Why Is Student-Level P-20 Data Sharing Important?

What New England can learn from the Data Quality Campaign

NANCY J. SMITH

Many states currently share information between P-12 and higher education; however, the resulting reports are usually based on aggregate counts of students instead of individual student data. For example, in many states higher education institutions provide "high school feedback" reports to individual high schools that include figures such as the total number of the school's graduates enrolled in that institution and the percentage of those students enrolled in college remediation courses. While useful to administrators, these reports can be enhanced significantly when supported by student-level information.

As of the 2007 Data Quality Campaign (DQC) annual survey, 22 states reported having the ability to share P-12 student- level data with higher education systems. With the ability to match *student-level* records between P-12 and postsecondary systems, policymakers and educators will be able to know, for example:

- The percentage of each district's high school graduates who enrolled in public higher education institutions within 15 months after graduation, so that the data can be analyzed by student demographics or special services received.
- The percentage of last year's graduates from each high school or school district who needed remediation in college, and how this percentage varied by student poverty status and ethnicity.
- The percentage of students who met the proficiency standard on the state high school test and still needed remediation in the same subject in college.
- How students' ability to stay in and complete college is related to their high school courses, grades and test scores.
- The percentage of students receiving special education services in P-12 who go on to public higher education institutions in the state.

The desire and commitment to answer questions such as these require a general culture change among educators and policymakers throughout the P-20 spectrum across the country in terms of how they view and use data systems.

Educators and policymakers are realizing the power of using data beyond the purpose of meeting state and federal reporting requirements. To support national and state goals to increase student achievement, education leaders are increasingly acknowledging the key role of data, not only for accountability and transparency, but also to inform continuous improvement.

Student-level longitudinal data should be collected and used for strategic planning of course offerings, both in P-12 and higher education, cost-benefit analysis of district and school-level programs and evaluation of course and program effectiveness. In addition, with the ability to connect P-12 student data to higher education data and teacher data, a state can evaluate the effectiveness of its teacher preparation programs.

P-20 systems can create a rich picture of student progress and school, district, and program performance. Longitudinal data helps monitor student progress across years, diagnose difficulties in specific areas of learning and prescribe interventions, conduct internal (to the school or district) and external benchmarking studies, predict which students are likely to succeed at the next level, and evaluate school, district and program effectiveness.

Observed Challenges to Forming a P-20 Data System

1. *Perceived barriers due to Family Educational Rights and Privacy Act (FERPA) regulations inhibit data sharing.*

In recent years, as more states have begun to build student-level longitudinal data systems, increased attention has been focused on clarifying FERPA and its regulations about using and sharing data. FERPA was enacted in a time when local education agencies (LEAs) were the primary collectors and users of student data, before state education agencies (SEAs) found it necessary to collect individual student data to fulfill U.S. Department of Education (USED) reporting requirements and long before technology provided the hardware and software to collect, clean, analyze and report massive amounts of data at a time. The economies of scale that current technology allows SEAs to provide their LEAs and the emphasis on using data for outcomes-based research have enabled a huge shift in the culture of education, but the FERPA language has not experienced a similar shift to address current goals, concerns and abilities in terms of educational data use.

In order to help states navigate the ambiguous and sometimes-conflicting rules, the DQC, through the Washington, D.C law firm of Holland and Knight,

In general, the SSN is collected but is not used as the primary unique student ID for a variety of reasons. For privacy and data quality purposes, it is recommended that states randomly assign each student a unique student ID.

LLP, conducted a detailed analysis of FERPA in relation to student-level longitudinal data systems. (See *Maximizing the Power of Education Data while Ensuring Compliance with Federal Student Privacy Laws: A Guide for Policymakers*, online at www.dataqualitycampaign.org.

Essential findings in analysis of FERPA:

- Sharing student data that are not personally identifiable is permissible.
- Even in instances in which personally identifiable information on students is shared, there are several types of disclosures that are permissible:
  - Evaluating/auditing state and local programs and implementing school and district accountability;
  - Monitoring and analyzing assessment, enrollment and graduation data;
  - Performing studies to improve instruction; and
  - Sharing student records among schools.

2. *Perceived problems with collection of student Social Security numbers (SSN) affects matching data between P-12 and higher education.*

Many states make the collection of the SSN an option for schools and parents, while others have policies in place that prevent the collection of the SSN. Rarely is the SSN used as the primary unique student identifier (ID), even in states that do collect it. Other than state-specific laws or policies, however, there are no strong objections from experts in the education data community to collect and use the SSN as one of a series of identifiers when assigning unique student IDs and matching student records over time and/or across schools or districts.

In general, the SSN is collected but is not used as the primary unique student ID for a variety of reasons. Primarily, there are concerns about student privacy, but there are also data quality concerns since not all students have an SSN and parents sometimes provide the same SSN for all students in their family. So, for privacy and data quality purposes, it is recommended that states randomly assign each student a unique student ID.

The SSN, while not 100 percent unique, does increase the probability of tracking the same student across databases. States usually use an algorithm of matching identifiers such as first name, middle initial, last name, date of birth, gender and SSN to create a unique ID. Given the plethora of common names and dates of birth (especially considering twins or other multiple births) and the existence of data entry errors, as well as the high mobility levels among students, it behooves each state to include as many options for identifying each particular student as possible. Analyses conducted in some states indicate that including the SSN in the matching algorithm significantly increases the match rate of student records from different sources.

The onus, then, if SSNs are included in the database, is on the data collectors, users, and researchers to make sure that they apply due diligence to not inadvertently release the SSN or any other personally identifiable information.

3. *Data system infrastructure is decentralized.*

Many SEAs are currently undergoing significant technology and cultural changes to accommodate a student-level longitudinal data system. Historically, data systems within the SEAs have been decentralized, such that each program area (e.g., Title I, Special Education, English Language Learner/ Bilingual) have collected their own data from school districts—often duplicating efforts. Usually, the information technology division also has periodic data collections to meet other reporting and information needs.

This type of data system infrastructure makes data sharing, analysis and reporting problematic, since different areas sometimes produce conflicting reports because they are not each using the same source data file and often ask for similar information in different ways. By having a centralized data collection system at the student-level that is used to feed each program area's analyses, the SEA is more likely to reduce the burden on district, free up staff time from so many data collections, and have more accurate data and reports.

Similarly, at the postsecondary level, if public postsecondary institutions fed their student-level data into a centralized higher education student data system, it would be easier to conduct research, create standardized reports, reduce analytical burdens across institutions, and share data with the P-12 arena.

Actions to Ensure Sustained P-20 Student-Level Data Sharing:

1. *Foster political buy-in from all levels of ongoing support to ensure that it remains after personnel and leadership changes.*

There should be a consistent mission and goal within the education community—from the governor, legislature, postsecondary regents and institutions to the state board of education, SEA and down to the local education agencies—that data are important and will be used to inform policy and funding decisions and improve student achievement and teacher effectiveness from P-12 through higher education. Without a common message and purpose about the value of data, it is unlikely that all stakeholders will work from the same perspective to create and use a true information system.

The P-20 Councils, SEAs and postsecondary organizations need on-going support to keep the P-20 data sharing initiative functioning and up-to-date over time. Whether or not the council and support for longitudinal data systems are provided for in state statutes, there should be clear and institutionalized statewide support from policymakers that is sustained long-term to prevent the weakening of the longitudinal systems and data-driven decision-making after a change in statewide leadership, such as a change in governor or commissioner of

either higher education or elementary and secondary education.

2. *Designate one or more organizations to act as an authorized evaluator/researcher of student and teacher data for the purpose of improving student achievement.*

The SEA and the state's higher education coordinating board each have a critical responsibility to evaluate their respective student data to ensure continued student achievement. By sharing data and working together and/or with other research organizations or state agencies, they should benefit from each others' resources and expertise to gain a broader and deeper understanding of the effectiveness of the education system the state.

If the SEA does not have the capacity or staff skill set to conduct thorough research and evaluation activities, they can designate the higher education coordinating board or a postsecondary institution as an authorized evaluator of student data (P-20) in order to abate perceived FERPA obstacles.

3. *Establish a Memorandum of Understanding (MOU) between P-12 and higher education to share data.*

This MOU should specify what data elements will be shared with whom and by whom, the timeline for and frequency of sharing specific files, and the products (e.g., reports, datasets) that higher education and P-12 will produce for each other to enable their own research or reporting activities, as well as any limitations on how the data will be used and shared by each partner.

4. *Create a technical subcommittee to P-20 council.*

A technical subcommittee comprised of technology and information systems staff and program area representatives from both higher education and P-12 should be formed to work through the logistics of how to match and share the data, as well as develop common data definitions and discrepancies that will affect interpretation.

If public postsecondary institutions fed their student-level data into a centralized higher education student data system, it would be easier to conduct research, create standardized reports, reduce analytical burdens across institutions, and share data with the P-12 arena.

We frequently hear business people and postsecondary educators say that high school graduates come to them unprepared to succeed at the next level. Educators and administrators say that their graduates have passed the statewide assessments and met the state-defined diploma requirements. If both statements are true, where is the crack in the system?

In order to find the crack and figure out how to fix it, educators, policymakers and analysts need to work together to ask some probing questions and evaluate the data. Before the appropriate analyses can occur, however, we need to make sure that our educational data systems in both the P-12 and postsecondary arenas are robust enough to track students' educational and performance histories and can be connected across the education pipeline.

Without robust, student-level data systems, it will not get any easier to determine why the postsecondary community says that students come to them unprepared, even though students have received a CCR-driven diploma. Without the ability to know which classes a student took, what grades they received, and how well they did on high school exams and the ability to connect that information to how the student performed in higher education, the most we can say is that some students are ready and some students are not.

The conversation about why students are not ready for college and the workforce has gone beyond just an educational topic; it is concern for state and national economic development. The student-level longitudinal data systems are essential to informing educational policy and practices at both the local and statewide levels.

Nancy J. Smith *is deputy director of the Data Quality Campaign. She oversees the operations of the Data Quality Campaign (DQC) at the National Center for Educational Achievement (NCEA), including research, newsletters, publications and the annual survey of state data systems. Email: nancy@dataqualitycampaign.org*