to providing a better understanding of education outcomes for students and how well the different institutions in the education system are serving them.

Three types of data are required for each student that enters the state education system at any level in order to compute this broad range of graduation and dropout measures: socio-demographic characteristics and where and what is the individual doing in order to determine the educational status of student (see Table B.1 in Appendix B). Relevant socio-demographics characteristics include all those mentioned in the preceding paragraph. Data requirements for where and what is an individual doing at any given point in time include whether the student is enrolled in public or private school, college or other educational institutions or program, the grade and institution enrolled in, the location of the institution, as well as whether she/he has obtained a high school and college degrees, GED or other alternative certificate.

Education Program Evaluation

Maximizing the graduation rate and/or minimizing the dropout rate of high school and college students are among the key desired outcomes (a major other outcome is student achievement) of state policies such as the Quality Education Investment Act (QEIA) and programs such as class size reduction, dropout prevention, charter school, college preparation, remedial, tutoring, and after-school programs. Currently, the state, and for that matter, district superintendents and principals have currently no means to systematically evaluate the effectiveness of these programs and of district and school instructional or other initiatives that they may have undertaken.[12] Evaluating policies and programs is expensive because it requires collecting information often not available otherwise via surveys and interviews of educators, parents, and students and classroom observations and the legislature has been reluctant to provide the funding needed. The result is that few California education policies and programs are ever assessed for their effectiveness.

SUR data systems that contain longitudinal individual student information are particularly well suited to evaluate ongoing and new education policies and programs. No new data would need to be collected each time a new policy or program is to be evaluated. Such a data system lends itself for use in rigorous evaluation using quasi-experimental

---

[12] California Policy Convening (2007).

techniques such as "before and after," "differences-in-differences" and regression discontinuity techniques. Most importantly, it also can support experimental designs even though schools and students may be assigned randomly to treatment and control groups as long as the schools and students participating in the experiment can be identified.

In order to determine whether policies and programs have the effects desired, additional information to that listed in the preceding section is required for each individual student including every program a student participates in, length of participation in each of the program, courses taken, grades obtained in each course taken, college (SAT, ACT), state, and other test scores (see Table.B.1 in Appendix B).

Identifying Students at Risk and Program Interventions

Understanding why students do not graduate and/or why they dropout from school or college can help educators design and target available interventions to those students that are most at risk. Research suggests that reasons for not graduating or dropping out from high school fall into three interrelated categories:[13] (1) Student-related reasons include course failures, low attendance, getting poor grades (most particularly in mathematics and English), being a limited-English-proficient student, being a special education student, and/or a student that exhibits certain behavioral traits such as requiring close supervision, having poor social interaction skills, being difficult-to-teach, and having low motivation as assessed by teachers; (2) Family-related reasons include being from a single parent household, having parents who did not graduate from high school, having an older sibling who dropped out of school, coming from a low income family, and/or spending three to four hours home alone after school; (3) school-related reasons include teachers' lack of shared commitment to improve the school for all students, low coherence of instructional program (both as reported by teachers), students' low level of trust for their teachers, and in schools where teachers' personal support for individual student, personalization in the classroom and press on students towards academic achievement are low (as assessed by students). Schools with these attributes have been found to have lower attendance and higher course failures than schools without these attributes.[14]

---

[13] See Rotermund (2007b); Rumberger (2007b); Neild and Balfanz (2006); and Larson and Rumberger (1995).
[14] See Allensworth and Easton (2007).

Some of the information-- including courses taken, grades, IEP and LEP status, and low-income family status--needed to identify students possessing one or more of the characteristics listed above was already identified as needed for computing graduation and dropout rates or to evaluate the effectiveness of programs. Additional student level information that would need to be collected to identify at-risk students include more detailed information on student's family characteristics and student traits as assessed by teachers and selected measures of school climate as assessed by student and teachers as listed above (see Table B.1 in Appendix B).

Assessing Workforce Preparation and Return on Education Investments

Spending more than $60 billion annually on education, California has an interest to ask whether its students are adequately prepared to meet the state's workforce requirements now and in the future and to be full participants in the state's political and social life. It also has an interest in knowing about the costs of a failure to adequately prepare students for adult life. Doing so requires data on individual activities that go beyond the students' high school and college years. The information required for this purpose includes the major students graduated in, the type of employment obtained after their graduation and beyond, their earnings, and the demand they place on social and other state programs including welfare, Medicaid, foster care, child and adult welfare, corrections, and health services (see Table B.1 in Appendix B).

Much of the information identified above as needed for educational accountability, policy and program evaluation, identification of students at risk and assessment of the ultimate outcomes of improving education for the economy and the use of social, health and other services is currently being collected and stored, albeit by a large number of different institutions with little or no current connection between them. We now turn to taking stock of the information actually contained in California's multiple existing student data systems and assess their potential use for the purposes described above.

CAPABILITIES OF CALIFORNIA'S EXISTING STUDENT DATA SYSTEMS

All four of the California education segments—K-12, Community Colleges, California State University, and University of California—already maintains or is developing longitudinal student record data systems. There are five such systems:

- **The California School Information System (CSIS)** has the responsibility to assign a unique, non-personally identifiable random student identification (CSID) number to all students in California public schools from kindergarten to grade 12 since 2004. It primarily maintains enrollment data on all students in the K-12 public education system since 2006.

- **California Longitudinal Pupil Achievement Data System (CALPADS)** was authorized by the legislature in 2003 to meet the federal, state, district and school reporting requirements of the *No Child Left Behind Act of 2001*.[15] It is under development and will not be completed until 2009 at the earliest.

- **California Community Colleges Management Information System (COMIS)** contains selected records on all students attending one of the 109 community college campuses throughout the state since 1992.

- **California State University Enrollment Recording System (ERS)** contains selected standardized records on students from all of the 23 CSU campuses since 1985, although a lesser set of records are available since 1979.

- **University of California Corporate Student System (CSS)** contains selected records on students from the nine UC campuses since 1999, although a lesser set of records are available since lesser complete records date since 1994.

These systems were primarily developed to support federal and state reporting requirements including IPEDS' reports to the U.S. Department of Education and for state budget requests. They were also developed to track students within the respective four segments. They are also used for internal management and research purposes.

In addition to the systems listed above, the **California Postsecondary Education Commission** maintains a student record data system that contains information from all

---

[15] CALPADS was authorized by the legislature primarily to meet the requirements of the *No Child Left Behind Act of 2001*. It directs the State Department of education to include in the system only the data required to meet the *NCLB* requirements (Hansen, 2007).

three postsecondary segments since 2005. This data system includes a limited subset of the records contained in the respective student data systems of the three state's postsecondary segments. This system, mandated by the state legislature in 1999 (AB 1570, 1999), was developed primarily for documenting the performance of postsecondary education institutions and for research.

The above data systems use two different unique student identifiers to track students across institutions within the four respective segments. The social security number (SSN) is used by the three postsecondary segments and by CPEC, while CSIS and the to-be CALPADS uses a 10-digit random California state identification (SSID) number.

Use of Existing Data Systems for Accountability

All of the above existing student record data systems have (or for CALPADS will have) the capability to support computation of state, district, and school or college on-time freshman cohort, completion, and dropout rates *within* each of the four segmental levels using their respective student identifiers.

The California State University and the University of California are currently using their respective student data system to compute the IPEDS graduation and completion rate at their respective institutions as part of their participation in the Consortium for Student Retention Rate Exchange (CSRDE).[16] For instance, CSU estimates that its average four-year on-time freshman graduation rate for its 2000 student cohort was 13 percent and its five and six-year completion rates were 36 and 48 percent, respectively. These rates count only graduates who started their college studies and earned a degree at the same CSU campus they started at; it does not account for students who transferred across university campuses within CSU or elsewhere.[17] In other words, it provides an *average* campus graduation or

---

[16] The CSRDE, founded in 1994, is a consortium of two-year and four-year colleges and universities "dedicated to achieving the highest levels of student success through their collaboratively sharing data, knowledge, and innovation. It publishes retention and graduation rates annually for hundred of educational institutions. See http://www.ou.edu/csred/consortium.html (As of December 17, 2007).

[17] http://www.calstate.edu/as/ (As of December 17, 2007). The California Postsecondary Education Commission (2006a) has also computed the 2000 cohort 4- and 5-year freshman cohort completion rates estimating them to be higher placing them at 17 and 40 percent respectively. Part the discrepancy may be due to differences in defining the freshman cohort. CSU defines it as including 34,454 freshmen who took a 12 credit workload in the first semester whereas CPEC defined it as 26,140 freshmen who took a full credit load for the full first year (27 semester units). Once again, this example underlines the importance of using a consistent definition of the group of students going into computing a graduation, retention or

completion rate, and is an underestimate of the true graduation rate because students who left and reenrolled at another CSU or other campus and eventually graduated, are not accounted for in the numerator, but are in the denominator. For K-12, such rates cannot be computed until 2010 when the first high school freshman cohort that received an SSID will graduate.

Although it does not provide a true graduation rate, the "IPEDS" rate allows to draw valid comparisons between schools or campuses *within* a segment as well as schools and campuses in other similar segments because every institution participating in the CSRDE is using the same definition.

Each of the segments' student data systems also currently have the capacity to compute graduation and completion rates *within* their segments, that is accounting for transfers from one K-12 school to another or one California state university to another California state university campus to another campus, although none of them have published these rates if they actually compute them. These rates would still not account for transfers to a private school, college or university within the state or an educational institution in another state and, in addition at the postsecondary level, for transfers across segments--such as students who transferred from a CSU to a UC campus and graduated from the latter.[18]

The state has two options to eventually account for transfers between segments, to private schools or colleges within the state, and out-of-state institutions. It can either link or integrate its student record data files across segments to account for transfers across segments and link the *resulting California integrated student record data* system with the student record data systems of other states and those of California private schools, colleges (discussed in the next section). Alternatively, it can develop a standardized algorithm to impute the graduation rates of students who have transferred. Implementing one or the other of the options is desirable considering that up to 20 to 25 percent of high school graduates going on to postsecondary education cross state lines and one out of five students

---

dropout rate. Unlike CSU, CPEC does not compute these rates for each new cohort. Also it appears to account for graduation from any of the CSU campuses not just from the ones freshmen started at, but CPEC does not specify. CPEC also computed 4-year graduation rates for students who took a full-time load at 30 percent—instead of 17 percent for all students—further underlining the importance of carefully defining the cohort for which these rates are measured.

[18] The California Postsecondary Education Commission (2006b), however, has done so for students who transferred from Community Colleges to CSU and UC campuses.

matriculated in one state will attend a college or university in another state before graduating (Edelman, 2007).

Imputing graduation rates of students who transfer involves estimating the probability that the individuals who have transferred out of one of the four California education segments have graduated from another school or college in another segment in the state or in another state. Such estimates are based on the assumption that transfer students have the same probability of graduating as the students of similar socio-economic and other characteristics (including grades or test scores) who have stayed.

Use of Existing Data System for Other Purposes.

Existing student record data systems are of limited use for the other three purposes discussed in the preceding section. They can be used in limited ways to identify students at risk as they contain a few of the data elements that are associated with students at risks such as IEP and LEP status, parent's education, family income. However, more potent factors associated with at-risk students such as attendance, grades in mathematics and English, and selected personal traits are not currently included in the segmented student data systems.

Similarly, existing student unit record data systems are of limited use to assess the effectiveness of preparation of students for college because of a lack of a link with the postsecondary data systems. They are also of extremely limited use to assess the effectiveness of the multiplicity of education improvement programs the state is funding, other than programs in which all students within a school or college participate such as charter schools. This is because information about which student participated in which program and for how long is currently not recorded in any of the existing student data systems.

Finally, current student data systems are of no use to assess return on investments simply because they are not linked with data systems that contain information on post-education activities, earnings, and federal and state program participation.

DESIRABLE ENHANCEMENTS TO EXISTING STUDENT DATA SYSTEMS[19]

To render the existing California student data systems more useful in the future, several enhancements need to be made:

- Add specific data elements

- Integrate the existing student unit record data systems from the four California education segments

- Link the integrated student unit record data file with other data files

- Broaden access to the resulting data set

- Improve data quality

- Enhance analytical capabilities

Add Specific Data Elements[20]

Ideally, each of the four California education segments should collect the same type of information on their students using definitions that are standardized across the four segments. Currently all four segments collect some socio-demographic data, but not always using the same set of measures or the same definition for the same measures. For instance, CALPADS will contain the immigration status of K-12 students while the postsecondary segments are collecting data on citizenship status. Although, these types of data are required to meet reporting requirements or for programmatic reasons, it would also be desirable that each segment collects and maintains a similar core set of socio-economic information using the same definitions.

Adding Data Already Being Collected. Information that is not currently stored or planned to be stored on the student unit record data files of one or more of the four

---

[19] The discussion in this subsection and the remaining of the paper is based primarily on a recent study conducted by the author and colleagues at the RAND Corporation. Twenty-nine state-level stakeholders, leaders and data systems administrators of each of the four education segments, and selected users of their respective data systems were interviewed for the study. We refer to them as "respondents" in the remainder of this paper. For a more comprehensive discussion of these issues involved in building a K-20 data student data system in California, see Vernez, Krop, Vuollo, and Hansen (2008).
[20] Two recent policy briefs, Loeb et al. (2008) and Perez (2008) make similar recommendations focusing exclusively on CALPADS, however.

segments, but that are already being collected by individual schools, districts or campuses include:

- Courses completed (CALPADS, CSU and UC)
- Grades for courses completed (CALPADS, CSU and UC)
- Dropout status and withdrawal (CSU and UC)
- College preparation test scores such as AP, ACT and SAT (CALPADS)
- Attendance (CALPADS).

Since these data elements are currently being collected by schools, colleges and universities or by testing companies, there would be minimal cost to add this information to the existing segmental student unit record data systems.

Data That Would Have To Be Collected. The most critical additional student information needed is *participation in the various education improvement programs* implemented by the state, the segments, or individual districts and colleges such as summer school, tutoring, remedial classes or courses, high-school-to-college transition programs, dropout prevention, class-size reduction programs. Without this information, the effectiveness of these policies and programs cannot be evaluated on an ongoing basis, with the alternative requiring the conduct of one-time evaluations, an approach that requires costly primary data collection on a program-by-program basis as was noted earlier. This information must be included at the student level if not all students within a school or college are participating in a specific program and even when all students in a school or college campus are participating in a specific program because some of the students in the school or college may transfer to another campus. By and large, educators at the school and college levels know which students participate in which program so that adding it to the information already collected at this school and college levels would not be a major additional burden.

Other desirable information that would require new and relatively expensive data collection would include teacher assessments of student traits that research indicate may help identify student at-risk and in need of focused attention or referral to special programs and school-related barriers and supports that are known to affect educational outcomes.[21] While desirable, collecting such information on an ongoing basis should probably wait until the reasons for dropping out of school or college are better understood.

---

[21] For instance, see Austin and Benard (2007)

Link the Four Segmental Existing Student Unit Record Data

To follow students' progress in their education from kindergarten through college requires linking or integrating the existing SUR data files of the four public California education segments in a first stage. It will also require establishing linkages with student unit record data systems maintained by other states and by private schools, colleges and universities within the state in a second stage (see next subsection). Without these linkages to the latter sets of data, graduation and completion rates that account for transfers to private institutions within the state and to out-of-state educational institutions cannot be computed accurately.

There are no technological barriers to the development of such an integrated so called K-20 and beyond SUR data system in California as the existence of such data systems in several states—including Florida, Utah, Virginia and Wisconsin—bear witness.[22] But, many interrelated issues—from major challenges concerning cooperation between the various California education segments and governance of the integrated student record data system to questions of unique student identifier and student privacy--will have to be resolved. These issues are discussed below along with suggestions for their resolution.

The main challenges to developing an integrated K-20 and beyond student record data system in California are political, not technological. It will require overcoming the culture of independence and the protective mindset of the four education segments and build a consensus from all interested parties that such a data system is desirable. Unlike any other state in the nation, the governance of California's public education system is divided between four relatively independent and powerful segments:

- a K–12 segment, under the administration of an elected state Superintendent of Education and a State Board of Education

---

[22] Lessons learned from the experience in developing statewide longitudinal student unit record data in these four states is available at http://dataqualitycampaign.org. Because California is the only state in the nation that has four quasi-independent education segments, the coordination and governance issues involved in building an integrated student data system are more complex in this state than any other states. As of 2006, 18 states had the capability to link their K-12 and postsecondary data and several others are developing this capability (Ewell and Boeke, 2007).

- the California Community Colleges, under the administration of a Board of Governors

- the California State University, under the administration of a Board of Trustees

- The University of California, under the administration of a Board of Regents.

In addition, the governor has recently added an appointed Secretary of Education in the governor's office to coordinate state education policies.

What are particularly unique about California are its three postsecondary segments:[23] Each has its own policies and practices for information collection, storage and dissemination. Even more fundamentally, the major policy instruments and practices to effectuate change, such as finance, accountability, curriculum, and standards, are not coordinated across the segments. As a result, the four segments have developed strong separate cultures and identities. According to our respondents, they do not have a tradition of collaboration and are protective of their independence and control the data they let out. They are suspicious of outside organizations evaluating their performance out of concern that the intent may be to embarrass them rather than to provide facts and information. These tendencies will have to be overcome before an integrated student record data system can move forward.

Under these circumstances, getting all stakeholders to agree on the desirability of, and on an approach to, developing, accessing, and using a California K–20 student data system will be difficult.[24] A legislative or executive mandate would go only so far. In the end, cooperation across segments and stakeholders will be required and it can only be gained over time. Experience in other states suggests that cooperation will come only if all

---

[23] The majority of states have only one postsecondary public segment and only five states have two postsecondary segments.

[24] Although gaining the broad support needed among stallholders in California is a sizeable challenge, there are some positive signs that it may be generated. The rapid increase in membership in the consortia of schools and colleges served by the California Partnership for Achieving Student Success (CAL-PASS) suggest that there is a growing interest at the ground level in using data-driven analysis to enhance student preparation and improve course articulation across educational segments. CAL-PASS maintain longitudinal student data on behalf of 11 regional consortia of K-12 schools, colleges and Cal State and UC campuses throughout the state and provide analytical support. Participation in these consortia, however, is voluntary and analytical reports are only shared among members of the consortia and are not made public (See Vernez et al. 2008).

stakeholders see the eventual benefits to them of having and using an integrated student record data system.[25] A majority of our respondents similarly emphasized the importance of showing the eventual long-term benefits of such a system to obtain buy-in.

Gaining the broad support and building a consensus among all stakeholders for such a student data system in California will be a sizable challenge. Experience in other states as well as interviews with stakeholders in California suggests that doing so will require an influential individual or group of individuals to champion its development, gather the necessary support from all interested parties and the public, and develop a consensus on its design.[26] Potential candidates for this role might be the governor, an influential legislator, or a commission that might include a representation of committed stakeholders and potential users.[27]

Beyond the issues of what data elements the integrated student data system ought to contain (discussed in the previous section) and who should have access to the system once developed (discussed in the next section), there are three additional key system design issues that would need to be addressed and resolved: *who should have decision-making authority over the design and access to the data system, who should operate and maintain the system, and what unique identifier should be used by the data system.*

The design and maintenance of an the integrated California student data record system, will require shifting some of the decision-making authority over several issues-- ranging from developing and overseeing the process of obtaining access to the data to addressing ongoing FERPA related concerns, determining the common set of data elements, standardizing data elements, defining data auditing requirements, and making IT decisions-- now held by the four segments to another "single organization." Currently, the only mechanism to address this kind of data system issues across the four education segments is

---

[25] Data Quality Campaign (2006a, 2006b, 2006c, 2006d).

[26] Vernez, Krop, Vuollo, and Hansen (2008). In Florida, Utah, and Wisconsin, the legislature of these states took the lead in authorizing and funding the development of their respective integrated or linked student data systems, prompted by the Governor in Florida and by the state education superintendent in Wisconsin while the governor took the lead in Virginia. Data Quality Campaign (2006a, 2006b, 2006c, 2006d).

[27] In its 2008 State of the State message, the governor announced his intent to establish an Education Data Commission to make policy recommendations for the development and implementation of an education data system for California public schools. Its focus would seem to be limited to enhancing CALPADS and the California Longitudinal Teacher Integrated Data Education System (CALTIDE). See http://www.gov.ca.gov/sots/text/education

the IT Inter-segmental Council. IT staff members from each segment are represented on this unofficial council. However, this council has no decision-making authority. A coordinating board or a data governing commission with representation from the key parties involved might serve this purpose provided that it is empowered with the appropriate authority.

A separate, but related, issue is who should be responsible for the day-to-day maintenance of the system. Because of the restrictions imposed by the federal Family Educational Rights and Privacy Act (FERPA) that strictly limits the sharing of identifiable students records, this issue is not trivial. FERPA does not permit the release of identifiable student data beyond a third party. Under this federal law, students are the first party (the owners of their records) and school districts, colleges, universities, the California Department of Education, and the governing boards of the community colleges, state universities, and University of California are the second party (responsible for maintaining *their respective* student data).[28] FERPA allows sharing of identifiable education records from second parties to third parties only under a narrow set of exceptions and third parties are not allowed to re-disclose identifiable data to anyone else.

The issue arises because of the fragmentation of the California education system. Two main options are available to address it.

If all that is desired is to maintain an anonymous integrated student data system, then the system could be lodged in any state, local, or private institution including anyone of the existing four education segments. Under FERPA, anonymous data can be disclosed without consent to anyone, although not without some restrictions to protect student privacy.[29] Developing this data system would require the assignment of a unique random identifier that is linked by a linking key to the identity of the student. Because the linking key would have to be kept by each of the four segments and anyone of the other state entities that may share their data, it would require *each* of the segments and other entities providing data to the integrated data system to generate the unique random identifier for their respective student

---

[28] There is no unanimity among the four California education segments on the interpretation of FERPA and privacy state laws when it comes to limitations and requirements for sharing data between segments as well as with outsiders.

[29] For instance, information that is based on few observations and, hence, that may allow someone to identify the identity of an individual(s) involved, cannot be released.

records. This requirement may prove to be demanding on the segments and other parties and unwieldy to coordinate.

If, to the contrary, what is desired is to maintain an identifiable (and non-identifiable) integrated student data system, another option is to designate a new or existing state institution to take responsibility for the data system *on behalf* of each of the segments and other state entities providing data to the integrated data system. Under this arrangement, the designated institution would be considered a second party acting on behalf of each segment. The advantage of this option is that the identifiable data could then be made available to third parties within the limitations of FERPA. A potential model for such an option is the Florida Education Data Warehouse (EDW).[30] Similarly, Kentucky's state legislature is currently considering a bill that would authorize the development of such an education data warehouse.[31]

The last issue is the question of what unique student identifier to use to facilitate the linking of the student records across the four education segments and, eventually, with other administrative data files. Clearly, using the social security number (SSN) would be the most effective option given that all three postsecondary segments and all other state agencies that might become part of the integrated data system, use it to link their individual records over time. The question arises of what to do with the SSID that is now used at the K-12 level as the common identifier.

One option would be to also start requiring use of the SSN at the K-12 level. Since the legislature only recently passed the legislation to create the SSID, it is unlikely that it can be undone. Another, only partial but probably more feasible option, would be to require that the SSID, in addition to the SSN, be provided to the postsecondary segments where the students' SSID could then be matched with their SSN. This option, however, would leave out of the system all students that leave the state to continue their education or remain in the state but do not continue their education past high school as well a students in private schools and colleges. However, these individuals could potentially be linked using a key

---

[30] http://edwapp.doe.state.fl.us/doe/ (December 18, 2007)
[31] Interviews with state staff, April 24, 2007. The Kentucky Department of Education, Postsecondary Council, and Education Professional Standards Board submitted a proposal for a "P-16 Seamless Data Warehouse Project," seeking an initial $3 million in the 2006–2008 budget periods, to be followed by $4 million in the subsequent budget period. However, the proposal has yet to be funded.

made up of some portion of the student's first and last name, birth date and some other demographic characteristics.[32]

Link with Other Data Files

As noted in the preceding section, for computing comprehensive graduation and completion rates, for identifying the economic activities that student engage in after leaving education and for assessing the performance of California students in their adult life and returns on educational will require also linking the eventual integrated student records with a broad set of other federal, state and other individual data records systems. These include:

For students who transfer out of the California education system into other educational institutions: The National Student Clearinghouse maintains records of from some 3,200 high schools and more than 2,000 colleges or universities that students throughout the nation attend (or have attended) and of the degrees that they have obtained.[33] It does not, however, contain information on courses taken, GPA, or other information. For this information, it would require linking with other states student data records. Since most students resident of California attend schools in neighboring states—Arizona, Nevada, Oregon, and Washington—linking with student from these states may be sufficient. Finally, the Association of Independent California College and Universities maintain records on students that attend California private colleges and universities. It includes more than 90 state colleges and universities with more than 250,000 students, about 21 percent of the state's college students.[34]

For employment, unemployment, and earnings. For California, the primary source of data on employment, unemployment and earnings and participation in training programs for residents is the California Employment Development Department. Beyond California, other potential sources of information on employment and wages after leaving education are large employers such as the Department of Defense, Federal government, and the postal services.

---

[32] The California Partnership for Achieving Student Success (Cal-PASS) is using such a key to match student data it obtains from participating schools, colleges and universities with reportedly a 90 to 95 percent matching rate.

[33] http://www.studentclearinghouse.org (As of January 15, 2008).

[34] http://www.aiccu.edu (As of January 15, 2008).

For social and health services: Several state agencies maintain individual records of the use of their services by state residents including the California Health and Human Services Agency—covering child support, health care, mental health, rehabilitation and social services—and the California Department of Corrections—covering prisons, juvenile delinquency and vocational rehabilitation.

In the long-term, it may also be desirable to integrate the K-20 SUR with existing data about individual teachers,[35] faculty, and institutional financial information. Integration of these data with student data raises union and confidentiality issues that could not be explored in this study.

The issues raised in linking these data sets with an integrated K-20 student record data system raise similar issues to the ones discussed about linking or integrating the existing four segmental student record data systems.

Unlike with developing and maintaining an integrated student data record system, there is yet little experience with linking longitudinally students records with their adult data records. Only one state—Florida—has developed considerable experience doing so. Beginning in 1988, it has matched its student records within its Education and Training Placement Information Program (FETPIP), with records of the postal services, department of defense, and the office of personnel management at the federal levels, and those of four state agencies: the Department of corrections (for incarcerations and recidivism), the State Department of Children and Families (for public assistance), the State Department of Administration (for career service employment) and the Florida Department of Labor and Employment Security (for employment and wages).[36]

---

[35] A parallel longitudinal data system for teachers, The California Longitudinal Teacher Integrated Data Education System (CALTIDE), was authorized by the legislature in 2006 (SB 1614). As for K-12 students, teachers are to be assigned a unique California ID. CALTIDES is meant to contain information on teacher credentials including teacher preparation program, certification including NCLB highly qualified status for specific subject matters, selected demographic characteristics, course assignment, in service-education and staff development. It will not include teacher name, SSN or other identifiable information. Although the legislation specifies that the data "may not be used, solely or in combination, with data from CALPADS, for purposes, of pay, promotion, or personal evaluation of an individual teacher or groups of teachers… (SB 1649, Sec. 3.c)," it does not seemingly prevent its eventual use in combination with CALPADS to assess the effectiveness of teacher preparation and development programs.
[36] See Pfeiffer (1999) for a description of the Florida Education and Training Placement Information Program.

Ease Access to the Data

Currently, limited use of even the four segments student record data systems is being made to drive decision-making and to conduct rigorous evaluations of educational programs and investments for several reasons. Ready access to the data is limited to designated personnel within each of the segments and the data is used mainly for the preparation of federal and state required yearly reports. Also, segmental staff has little time to engage in research other than may be required for institutional management purposes. And even though analysts outside the segments, including research organizations, can request access to the segments' student data they rarely have done so because of the cumbersome request process involved: For fear of litigations, the segments have been interpreting federal and state student privacy statutes conservatively, even when the request is for anonymous student level data.[37]

It is promising that the legislature has indicated its intent in the legislation that authorized the development of CALPADS that its data will be accessible not only to schools, school districts and state agencies responsible for education, but also to legislative policy analysts and researchers:

> It is the intent of the legislature to . . . establish state data management practices that require the development of specific categories of users and uses for pupil data and establish responsibility for establishing and servicing users as well as responsibility for establishing and posting protocols, criteria, and procedures for use that are developed in a manner that is consistent with recommendations of the State Department of Education's advisory committee on privacy and data protocol. *Approved users should include schools, charter schools, state agencies with responsibility for education, legislative policy analysts, evaluators of public school programs, and education researchers from established research organizations* (emphasis added).[38]

However, it remains to be seen whether the recommendations of the State Department of Education's advisory committee on privacy and data protocol will follow suit and, if it does, whether the same level of accessibility would be provided for data from the postsecondary segments and for data from other state agencies. In federal FERPA and California state laws, there appears to be no barriers to making non-identifiable longitudinal

---

[37] An exception is the Community Colleges that have allowed access to the COMIS data in its anonymous form.
[38] SB 257, Chapter 782, section 1 (c).

individual student level data to non-state educational organizations such as research organizations and individual researchers.[39] Non-identifiable student data is all that is required for nearly all possible use that researchers and other non-educators would need it for.

Under the FERPA "studies exception," personally identifiable data may also be provided to a broad set of third-party users, including researchers of state and local educational institutions. An educational agency (or institution acting on behalf of an educational agency) may release personally identifiable data to an organization conducting studies on its behalf. However, the purpose of the studies must be specifically to "develop, validate, or administer predictive tests, administer student aid programs, and improve instruction."[40] Most interpretations of the FERPA studies exception also allow for independent research to be conducted so long as precautions are taken:

> "This authority may be implemented to include disclosing personally identifiable information to organizations for *independently initiated studies* by promulgating state regulations or guidelines to provide access for authorizing studies initiated by third parties in which the local education agency or school has a clear interest (whether or not it funds the study) and including strong privacy protections against re-disclosure, consistent with FERPA."[41]

In brief, and subject to the appropriate procedural precautions and purposes specified for studies, there appear to be no privacy barriers encoded in national or state laws that would prevent third parties from accessing individual student data, identifiable or not—including researchers from organizations outside California's four education segments.

States that have developed integrated or linkable longitudinal data systems have made them accessible to outside researchers. Florida does so on a quid pro quo basis requiring that the research questions be of interest to the state and retaining the right to review the methodology and findings. Research organizations that have used the Florida data include Harvard University, the Manhattan institute, the University of Chicago, and Florida University, among others. Texas has provided its integrated student data to researchers at the University of Texas and the state demographer. In addition, Texas has provided funding

---

[39] Greene and Perkins (2007) and Coleman, Palmer and Winnick (2006).
[40] Greene and Perkins (2007), p. 14.
[41] Greene and Perkins (2007), p. 15.

to establish up to three education research centers that will be expected to provide access to the Texas integrated student data system to outside researchers in addition to its own staff.[42]

Enhance Data Quality

Users of the existing segmental student data systems say that the quality of the data varies broadly depending on data elements. Some data elements are unevenly reported across schools and campuses, others are inconsistently coded, and some districts and campuses use different definitions for the same data elements. Data that have financial implications—such as enrollment--for the respective institutions are generally accurate while other data elements are less so. Because no one school or campus uses its respective segmental data system and the data it sends rarely comes back to them, the data at the source rarely get cleaned up.

States with experience with developing and maintaining student record data systems indicates that quality improves as the data are used for resource allocation and other public reporting at the school and campus levels, that is, when the schools and campuses have a stake in the accuracy of the data released.[43] Quality, reportedly, also depends on how experienced and trained the school or campus-level staff collecting, transferring, and using the data are. Increasing the quality of the data collected for CALPADS and the three postsecondary segments will require providing schools, districts, and campuses access to the data. It will also require providing them with the staff resources and training needed to collect, transfer and use the data, especially at the K-12 level.

Enhance Analytical Capabilities

To make maximum use of the enhanced and integrated K-20 student record data system as described in these pages not only would require making it broadly accessible to individual schools, colleges, universities, and faculty, state officials, and researchers within and outside the four education segments, but also would require the development of a dedicated analytical capability. States that have developed such student data systems have

---

[42] Vernez, Krop, Vuollo and Hansen (2008).
[43] Data Quality Campaign (2006a, 2006b, 2006c, 2006d).

recognized this need. Florida, for instance, has an analytical staff of eleven attached to its data system "warehouse" to meet request from its legislators and educational institutions in addition to making its data available to outside researchers as does Texas. The latter's legislature has provided funds to create an internal 'think tank" to insure that maximum use is made of its data as well as to establish three education research centers in the state.[44] These centers will have access to the K-20 Texas SUR data and will be expected to provide outside researchers access to it. Several of our California respondents have indicated the similar need to develop an adequate capability dedicated to analyzing student record data. Underlining this need, one of our respondent stated "California has data, but not the information needed for policy decisions."

CONCLUSIONS

To track the progress of its students, accurately measure the performance of its students, schools, and colleges, and assess the effectiveness of its education programs and investments, California needs to develop a longitudinal student record data system integrating information from its four education segments with added linkages to a broad set of other in-state and out-of-state data sets.

The state already has the basic infrastructure needed to develop such a data system. Each of its four education segments—elementary and secondary, community colleges, state university and university of California—already have (or in the case of K-12 is building) a student record data system that permits tracking students over time and across schools, colleges and college and university campuses within each of these segments, but not across the segments. However, the information that these data systems currently contain are not sufficient for: measuring performance; evaluating programs; identifying at-risk students and designing program interventions; assessing the preparation of students for the workforce; and assessing the returns on education investments.

To increase the utility of the current segmental student data systems, several enhancements are required:

---

[44] State staff interview, April 24, 2007.

Student information (including grades for course completed, college preparation test scores, student attendance, education programs students participate in) that is currently collected by schools, districts, and or colleges ought to be added to the existing student record of each of the segmental data systems.

The four segments' existing data systems enhanced with the added data elements listed above ought to be linked, or preferably integrated into a unique student data record system for the state. Over time, the capacity to link this integrated student data system with other in-state and out-of-state data systems containing information on students transferring to private or other states educational institutions, and on students activities after they have completed their education, ought to be added incrementally.

Access to the existing, and eventually to the integrated, student record data systems, ought to be made available (in an anonymous format) to schools, districts and campuses, as well as to outside interested parties and researchers. While FERPA and state laws limits the options under which the integrated data system can operate, there are few restrictions to allowing access to anonymous student data.

Standardization of the data across segments, and the quality of the data collected, needs to be given more focused attention. This will require providing staff resources and training at the school, district and campus levels.

An independent dedicated capability that is adequately funded ought to be established to analyze on an ongoing basis, the data from the integrated student data record system, provide objective measures and assessments of the state of California's education, and support data-driven decision making. The capability could be part of the entity that eventually operate the integrated student data system or could be lodged in an existing research center or university.

Doing so will require addressing several challenges, including overcoming the culture of independence and the protective mindset of the four education segments, shifting some of the decision-making authority on data requirements and IT decisions away from the four segments' governing boards, establishing an entity that has the authority to make decisions on the design of the integrated student data system, and another (or the same) entity that is given the responsibility to maintain and operate the data system. Decisions will also have to

be made regarding a unique student identifier across the four segments and about who will have access to the data under what circumstances.

Experience with other states, and our own study of the feasibility of developing an integrated student record system for California,[45] suggests that a strong advocate—the governor, a legislator, a well-connected business leader—will need to promote for the development of an integrated student record data system for the state. Legislation will be required because many of the issues involved are sensitive and important, and need to be aired, debated, and eventually resolved in an open public debate. It is not likely that the disparate views about governance of the proposed data system and access to the data could be resolved without legislation authority.

---

[45] Vernez, Krop, Vuollo, Hansen (2008); and Data Quality Campaign (2006e).

REFERENCES

Allensworth, Elaine M., and John Q. Easton, *What Matters for Staying On-Track and Graduating in Chicago Public High School*, Consortium on Chicago School Research At the University of Chicago, July 2007. As of January 21, 2008: http://ccsr.uchicago.edu/content/publications.php?pub_id=116.

Austen Gregory, and Bonnie Benard, *The State System to Assess Learning Barriers, Supports, and Engagement: Implications for school Reform Efforts*, Los Alamitos, Calif.: West ED, 2007. As of April 28, 2008: http://www.californiaschoolfinance.org/portals/0/PDFs/Policy/WestEd_brief.pdf

California Policy Convening, *Addressing California's Data Needs: Implementing Comprehensive, Longitudinal Systems at the Local and State Levels*, submitted by Practitioner Working Group for the October 19, 2007 California Education Policy Convening, 2007. As of January 21, 2008: http://www.californiaschoolfinance.org/tabid/169/Default.aspx.

California Postsecondary Education Commission, *Have They Finished? Time-to-degree of Freshman Entering UC and CSU in Fall 2000*, Working Paper WP/06-02, 2006a. As of January 21, 2008: http://www.cpec.ca.gov/completereports/2006reports/WP06-02.pdf.

----------, *California Higher Education Accountability: Goal-Student Success: Measure: Four-Year Degrees Conferred on Community College Transfer Students,* Working Paper WP/06-18, 2006b. As of January 21, 2008: http://www.cpec.ca.gov/completereports/2006reports/WP06-18.pdf.

Coleman, Arthur, Scott R. Palmer, and Steven Y. Winnick, *State Longitudinal Data Systems and Student Privacy Protections Under the Family Educational Rights and Privacy Act*, November 2006. As of November 27, 2007: http://www.hklaw.com/content/whitepapers/StateLongitudinalsDataSystems2.pdf.

Data Quality Campaign, Newsletter, Volume 3, Issue 1, January 2008. As of February 6, 2008: http://www.dataqualitycampaign.org/files/Publications-Newsletter_Jan08.pdf.

----------, *Florida Case Study: Building a Student-Level Longitudinal Data System*, 2006a. As of November 27, 2007: http://www.dataqualitycampaign.org/files/State_Specific-Florida_2006_Site_Visit.pdf.

----------, *Utah Case Study: Building a Student-Level Longitudinal Data System*, August 2006b. As of November 27, 2007: http://www.dataqualitycampaign.org/files/State_Specific-Utah_2006_Site_Visit.pdf.

----------, *Virginia Case Study: Building a Student-Level Longitudinal Data System*, August 2006c. As of November 27, 2007: http://www.dataqualitycampaign.org/files/State_Specific-Virginia_2006_Site_Visit.pdf.

----------, *Wisconsin Case Study: Building a Student-Level Longitudinal Data System*, August 2006d. As of November 27, 2007: http://www.dataqualitycampaign.org/files/State_Specific-Wisconsin_2006_Site_Visit.pdf.

------------, *Building Student-Level Longitudinal Data Systems: Lessons Learned from Four States*, August 2006e. As of November 27, 2007: http://www.dataqualitycampaign.org/files/State_Specific-2006_Site_Visit_Synopsis.pdf.

Education Week, *Diplomas Count: An Essential Guide to Graduation Policy and Rates*, Washington, D.C.: Editorial in Education, June 2006.

Ewell, Peter, and Marianne Boeke, *Critical Connections: Linking State's Unit Record Systems to Track Student Progress*, Indianapolis, In.: Lumina Foundation for Education, 2007.

Loeb Susanna, Tara Beteille, and Maria Perez, *Building an Information System to Support Continuous Improvement in California Public Schools*, Berkeley, Calif.: Policy Analysis for California Education, Policy Brief 08-2, 2008. As of April 28, 2008: http://lmri.ucsb.edu/publications/reserachreports.php.

Greene, David, and Paula Perkins, Legal Memorandum to Pamela Burdman, Program Officer at the William and Flora Hewlett Foundation, First Amendment Project, Oakland, Calif., February 9, 2007.

Hall, Daria, *Getting Honest About Grad Rates: How States Play the Numbers and Students Lose*, the Education Trust, June 2005

Hansen, Janet S., *Education Data in California: Availability and Transparency*, Palo Alto, Calif.: Stanford University: Institute for Research on Education and Practice, 2007. As of August 7, 2007: http://irepp.stanford.edu/documents/GDF/STUDIES/15-Hansen/15-Hansen(3-07).pdf

Heckman, James J., and Paul A. LaFontaine, *the American High School Graduation Rate: Trends and Levels*, Cambridge, MA: National Bureau of Economic Research, Working Paper 13670, 2007. As of February 6, 2006: http://www.nber.org/papers/W13670.

Integrated Postsecondary Education Data System, *Graduation Rates for 4-Year Institutions for Academic Reporting*, no date. As of January 21, 2008: http://nces.ed.gov/ipeds/web2000/printforms.asp?formid=13.

Larson, Katherine A., and Russell W. Rumberger, *ALAS*: "Achievement for Latinos Through Academic Success." In Helen Horton (Ed.), *Staying in School: A Technical Report of Three Dropout Prevention Projects for Middle School Students with Learning and Emotional Disabilities*, Minneapolis, MN: University of Minnesota, Institute on Community Integration, 1995.

Mishel, Lawrence, and Joydeep Roy, *Rethinking High School Graduation Rates and Trends*, Washington D.C.: Economic Policy Institute, 2006.

National Governors' Association, *A Compact on State High School Graduation Data*, no date. As of January 3, 2008: http://www.nga.org/Files/pdf/0507GRADCOMPAC.

Neild, Ruth Curran, and Robert Balfanz, *Unfulfilled Promise: The Dimensions and Characteristics of Philadelphia's Dropout Crisis, 2000-2005*, Philadelphia Youth Network, The John Hopkins University, and University of Pennsylvania, 2006. As of January 21, 2008: http://www.projectuturn.net/downloads/pdf/Unfulfilled_Promise_Project_U-turn.pdf.

Perez Maria, Larisa S. Shambaugh, and Tom Parrish, *Answering the Knock of Opportunity: Addressing the Needs for California's English Learners*, Pala Alto, Calif.: Institute for Research on Education Policy and Practice, Stanford University, February 2008. As of April 28, 2008: http://lmri.ucsb.edu/publications/reserachreports.php

Pfeiffer, Jay J., *Student Follow-up Using State and Federal Administrative Records: Lessons from Florida's Education and Training Placement Information Program (FETPIP)*, Tallahassee, FL: Florida Department of Education, 1999.

Rotermund, Susan, *Alternative Education Enrollment and Dropouts in California High Schools*, California Dropout Research Project, UC Santa Barbara; Statistical Brief 6, 2007a. As of January 17, 2008: http://www.lmri.ucsb.edu/Dropouts.

Rotermund, Susan, *Why Students Drop Out of High School: Comparisons from Three National Surveys*, California Dropout Research Project, UC Santa Barbara; Statistical Brief 2, 2007b. As of January 21, 2008: http://www.lmri.ucsb.edu/Dropouts.

Rumberger, Russell W., *What is California's High School Graduation Rate*, California Dropout Research Project, UC Santa Barbara, Statistical Brief 4, August 2007a. As of January 17, 2008: http://www.lmri.ucsb.edu/Dropouts.

Rumberger, Russell W., *Early Predictors of High School Graduation and Dropout*, California Dropout Research Project, UC Santa Barbara, Statistical Brief 3, June 2007b. As of January 21, 2008: http://www.lmri.ucsb.edu/Dropouts.

Vernez, Georges, Cathy Krop, Mirka Vuollo, and Janet S. Hansen, *Toward a K-20 Student Unit Record Data System for California*, Santa Monica, Calif: The RAND Corporation, MG-695-WFHF, 2008.

APPENDIX A

California Measures of High School Graduation Rates

California reported two widely different high school graduation rates for the state in 2005-06: 83.2 and 67.1 percent.[46] The first estimate is based on school districts provided dropout figures and was reported for the purpose of compliance with *No Child Left Behind* while the second estimate is based on approximating the progress made by a cohort of high school freshman through graduation within four years. Still, others have placed California's high school graduation rate somewhere between 71 and 74 percent.[47]

The first estimate noted above uses a methodology based on aggregated dropout data reported by school districts each year during the 9th grade, then the 10th grade, and so on to the 12th grade.[48] This methodology is likely to underestimate the true dropout rate because schools have no sure way of knowing whether a student who has left has actually dropped out or is actually attending another school in another district or in another state unless the parents tell them.[49] Also, it does not include students that obtained a GED, are pursuing a GED or other education program, or are attending a college offering an associate degree or baccalaureate.[50]

The lower estimate of 67.1 percent noted above is derived simply by dividing the number of high school graduates at the end of the 12th grade in 2005-06 by the enrollment in grade 9 four years earlier to approximate the rate at which a cohort of students entering high school eventually graduates on time. This measure also has several limitations. It does not account for students that come in and out of the state during the four year period and may underestimate the true graduation rate when the student population increases during the 4-year period. Also, it does not account for students that take more than 4 years to graduate. Finally, it does not account for the significant bulge in the number of students at grade 9 because of the disproportionate number of students that repeat this grade. Grade repetition creates ambiguities in the count of graduates (the numerator) because it

---

[46] http://cde.ca.gov/dataquest (As of November 13, 2007).

[47] Rumberger (2007b)

[48] This high school graduation rate is based on the number of graduates at end of 12th grade divided by the number of graduates at end of 12th grade + dropouts in grade 9 + dropouts in grade 10 + dropout in grade 11 + dropouts in grade 12) * 100.

[49] Hall (2005).

[50] Students who have dropped out are reported by schools using a state definition that includes more than 20 different reasons for dropping out. Schools typically do not have the means and have little incentives to verify these reasons.

includes students that did not graduate in 4 years (those who repeated the 9th grade) and in the count of students in grade 9 (the denominator) because it includes students who are repeating that grade. In this state, substantial number of students repeat 9th grade as suggested by the 56,000 additional students enrolled in grade 9 in 2005-06 than in grade 8 in 2004-05.[51] When all 9th graders are included in the high school graduation rates, the repeating students are counted twice in the denominator, the first and second time they are in 9th grade. It would be more appropriate to use the number of students who are entering the 9th grade for the first time as the base denominator, a piece of information that is not currently available

Analysts have used different methods to adjust for these issues. To adjust for population growth or decline, U.S. Census Bureau or state data estimates of changes in the population of high school students have been used. These adjustments can only be made at the state or county level, however, and are based on net migration. This means that they do not account for the likely possibility that the students that transfer in to the State's education system may have different socio-economic and other characteristics than the students that transfer out and hence may have different graduation rates.

To adjust for repeaters of the 9th grade, one method is to use the number of students in the 8th grade as the denominator for computing the graduation rate. This method yields a 2005-06 high school graduation rate for California of 75.7 percent. This rate includes in the count of graduates students who have returned from private to public schools and may overestimate the true graduation rate. Another approach, preferred by analysts, is to use an average of reported 8th, 9th, and 10th grade enrollment as the denominator. Using this approach results in a somewhat lower 2005-06 graduation rate of 71.2 percent than the preceding method, but a higher rate than 67.1 percent based on 9th grade enrollment alone.

Yet, a different way of computing the high school graduation rate based on aggregate student data and developed by *Education Week* is to create an artificial synthetic cohort made up of students in successive grades. Dropouts are accounted for by examining the ratios of the numbers of 10th, 11th, and 12th graders and graduates in a given year to the numbers of 9th, 10th, 11th and 12th graders the previous year. Multiplying these ratios together yields an estimate of the on-time

---

[51] A share of this increase may be accounted for by students who are returning to public schools from private middle schools.

32

graduation rate for 9th graders.[52] This methodology has the advantage of accounting for migration in and out of the state but, does not deal with the repeaters issue. This approach results in a 2005-06 graduation rate of 64.5 percent for California.

The above five different methods of computing the California's graduation rate yields five different graduation rates that for 2005-06 varied from a low of 64.5 percent and a high of 83.2 percent, a spread of 19 percentage points, each with limitations that makes it impossible to know which is the true high school graduation rate in the state. Most of the issues discussed above in estimating the California's graduation and dropout rates can be addressed, however, were a SUR data system developed and maintained for the state.

---

[52] Education Week (2006).

APPENDIX B

Table B.1
Key Data Elements for a K-20 SUR Data System

Data Elements	Accountability	Program Evaluation	Students at Risk	Return on Investments
Demographic / Socio-Economic				
Name	✓			
Age	✓			
Gender	✓			
Race / ethnicity	✓			
IEP status	✓		✓	
LEP status	✓		✓	
Immigrant status	✓			
Family income	✓			
Single parent HH			✓	
Parents' education			✓	
Dropout older sibling			✓	
Alone after school			✓	
Student financial				
Enrollment				
Enrollment status	✓			
Temporary out-of-school status	✓			
School / college Id	✓			
Grade	✓			
Attendance			✓	
Achievement Scores				
College test scores (ACT, SAT)	✓	✓		
State test scores	✓	✓		
Graduation Status				
HS / college degrees	✓			✓
GED	✓			✓
Alternate certificate	✓			✓
Dropout status / withdrawal reasons				✓

Table B.1
Key Data Elements for a K-20 SUR Data System

Data Elements	Accountability	Program Evaluation	Students at Risk	Return on Investments
Courses & Programs Participation				
Courses completed		✓	✓	
Course grades		✓	✓	
Programs participating in		✓		
Length of time in programs		✓		
College major				✓
Outcomes				
Employment by type				✓
Wages / income				✓
Participation in:				✓
Social programs				✓
Health programs				✓
Corrections				✓
Military service				✓
Students Traits				
Supervision requirements			✓	
Social interaction skills			✓	
Difficult-to-teach			✓	
Motivation			✓	
School Climate				
Teacher commitment to improve school			✓	
Coherence of instructional program			✓	
Personalization in the classroom			✓	
Press on students for academic achievement			✓	