

**Student Unit Record Systems and Postsecondary Accountability:
Exploiting Emerging Data Resources**

Peter Ewell
National Center for Higher Education Management Systems (NCHEMS)

peter@nchems.org

Draft: Please do not cite without permission from the author

Prepared for the American Enterprise Institute Conference, “Increasing Accountability in American Higher Education”
November 17, 2009

The collected papers for this conference can be found at www.aei.org/event/100134.

Draft: Please do not cite without permission from the author.

The past decade has seen unprecedented growth in the quantity and scope of electronic student record data available to states and systems in higher education. All but eight of the fifty states have such records and collective experience is growing about how to harness them to yield information useful for building policy and exercising appropriate accountability.¹ The purpose of this chapter is to review these developments by a) describing the characteristics and contents of extant systems, b) describing how they have been used to create indicators of institutional effectiveness (frequently by linking student data with parallel high school records and data about employment and earnings), c) proposing a number of standard statistical performance measures that can be calculated for colleges and universities for accountability and consumer information purposes, d) exploring the growing promise of linking databases across educational sectors for multiple states in a multi-state region and, e) discussing a number of commonly-encountered challenges associated with using such data effectively and how to address them.

Background

A Student Unit Record (SUR) database is a collection of electronic records about the enrollment behavior of every student enrolled in a given set of colleges and universities for each enrollment period (for example, semester, quarter, or year). SUR databases date from the early 1980s when states needed a way to systematically and accurately obtain student enrollment and credit hour counts each term in order to drive enrollment-based institutional resource allocation formulae. A second motivation, largely in the southeastern states, was to enable comparative tracking of graduation rates for African-American and Caucasian students, as mandated by a number of Consent Decrees on racial desegregation imposed by the federal courts system. Whatever their origin, such systems have grown steadily in scope and complexity. Originally

Draft: Please do not cite without permission from the author.

designed to only to cover enrollments in public institutions, at least twelve SURs now contain at least some information on independent and proprietary institutions—coverage made possible through the leverage induced by these institutions’ participation in lucrative state scholarship programs—and many more states indicate plans to do this. With regard to data coverage, these databases all contain the data elements needed to generate institutional statistics on such matters as annual or term-to-term retention, degree completion, time to degree, and inter-institutional transfer (within the boundaries of a given state) on a consistent and comparative basis. About half of them contain additional data on individual course enrollments and performance for every student each term. This allows the development of performance measures on such matters as placement in developmental work, the completion of developmental work (and the time needed for this to happen), and student performance in key “gate-keeper” courses like English Composition, College Algebra, and Calculus.

Postsecondary SURs are also being linked together on a more frequent basis, enabling student tracking across a multi-state region. At the same time, they are being matched with similar unit record enrollment databases in K-12 education and with Unemployment Insurance (UI) wage records. The first linkage allows K-12 systems and schools to evaluate their relative effectiveness in preparing graduates and former students for collegiate study by looking at whether these students need remediation when they enter college and how well they perform. Linking postsecondary SUR data to wage records, in turn, allows states and systems to construct additional collegiate performance measures such as the proportion of graduates and former students working in the field in which they took courses or earned degrees, as well as documenting any gains in income that might be attributable to postsecondary study.

Draft: Please do not cite without permission from the author.

One impetus for the development of linked systems of this kind has been the implementation of federal “No Child Left Behind” (NCLB) legislation, which suggests the growing desirability of tracking students longitudinally through their elementary and secondary schooling and into college. To help states implement such systems, the Department of Education has recently made substantial investments in developing the capabilities of SUR databases through State Longitudinal System Development grants to State Education Agencies. These grants will be supplemented with even larger investments later this fall, with preference given to systems that can link educational records in K-12 to postsecondary data systems and wage records.

Despite this steady progress, all has not gone completely smoothly. When the National Center for Education Statistics (NCES) failed to get Congress to go along with the establishment of a national unit record database for postsecondary education some five years ago, it was widely seen as a major setback. Despite its potential benefits, the proposal was successfully resisted on privacy grounds by much of the higher education community, led by independent college lobbying organizations. Given continued opposition, it appears that the proposal is not likely to be revived any time soon. Fortunately, the linking of state databases through regional consortia, together with the data holdings of the National Student Clearinghouse (NSC)—an organization that has compiled at least some data on enrollment and degree granting for over ninety percent of the nation’s college students—are beginning to fill the gap with respect to a national longitudinal tracking capability, though they will never fill it completely.

Draft: Please do not cite without permission from the author.

SURs and Institutional Performance Reporting for Accountability

When states began developing performance indicators and using them for accountability and performance funding in the late 1980s, graduation rates calculated using their SUR databases were the most popular measures in place.² But each state was essentially operating on their own with respect to how these measures were constructed. Graduation rates were given a good deal of definitional consistency in 1989, however, when Congress passed the Student Right-to-Know and Campus Security Act (SRK). This legislation required all institutions to report cohort-based graduation rates calculated on the basis of first-time full-time starters tracked out to six years for four-year institutions and three years for two-year institutions.³ These statistics are disaggregated by gender and by race/ethnicity. Reporting also included year-to-year retention and allowed institutions to report successful transfers to other postsecondary institutions “at a higher level” if such transfers could be documented. SRK reporting was later incorporated into the Integrated Postsecondary Education Data System (IPEDS) maintained by NCES in the form of an annual Graduation Rate Survey (GRS). The SRK/GRS definitions were quickly adopted by all states for their performance indicator systems and helped rationalize completion rate reporting on a national basis.

Student Right-to-Know, as its name implies, was also a milestone because it was explicitly structured as a “consumer protection” measure. Accordingly, institutions were required not only to calculate these statistics but also to display them prominently in catalogue and recruitment materials. From this point forward, federal officials moved increasingly toward an accountability approach centered on the provision of consumer information. This development reached its climax with the clamor for public reporting of student learning outcomes that arose during the deliberations of the Secretary’s Commission on the Future of

Draft: Please do not cite without permission from the author.

Higher Education (popularly known as the “Spellings Commission”) three years ago. A similar impetus was behind additional federally-mandated performance reporting which was then picked up by the states. Carl Perkins accountability legislation, for example, required vocational programs funded by this program to track student employment placement in the field in which they received training and the short-lived performance measures established by the 1992 Reauthorization of the Higher Education Act (HEA) included a measure of income gain in relation to tuition charges for vocational program graduates.⁴

The logic of this approach as a matter of public policy is simple: accurate information about collegiate outcomes like the probability of completing a program will influence student and parent choice about where to attend, and will consequently leverage institutions to improve their performance through the market. It remains unclear, though, if this logic is sound if directly applied, because the vast majority of students make decisions about where to attend college on the basis of location and price.⁵ This substantially diminishes the actual market effect of college rankings like those prepared annually by *U.S. News and World Report* (USNWR) as well. According to one study, only seventeen percent of U.S. college students consulted the USNWR rankings in making college choice decisions and, when they did, the information was used largely to verify a preexisting selection.⁶ But the same source indicates that the USNWR rankings have a strong effect on presidential behavior and institutional reputation, which may in turn influence legislative allocations. That is, such reporting devices may function better as direct accountability measures than indirect market levers.

It is interesting to compare the interest of the states—which, after all, have more financial stake in higher education—in following the federal logic. Forty-one of the forty-eight SUR databases surveyed by NCHEMS in 2006 used their SURs to calculate and report comparative

Draft: Please do not cite without permission from the author.

graduation rates using some variation of the SRK/GRS methodology. But their applications were directed more toward direct accountability reporting than toward guiding consumer choice. For example, the Kentucky Council on Postsecondary Education maintains a highly visible aid-to-student-choice utility on their website entitled “KnowHow2GOKy,” one click away from the agency’s home page, through which students can access a range of information about program offerings and affordability across Kentucky’s public and private institutions.⁷ But it contains no information about performance. This is not because the agency has none. Indeed, its staff includes some of the best practitioners in the country when it comes to using longitudinal data to conduct policy analyses. Detailed comparative data on retention, transfer, and completion rates are prominent in the agency’s accountability reports—also available on the Council’s website⁸—but casual student users would never know where to look for them. The website of the Minnesota State College and University System (MnSCU) has a similar utility for college choice that allows users to access a statistical “profile” for each institution in the system. Once again, though, MnSCU has produced an eye-catching set of performance measure displays that allow users to look at comparative persistence and completion rates, licensure examination pass rates, and in-field job placement rates by program by institution in a different place on the website labeled “accountability,” where prospective students are unlikely to go.⁹ There is nothing inherently misguided in taking this approach. It merely recognizes that the leaders of state university systems believe that performance measures calculated from data contained in their SURs are more useful in steering institutional behavior directly through accountability than indirectly by informing the market for student choice.

Draft: Please do not cite without permission from the author.

There are many good examples of how states and systems have used SUR data resources for accountability purposes—either through public reporting of such statistics as completion and job placement rates or by linking such measures to resource allocation systems for public institutions. One of the most effective among the latter is the so-called “Momentum Points” performance funding scheme recently put into place by the State Board of Community and Technical Colleges (SBCTC) in the state of Washington. This approach rewards institutions with additional funding to the extent that they move students through important intermediate outcomes or milestones on their way to a degree or certificate, information that is all contained in the system’s SUR database. Intermediate outcomes include successful completion of a General Education Diploma (GED) or literacy module, successfully passing out of developmental study, and the achievement of significant milestones in earning college credit. These intermediate outcomes are assigned particular point values that determine the amount of additional funding awarded to each college based on performance. This approach has the virtue of allowing colleges to demonstrate success at every point in the longitudinal enrollment pipeline—an important property for community colleges where, for a number of reasons, large proportions of newly-enrolled students do not earn associate degrees.

Another good example of a performance funding application is provided by the “Brain Gain” initiative of the Oklahoma Board of Regents.¹⁰ Under this approach, institutions are awarded additional funds for each student earning a degree beyond an established threshold set by the Board on the basis of an analysis of existing retention and completion information about other colleges and universities. This analysis creates a regression model to estimate the “expected” performance level for the institution in question, based on such variables as average admissions test scores, gender and race/ethnicity, and a range of enrollment behavior factors like

Draft: Please do not cite without permission from the author.

full-time/part-time attendance. Institutions are awarded performance funding dollars in proportion to the amount that they outperform their statistically expected performance with respect to graduates. Bonus dollars are provided for degrees earned by students drawn from at-risk backgrounds. Similar schemes in other states provide such “bonus points” for completions in designated fields of study deemed important for state economic development purposes.

Both these examples contain “best practice” features that other states would do well to look at. First, they are transparent and simple. The benchmark statistics are not indirect estimates of an underlying property or performance (like test scores) but rather the actual output itself, and the calculations required are easy to understand. Even the statistical model that drives the Oklahoma approach is easy for a lay person to grasp because it can be presented graphically (see the link provided in note 10). Second, the outcomes in question are linked demonstrably to the achievement of particular state policy objectives such as closing achievement gaps or addressing workforce needs with respect to literacy and skills. Third, both schemes only count a couple of things so they send a fairly straightforward message to institutions about what the state wants from them. The ill-fated performance funding approach adopted briefly by South Carolina about a decade ago, in contrast, rested upon thirty-seven separate performance measures and made it difficult for institutions to determine state priorities. Finally, these outcome measures are easy for a state with comprehensive SUR resources to calculate consistently and fairly under circumstances in which it is in the institutions’ interest to cheat. Tennessee’s long-standing performance funding system, in contrast, is based on institution-reported measures—a feature that requires state authorities to invest a good deal of time and effort checking up on the statistics reported each year.

Draft: Please do not cite without permission from the author.

Finally, there is a growing trend for SUR-based graduation and job placement rate statistics to be tied directly to a statewide strategic plan or “public agenda” for higher education in the form of an accountability dashboard. For example, Indiana and Kentucky have both advanced statewide plans framed around citizen benefits provided by higher education such as increased earnings or higher proportions of younger citizens with a postsecondary credential. Websites at the higher education agencies in both states allow users to click on each goal or question that comprise the public agenda to access graphic displays of the relevant performance indicators.

Linking Databases Across Sectors and States

Most of the examples of accountability measures using SUR data described in the previous section involve a single sector (postsecondary) in a single state. But a good deal more analytical and reporting power can be generated if data about multiple sectors in a multi-state region can be linked together. Florida, for example, has pooled data drawn from K-12 education, higher education, private and federal employment, military enlistments, tax and public service support, and corrections to produce a powerful picture of such topics as the impact of education on social mobility and burden, or the importance for college success of taking the right kinds of courses in high school.¹¹ NCHEMS, moreover, recently completed a pilot project involving four states—Ohio, Kentucky, Tennessee, and West Virginia—to examine patterns of student migration across state lines and between two-year and four-year colleges¹². All of these require a reliable mechanism for linking large datasets using common record identifiers.

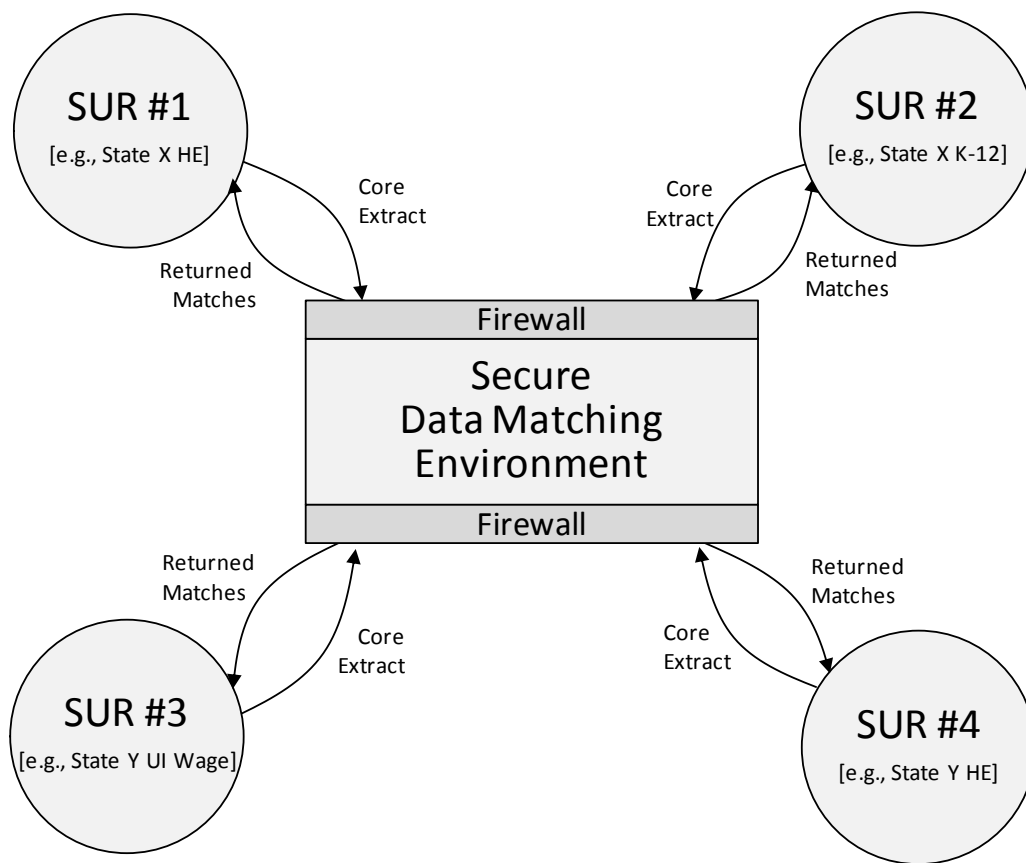
Extant experience suggests that the development of a multi-sector multi-state data-exchange capability is best conceptualized as a single centralized database containing data

Draft: Please do not cite without permission from the author.

extracted from multiple “source” SUR databases¹³. But the actual implementation of such an approach should be undertaken incrementally, and be advanced under the rubric of “data exchange” rather than “building a centralized database.” The basic concept for accomplishing such an exchange is quite straightforward, as portrayed by Figure 1. Its basic elements are as follows:

Figure 1

Basic Concept for Linking Data Bases



- Multiple SURs maintained independently for different sectors by participating state agencies. These will necessarily differ with respect to data contents, data structures, and operating environments.

Draft: Please do not cite without permission from the author.

- One or more unique identifiers that enable records corresponding to the same individual or entity in two or more databases to be identified. The most common data element used for these purposes is the Social Security Number (SSN), which is the only such key link that enables matching with UI wage-record data. Most state postsecondary SURs also continue to use the SSN, although they are moving away from this practice because of privacy concerns. State education departments are all developing unique identifiers for K-12 student records and these are increasingly being incorporated into postsecondary SURs as well.
- A secure data environment within which records can be matched to determine if the same student is present in multiple SURs. Selected unit record data supplied by individual participating states and drawn from their SURs enters this environment through a secure Web-based access portal and matching takes place behind a firewall. Alternatively, this environment can consist of a particular state agency’s database itself, so long as it has the requisite size and capabilities including secure Web-based access.
- A set of core data elements defined and coded in a standard format that constitute the substance of the records to be exchanged. Only a few such data elements will be needed to make this approach viable because most accountability measures can be constructed using only a few key pieces of data. These describe the “dependent variable” of interest in such exchanges—in essence whether and how a given student is enrolled in or has completed a degree at an institution covered by another SUR database, became employed in a particular industry at a given wage rate, and so on. “Independent variables” of interest, in contrast, will be much more extensive and varied but will for the most part remain within the individual SURs of participating states. Examples include

Draft: Please do not cite without permission from the author.

demographic descriptors, receipt of state scholarship funds, and participants in particular programs of interest to the state. After matched records are returned to participating states, they can be linked with any desired combination of “independent variables” of this kind within each state’s own SUR environment with no need for linkages or common data structures across states.

- An input protocol for participating states to submit core data elements. Under full-scale implementation, all participating state agencies submit records for all students with active records contained in their SURs on an annual basis. Under more limited forms of implementation such as bi-lateral exchanges, states might only submit records for the particular set of students they want to match.
- An output file that returns the results of the matching process to participating states. Ideally, this takes the form of a unique unit record that contains the core data elements, generated for each student enrollment event at each institution outside the original source agency. For students who enrolled in three different institutions outside the original source state, for example, three such records are generated and returned to the source state.
- One or more Memoranda of Understanding (MOU) that formally spells out the rules under which data can be accessed and used by each party. Some matches are governed by a single MOU, which is quite feasible to establish within the framework of state law. If more than one state is involved in the exchange, however, it is usually best to craft a series of bilateral arrangements between pairs of participating state agencies or between each agency and the third-party data manager because of differing legal environments in different states.

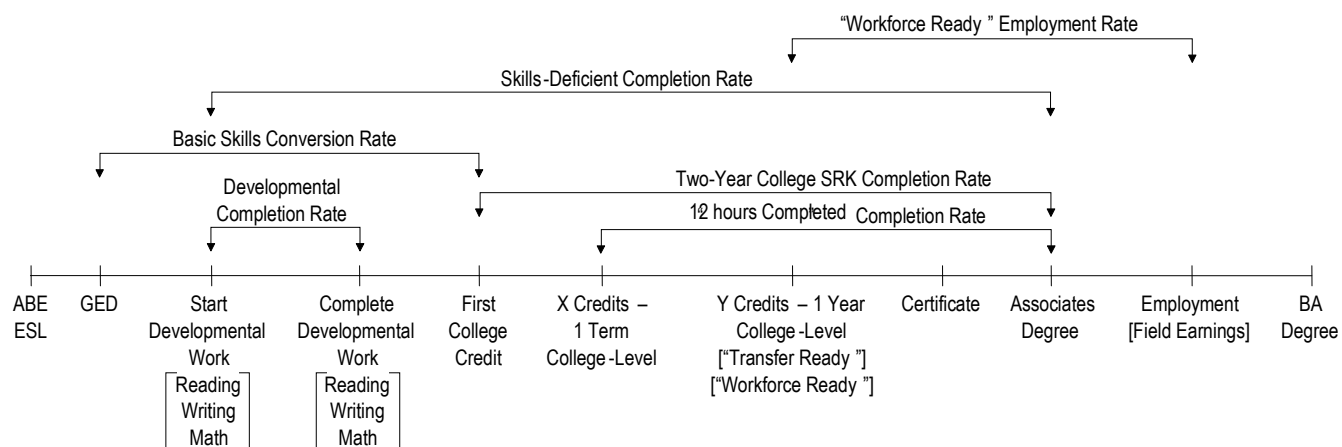
Draft: Please do not cite without permission from the author.

Adopting such an approach to data linking across sectors and agencies has the advantage of keeping the amount of data actually exchanged to a minimum and of allowing each participating state to construct its own performance measures using the augmented outcomes data generated by the matching process. But whatever approach is used, state agencies can benefit substantially with respect to the range of institutional performance that can be measured if the data contents of several SURs are tapped.

Statistics for Accountability

SUR databases can and have been used to produce a considerable array of performance measures that can be used for accountability reporting in higher education. With the exception of the graduation rate measure established by SRK and reported through the GRS, states and systems currently lack common definitions for these measures. This section describes a common set of benchmark statistics that all states and systems should eventually be able to produce by tapping and linking their SURs.¹⁴ These recommended performance measures are defined longitudinally within a given cohort in terms of the relationship between two or more events in a given student's enrollment history within a given period of time. For example, the "Two Year SRK Associates Degree Completion Rate" relates a given student's achievement of an Associates degree with his or her first credit enrollment within a specified time period. As another example, the "Skills Deficient Completion Rate" relates the point at which a given student placed below the college level in one or more basic skills with his or her attainment of a degree. An illustrative chart of "milestone events" of this kind is presented in Figure 2.

Figure 2
“Milestone Events” in a Student Enrollment Pathway



These performance measures recognize the fact that that such “milestone events” may occur in a different order for different students. For example, students may enroll for their first college-level credit at a point either before or after their enrollment in a developmental course. Similarly, students may transfer before or after they have earned a credential or achieved “transfer ready” status. Given that they will be calculated from statewide databases, these measures also recognize the fact that these events may take place at different institutions. Each performance measure is calculated independently in this manner within a given cohort.

These measures can be presented in two groups depending upon the complexity of the data needed to construct them. Measures in the first set are fairly straightforward and depend largely upon the presence of basic enrollment and completions data in state or system SUR databases. The exceptions are proposed measures on developmental education, which may require some states to produce flags for enrollment and completion of developmental education. For institutions that have the requisite data, these flags can be constructed by tapping transcript-level detail for developmental courses; otherwise, they must be calculated and reported directly

Draft: Please do not cite without permission from the author.

to the state agency by each institution. None of the measures in the first set address non-credit instruction and none of them require linkages to other databases.

- Completion Rate (Credentials). The proportion of students in a cohort who earn a credential, tracked from the point at which they enroll for the first time in instruction that will count toward a credential, calculated at annual intervals out to six years.¹⁵ Credentials include degrees, certificates, diplomas or any other formal award. Students placed in developmental work are considered to have reached this start point if they are enrolled in a course of study that leads to a credential.
- Completion Rate (Degrees). The proportion of students in a cohort who earn an Associate or Bachelors Degree, tracked from the point at which they enroll for the first time for credit that will count toward a degree, reported at annual intervals out to ten years.¹⁶ Students placed in developmental work are considered to have reached this start point if they are enrolled in the appropriate course of study.
- Annual Persistence Rate. The proportion of students in a cohort who remain enrolled in a program leading to a credential or a degree at successive one-year intervals after first enrolling in instruction that will count toward a degree or credential, tracked annually out to six years.
- Developmental Success Rate I. The proportion of students tracked from the point they begin developmental work who successfully complete developmental work, reported annually out to five years. Reporting should distinguish: a) beginning developmental work in *any* field to the completion of *all* required developmental work and b) be broken down separately to reflect beginning and successful exit in reading, writing, and

Draft: Please do not cite without permission from the author.

mathematics. An alternative for states that have test scores or placement levels is to use students identified as deficient as the denominator for this measure.

- Developmental Success Rate II. The proportion of students tracked from the point they begin developmental work that successfully complete developmental work and enroll for credit at the college level, reported annually out to five years. Basic reporting should reflect *any* college-level enrollment after the successful completion of developmental work.¹⁷
- Transfer Rate. The proportion of students in a cohort, tracked from the point they first entered a degree program, that subsequently enroll in a degree-program at another degree granting institution,¹⁸ tracked annually out to ten years. A refinement of this measure would distinguish those who transferred to a higher level of study at a four-year institution.

Performance measures in the second set depend on the ability to access transcript-level information and to link postsecondary enrollment records to the UI wage record. Only a few states or systems are now able to generate these measures, either because of current limitations in the available data or because generating them would require considerable programming or analytical effort.

- Twelve-Hours Earned Rate. The proportion of students in a cohort who complete one term (that is, 12 Student Credit Hours [SCH] or equivalent) of college-level work, tracked from the point at which they enroll for the first time in instruction that leads to a credential, calculated at annual intervals out to six years.

Draft: Please do not cite without permission from the author.

- Developmental Success Rate III. The proportion of students tracked from the point they begin developmental work that successfully complete developmental work and complete one term (that is, 12 SCH) of college-level work, reported annually out to five years.
- Developmental Success Rate IV. The proportion of students tracked from the point they begin developmental work in Writing that successfully complete the college-level course in English Composition (or equivalent) and the proportion that begin developmental work in mathematics that successfully complete a first college-level mathematics course, reported annually out to five years.
- “Transfer Ready Status” Achievement Rate. The proportion of students in a cohort who achieve the enrollment milestone of “Transfer Ready” status, tracked from the point at which they enroll for the first time in instruction that leads to a credential, calculated at annual intervals out to ten years. “Transfer Ready Status” is achieved when the student has a) completed one year (that is, 30 SCH) of college-level work; b) has passed or placed out of all developmental work and; c) has completed English Composition, a college-level math course, and one college-level course in each basic discipline cluster (science, social science, and humanities).
- “Workforce Ready Status” Achievement Rate. The proportion of students in a cohort who achieve the enrollment milestone of “Workforce Ready” status, tracked from the point at which they enroll for the first time in instruction that leads to a credential, calculated at annual intervals out to ten years. “Workforce Ready Status” is achieved when the student has completed one year (that is, 30 SCH) of college-level work at least half of which consists of vocational courses related to a particular occupational field.

Draft: Please do not cite without permission from the author.

- Non-Credit Conversion Rate: The proportion of students who entered the cohort as non-credit students (that is, their “first contact” with the institution was enrollment in a non-credit offering) who subsequently enroll in instruction that will count toward a credential, reported annually out to five years. Reporting should distinguish between entering a degree program and entering a credential program. Reporting should also distinguish between different non-credit tracks including GED, Adult Basic Education (ABE), English as a Second Language (ESL), etc.
- Employment Rate. The proportion of students in a cohort, tracked from the point they first enrolled in instruction leading to a credential, who are employed in the third UI-Wage Record reporting quarter after having a) completed a credential or b) their last known enrollment in the program. Reporting should distinguish students who completed a credential from those who did not complete a credential. A subset of these rates is “in-field employment rate,” which means that the former student is employed in the field for which he or she received training.
- Post-Enrollment Earnings. The annualized earnings of a student in the third UI-Wage reporting quarter after having completed a credential or their last known enrollment in the program.
- Return on Investment. The ratio between the average total cost to the student to complete a given degree program and the average annual wage income earned by graduates of that program.

Mid-point measures such as the developmental success rates, achievement of credit-earned milestones, and the achievement of “transfer-ready” or “workplace-ready” status should be priorities for future development because of their importance in providing indicators of success

Draft: Please do not cite without permission from the author.

short of attaining a formal credential. Such measures will be especially important for community colleges.

Measures of both kinds can at minimum be calculated for each institution in the state or system as a whole. But performance benchmarking will be considerably more informative if these measures are broken down by a set of agreed-upon subpopulations. The following subpopulations are among the most important:

- Gender. Standard IPEDS.
- Race/Ethnicity. Standard IPEDS.
- Age. The following breakdown is typical:
 - Traditional Age (under 22 years old).
 - Younger Working Age Adult (22-35).
 - Older Working Age Adult (36-64).
 - Older Student (65 and above).
- Part-Time Status. IPEDS definition for the first term of enrollment.
- Financial Status. This is an increasingly important policy variable for disaggregation, but many states do not have the required data. The best proxy measure is whether or not the student is receiving a Pell grant.
- Transfer Status. Entering as first-time in college student or as transfer student (further breakdowns of entering transfer students are recommended at term-length intervals (that is, 15 SCH) of credits transferred).

Many other subpopulations of interest can be defined, but the above breakdowns should be sufficient for benchmarking and accounting for institutional and program performance.

Draft: Please do not cite without permission from the author.

Statistics like these can be used for many research and policy purposes, but it is useful to briefly review how states have used them specifically for accountability. Three main approaches have typically been used:

- Accountability Reporting. Virtually all states with SURs use them to generate basic institution-level statistics on retention and degree completion, generally using the SRK method.¹⁹
- Performance Funding. As already described for Oklahoma and the Washington SBCTC, some states have allocated additional funds to institutions that perform well on various outcome measures, if these measures can be constructed from data elements contained in their SURs. A straightforward approach here, so far not adopted by any state, would be to base some proportion of the allocation—now based only on credits registered for at an early census date—on the basis of credits completed successfully instead.
- Consumer Information. Most state systems have a utility on their websites designed to help prospective students look at public institutions to aid them in making a choice, but none of these provide any information about performance. As the MnSCU and Kentucky examples discussed earlier illustrate, though, the agencies clearly have and report this information. A simple link to it in the student choice utility could remedy this situation.

Challenges Associated with Developing and Using Such Measures

Growing state and system experience using SUR data in accountability reporting reveals a number of common challenges and dilemmas. These occur with sufficient frequency that they need to be discussed in any treatment of such applications. For each of these challenges, the

Draft: Please do not cite without permission from the author.

nature of the problem must first be explained, followed by some suggestions for addressing the difficulty.

Defining the Base Population. Virtually all of the performance measures discussed in the previous section are based on a population ratio of some kind. For example, the SRK/GRS graduation rate is defined as the proportion of first-time full-time students entering the institution in a given fall term that have completed a degree within a designated period. The trick in constructing such statistics is to set the denominator fairly. For example, community college leaders have long complained that the GRS rate under-reports their success with students because a large (and unknown) share of entering students do not intend to complete a degree in the first place. They therefore argue that these students should be backed out of the denominator before the completion percentage is calculated. Similarly, higher education officials in Utah successfully modified the originally-proposed SRK calculation procedures to enable them to remove from the calculation's denominator students who delayed completing their programs because they undertook a religious mission. A similar rationale was used for excluding students called up for active military service from the SRK denominator, as well as those who die.

The justification for making such adjustments rests on the principle that a good accountability statistic ought not to hold institutions responsible for things that are not under their control. The difficulty lies in determining just how far to go because each such adjustment distorts the true picture of performance, sometimes quite significantly. For example, NCES currently reports a three-year graduation rate for community college starters of just under twenty-two percent using the standard calculation. But an NCHEMS study using data drawn from the National Student Clearinghouse (NSC) yielded only about half this proportion of degree earners in this time period. The difference lies in the fact that the denominator used in the

Draft: Please do not cite without permission from the author.

NCHEMS study was almost twice as big as that used for the GRS calculation because no assumptions were made about who was “degree-seeking.” This assumption is frequently made by registrars at community colleges when calculating the GRS rates that they report to NCES. Which rate is more appropriate? On the one hand, backing out supposed non-degree students is “fair” because they are presumed to be just that, and are thus unlikely to complete. On the other hand, the excluded students were enrolled at least half time in credit-bearing courses that count toward a degree and there is no assurance that they might later decide to earn a degree using these credits. The latter, although it admittedly holds the institution responsible for something beyond its control, moreover, is the actual performance level attained.

One way to address this difficulty is to report more than one number. The Florida Community College System uses an accountability measure based on the associate degree completion rates of only those students in a GRS cohort who earned eighteen or more hours of credit.²⁰ But it also uses the standard first-time full-time GRS rate for research purposes and has reported that number as well.²¹ For the 2001 entering cohort tracked over four years, the first measure yields a rate of just over 50% while the comparable rate for the standard measure is only about half this level. The logic behind the modified rate is that students who have completed a significant amount of credit are more likely to be degree-seeking than those who only enrolled for a course or two and ceased attending when their goals for attendance were met.

Apples and Oranges. A second related challenge associated with graduation rates and similar performance statistics is that most of the variation across institutions can be accounted for by differences in institutional characteristics that have nothing to do with performance. The most telling example is entering student ability as reflected in admissions selectivity, which accounts for by far the largest share of explained variation in graduation rates across institutions.

Draft: Please do not cite without permission from the author.

Highly selective institutions routinely graduate more than ninety percent of an entering class, while open admissions institutions can fall to below twenty percent. Clearly it is not appropriate to directly compare these measures. Indeed, holding the latter institutions more accountable for their low performance may induce them to become more selective, contrary to the policy interests of the state.

To address this problem, states and systems have adopted a range of measures to adjust for factors like these. One of them, illustrated by the Oklahoma Board of Regents performance funding scheme described earlier, is to make statistical adjustments that remove the unrelated variation from the outcome measure. This is done in Oklahoma by constructing a regression model containing all of the institutional covariates to predict graduation rates for each institution. Each institution is then assessed on the extent to which its actual graduation rates are above or below its predicted performance. Another way to address this issue is to disaggregate outcomes for different student populations if student characteristics are known to be related to graduation. For example, Texas reports graduation rates for public institutions broken down by gender and race/ethnicity. The primary reason for doing this is the fact that reducing performance gaps across population groups is a prominent state goal for higher education. But because blacks and Hispanics typically have lower graduation rates than whites, it also helps sort out the fact that overall institutional graduation rates can be markedly influenced by any given institution's racial/ethnic mix. This feature also enables potential students to determine with more precision what their own chances of a successful outcome at a given institution may be. A final approach used by some states is to set individually tailored "stretch goals" for each institution on the basis of its operating environment. For example, one of the indicators used in the Washington SBCTC's original performance funding scheme was in-field job placement rates for program

Draft: Please do not cite without permission from the author.

completers. Because job markets vary across different regions of the state, SBCTC adjusted benchmark rates on a case-by-case basis to take away any advantages an institution might enjoy by being located in a region in which it was relatively easy to get an in-field job.

Student Mobility. A significant limitation to the use of SUR data for accountability is the fact that such data only cover students enrolled at institutions within a given state. If a student ceases enrollment and transfers to an institution in another state, it is not possible to identify the transfer using in-state SUR data alone. And this is far from an infrequent occurrence. Federal studies have indicated for years that of the seventy percent of bachelors' degree earners who attended more than one institution, more than forty percent of them crossed state borders at some point in their enrollment histories.²² The information on institutional transfer performance is therefore incomplete, a property unwelcome in an accountability statistic. This is why data sharing across states in a larger geographic region can be so important. Because most border crossing is local, a fair share of the missing data can be captured in the postsecondary SURs of neighboring states.²³ But there are other ways to capture this missing information as well, using the data resources of the National Student Clearinghouse (NSC).

To demonstrate the potential of NSC data, NCHEMS recently completed a pilot project in partnership with the National Center for Policy in Higher Education.²⁴ The intent of the demonstration was to create and assess the quality of state-level graduation rate statistics using NSC data alone. To do this, NCHEMS researchers defined a set of starting cohorts within the NSC database using SRK/GRS definitions and tracked these students through six years of history to determine when and where they earned degrees. Completion rate statistics were then calculated for four-year and two-year starters including the proportions earning a degree from a)

Draft: Please do not cite without permission from the author.

the institution at which the student first enrolled, b) an institution located in the same state and, c) an institution located anywhere else.

Findings revealed that on a national basis, six-year graduation rates for four-year starters are boosted just over six percentage points if all institutions in a state are included, and an additional four percentage points if institutions in all states are included. These patterns vary substantially by state. For example, states like Arkansas and Texas show relatively low out-of-state graduation rates, while states like North and South Dakota show relatively high rates. These variations are expected because of differences in general geographic mobility and differential tuition policies across states. Substantial variations are also apparent with respect to additional in-state completions. States like Georgia and Wisconsin show relatively high rates of in-state transfer completion, while states like Delaware and Vermont are comparatively low. These variations are due to a combination of geographic location and the structure of a state's public higher education system including the mix of institutional types and the presence of statewide articulation agreements.

Using a similar approach, growing numbers of states and systems are tapping NSC data to locate missing transfer students, thereby improving the quality of their longitudinal SUR-based accountability measures. The NSC is also under contract to the American Association of State Colleges and Universities (AASCU) and the Association of Public Land Grant Universities (APLU) to complete the graduation rate reporting portion of the performance "report card" developed as the centerpiece of the Voluntary System of Accountability (VSA) recently launched by these two organizations.

Points of Comparison. Statistics acquire meaning only when they are placed alongside meaningful benchmarks. To be useful as accountability measures, points of comparison are

Draft: Please do not cite without permission from the author.

needed. Three such benchmarks are typically used in performance statistics like graduation rates. The use of fixed criteria implies that there is a recognized standard that can be applied to everyone indiscriminately. This is rarely used by states and systems in reporting longitudinal performance measures like graduation or job placement rates because no such standards currently exist. The closest approximations are the cut values used to define performance zones (red, yellow, and green) for student success and job placement indicators on the MnSCU Accountability Dashboard.

Using normed or relative statistics implies a competitive situation in which absolute performance is valued less than how well one is doing compared to others. This is the most common approach to establishing points of comparison for performance indicators among states and systems. Usually the arithmetic mean across all institutions in the state or system is used as the point of comparison but some agencies report quartiles or whether or not a given institution's performance is above or below one standard deviation of the mean. For example, the performance website for the North Carolina Community College System arrays progression measures in tables so that users can compare the performance of any given college against others individually, or against the state mean.²⁵

Finally, measuring change over time implies a world free of both standards and competition, where "success" simply means steady improvement. The current Washington SBCTC "Momentum Points" performance funding scheme described earlier incorporates this feature in that each time an institution exceeds its past best performance, the target is recalibrated so that it needs to achieve more in order to get rewarded.

Random Variation. Performance indicators such as graduation rates, like any real-world statistic, always vary slightly from year to year. These variations are for the most part

Draft: Please do not cite without permission from the author.

independent of actual performance but if the reported statistics are part of a public accountability system or are intended to serve as an aid to consumer choice, they may be mistaken by observers for changes that have substantive meaning. To achieve greater stability, some states or systems reporting such statistics (for example, the MnSCU Accountability Dashboard and the North Carolina Community College System) use a three year rolling average. Others, like the Indiana Commission on Higher Education, find ways to show the limits of variation through box plots or whisker charts. Very few go to the length of actually testing differences for statistical significance, but good practice does suggest reporting the numbers of cases associated with any reported statistic so that readers can roughly assess significant difference. This is done in the great majority of cases. For example, the University System of Georgia allows users of its website to generate comparative graduation rates “on the fly” from an underlying database by varying parameters in an interactive user interface, and both numbers and percentages are shown on the resulting tables.

Analytical Capacity. A final challenge faced by growing numbers of states and systems under today’s economic circumstances is retaining the trained staff capable of constructing longitudinal databases from base SUR records and creating the performance measures needed to ground postsecondary accountability. Keeping these staff members is generally considered less important during bad budget times than retaining staff assigned to core functions like resource allocation and payroll. In fact, the NCHEMS survey of state SUR databases in higher education revealed diminished analytical staff capacity to be one of the top concerns of database administrators. This is because it is difficult to harness SUR data resources to conduct relevant policy research and create the kinds of accountability measures described in this chapter. There

Draft: Please do not cite without permission from the author.

is little point in going to the trouble to collect these data if all that will happen is they remain in electronic form in the agency's archives.

One promising way to address this condition is to recruit university scholars to undertake the necessary database and analytical work. Economic and educational researchers from Columbia, Harvard, the University of Pennsylvania, and Princeton have worked on SURs from Ohio, Florida, and Washington at the invitation of state and system administrators. In another instance, researchers at the Community College Research Center at Teachers College Columbia University recently helped the Washington SBCTC construct the "Momentum Points" performance funding scheme described earlier. The researchers gain access to a powerful pool of data that they can use to answer scholarly questions on topics like labor market supply and the role of education in earnings. In return, state database administrators can ask the researchers to undertake work on behalf of the state. This can be done effectively only if the requisite data access agreements are in place, with clear understanding of the details of the *quid pro quo* between the researchers and the agency.²⁶

Conclusion

The federal funds currently being made available to states to build new K-12 longitudinal student tracking capacity and to enhance existing capacity in postsecondary education will undoubtedly yield significant progress in higher education state and system capacities to produce powerful accountability reporting statistics. Reinforced by the increasing availability of additional unit record data from the National Student Clearinghouse on a national basis, together with the expansion of NSC data holdings into K-12 education now under way with the support of the Bill and Melinda Gates Foundation, we are rapidly witnessing the development of a seamless

Draft: Please do not cite without permission from the author.

national data resource. This means that whatever the fate of any future NCES proposal for a national student unit record system in higher education, those responsible for running state college and university systems will be able to produce increasingly powerful and useful accountability measures to steer institutional behavior. The result, hopefully, will be increases in levels of postsecondary attainment and more effective student learning.

¹ Peter T. Ewell and Marianne Boeke, *Critical Connections: Linking States' Unit Record Systems to Track Student Progress* (Indianapolis, IN: Lumina Foundation for Education, 2007). This comprehensive fifty-state survey of state SURs is the basis for the descriptions of the characteristics and capacities of these systems made in the three following paragraphs.

² Joseph C. Burke and Henrik Minnassians, *Performance Reporting: "Real" Accountability or Accountability "Lite"?* (Albany, NY: Rockefeller Institute of Government, State University of New York, 2003).

³ These tracking periods were established to represent 150 percent of the advertised catalogue length of the degree program.

⁴ The latter statistic was also included in the measures proposed for "triggered" institutions under the short-lived State Postsecondary Review Entity (SPRE) programs in many states. See State Higher Education Executive Officers, *Developing Quantitative Performance Standards for SPRE Reviews: A Technical Assistance Primer* (Denver CO: State Higher Education Executive Officers (SHEEO), 1995).

⁵ Robert Zemsky, "The Dog that Doesn't Bark: Why Markets neither Limit Prices nor Promote Educational Quality", in *Achieving Accountability in Higher Education*. J.C. Burke and Associates (New York: Jossey-Bass, pp. 275-295, 2005).

⁶ Patricia M. McDonough, Anthony Lising Antonio, Mary Beth Walpole and Lenor Perez, *College Rankings: Who Uses Them and with What Impact?* (Los Angeles, CA: Graduate School of Education and Information Studies, The University of California at Los Angeles (UCLA), 1997).

⁷ <http://cpe.ky.gov/forstudents/>.

⁸ <http://apps.cpe.ky.gov/reports/AccountabilityReport0708final.pdf>.

⁹ <http://www.mnscu.edu/board/accountability/index.html>.

¹⁰ <http://www.okhighered.org/studies-reports/brain-gain/braingain2008-update-screen.pdf>.

¹¹ See <http://www.fldoe.org/cc/OSAS/DataTrendsResearch/DT33.pdf>.

¹² NCHEMS found that increases in cross-state completions were significantly higher for institutions located near state borders than at other institutions. See National Center for Higher Education Management Systems, "Tracking Postsecondary Students Across State Lines: Results of a Multi-State Data Exchange Initiative", 2008, available online at <http://www.nchems.org/c2sp/documents/ResultsofMulti-StateDataExchange.pdf>.

¹³ National Center for Higher Education Management Systems, *Linking Student Enrollment Records Across State Lines: Extending the Concept*, 2004, available online at <http://www.nchems.org/c2sp/documents/Multi-StateDataExchangeConceptPaper.pdf>.

¹⁴ These measures were initially developed through a state data initiative undertaken in 2005 under the auspices of the Achieving the Dream (AtD) project funded by the Lumina Foundation for Education and the Community College Bridges to Opportunity project funded by the Ford Foundation. See Community College Research Center, *Community College Data and Performance Measurement Toolkit*, (New York, NY: Teachers College, Columbia University, 2008).

¹⁵ Because of system conversions or similar limitations in historical data, some states or systems will not be able to track students to the originally-proposed ten year mark, but all who can should be encouraged to do so.

¹⁶ Note that this is a subset of the previous rate.

¹⁷ Includes all of the comments associated with the above measure about establishing the beginning of developmental work.

¹⁸ Only aggregate information on transfers is available in many states. This may limit the subpopulation breakdowns that can be provided for this measure although all participating states should be able to generate the basic statistic for the state as a whole. Also, even with states that have unit-record data on transfers, only transfers to public institutions are typically included. Finally, many states can supplement their transfer records with data obtained from the National Student Clearinghouse.

¹⁹ See Ewell and Boeke, op. cit.

²⁰ See <http://www.fldoe.org/cc/OSAS/FastFacts/pdf/2009-03.pdf>.

²¹ Data obtained from the Florida Community College CPT Progression File System.

²² Clifford Adelman, *The Toolbox Revisited: Paths to Degree Completion from High School Through College* (Washington, DC: Office of Vocational and Adult Education, U.S. Department of Education, 2006).

²³ NCHEMS found that increases in cross-state completions were significantly higher for institutions located near state borders than at other institutions. See National Center for Higher Education Management Systems, "Tracking Postsecondary Students Across State Lines: Results of a Multi-State Data Exchange Initiative", 2008, available online at <http://www.nchems.org/c2sp/documents/ResultsofMulti-StateDataExchange.pdf>.

²⁴ National Center for Public Policy in Higher Education, *State Level completion and Transfer Rates: Harnessing a New National Resource*, (San Jose CA: National Center for Public Policy in Higher Education, 2009).

²⁵ <http://www.nccommunitycolleges.edu/Publications/docs/Publications/csf2009.pdf>.

²⁶ National Center for Higher Education Management Systems, "Harnessing the Potential for Research of Existing Student Records Databases: An Action Agenda", 2005, available online at <http://www.nchems.org/c2sp/Promoting%20Researcher%20Access%20to%20State%20SUR%20Data%20Resources.pdf>.