Achieving the Dream™

Community Colleges Count

Basics of Longitudinal Cohort Analysis

PRINCIPLES AND PRACTICES OF STUDENT SUCCESS

Rick Voorhees and John Lee

Longitudinal cohort analysis is a powerful tool for helping colleges understand student performance. It involves tracking students as a group or cohort over a specified period of time. The results allow administrators, faculty, and staff to identify groups of students who are succeeding or falling behind and the points in the educational pipeline where they may falter. This guide is designed to give colleges an overview of longitudinal cohort analysis and how it can be used to improve student success.

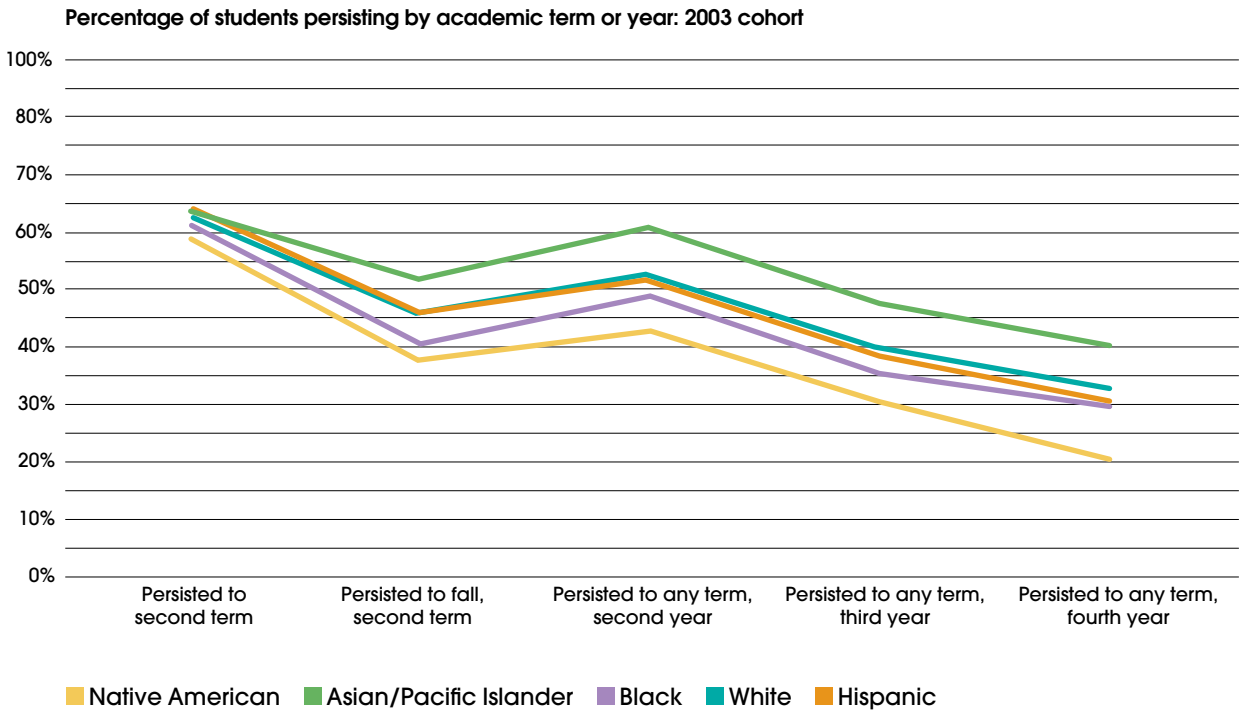
Defining longitudinal cohort analysis

A cohort is a group of students who enter a college or a program at the same time. One commonly followed cohort consists of first-time college students — those who enter at a particular point in time, with no previous college credits. Tracking so-called "first-time-in-college" students allows colleges to better understand the characteristics of their entering students — whether they are full- or part-time, referred to developmental education, or enrolled in an academic or vocational track — as well as their gender, age, race and ethnicity, income, and placement test scores. In later enrollment periods, cohort tracking allows colleges to examine the progress of students who begin at the same "starting line." For example, a college might want to know how many credits students attempt and complete, what grades they receive, and whether they succeed in important "gatekeeper[1]" courses, such as the first college-level math and English courses.

As data are added for each successive enrollment period, it is possible to identify when students change enrollment status, achieve certain educational milestones, or graduate. This information can provide a baseline for evaluating the effects of the college's efforts to increase student success. A new cohort can be started each year, making it possible to determine if measures of student progression and success improve over time.

Colleges may disaggregate data on a cohort to create subgroups defined by student characteristics, such as race, gender, or whether or not students in the subgroup were referred to developmental instruction. This allows colleges to see the gaps in progression and achievement among different student subgroups. For example, Figure 1 shows the persistence of students by race and ethnicity during a four-year period, drawn from data reported by Achieving the Dream colleges. In this example, Asian/Pacific Islander students persist at the highest rate, while Native American students have the lowest persistence rates.

Figure 1



Colleges generally obtain the data needed for longitudinal cohort analysis from the administrative records created through the registration and enrollment processes. Some colleges add additional data elements obtained when students apply, such as their objectives or intent for enrolling, whether they have a high school diploma or GED, and whether they have previous college credits. Additional information may be collected during the course of a student's enrollment, as explained below.

Cohort analysis will help your college

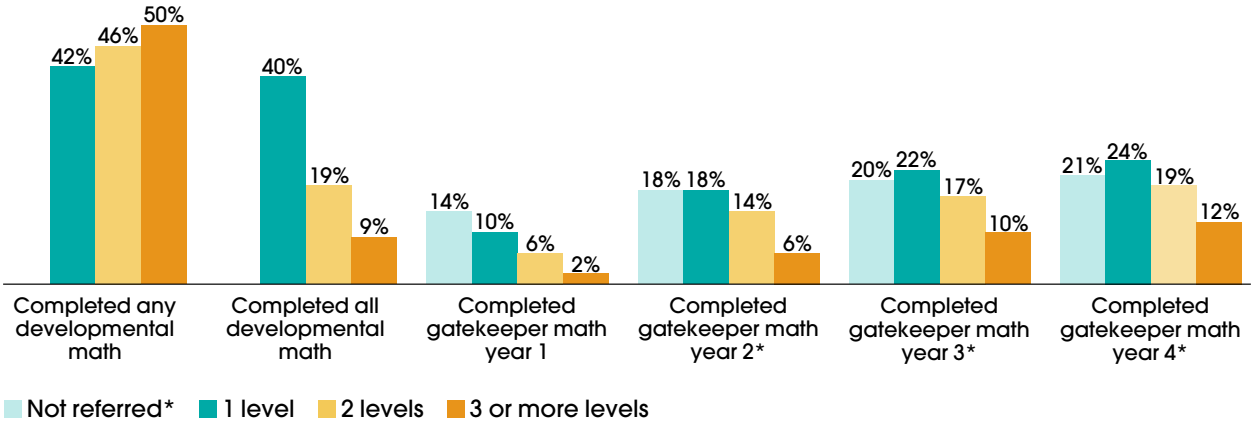
Achieving the Dream colleges use cohort analysis to answer the following key questions related to the initiative's five main performance measures for cohorts of entering students. These datapoints allow institutions to see the pattern of student progression for particular groups of students:

1. What percentage of students complete developmental coursework and advance to college-level courses?
2. What percentage of students enroll in and complete "gatekeeper" courses, the first-level college courses that are prerequisites for degrees and generally have high drop-out/failure rates?
3. What percentage of students successfully complete their courses (with a grade of C or higher)?
4. What percentage of incoming freshmen re-enroll for a second semester? What percentage re-enroll for a second year?
5. What percentage of students earn certificates and/or degrees?

Cohort analysis can help colleges improve their understanding of the points at which any given number of students fail or succeed and bring information on actual patterns of student progression and success to discussions that sometimes reflect personal biases or desired outcomes, rather than reality. This can help focus efforts to improve student success on areas where large-scale gains in achievement can be attained. For example, cohort analysis may suggest that if students can

Figure 2

Percentage of students completing all of the required developmental math sequence in two years, and percent completeing gatekeeper math by year: 2003 cohort



* Some students not referred to developmental education enrolled in developmental courses; however, they are not assigned a referral level and, as such, cannot be included in the developmental math measures reported here. Reported percentages of students completeling gatekeeper math is cumulative.

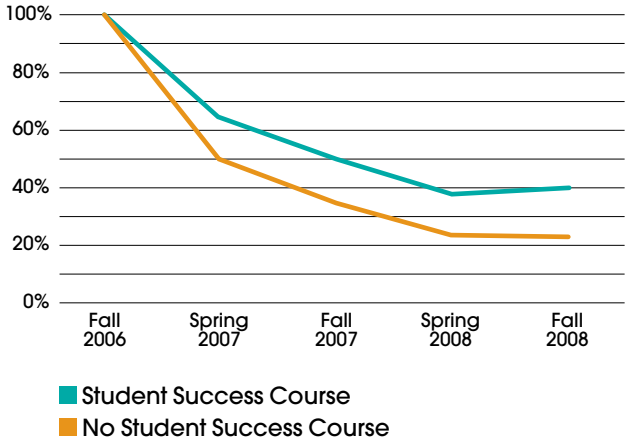
complete their developmental education courses, their chances of graduating are similar to, or better than, those of students who did not take developmental education. Such an analysis might reveal the need both to require developmental education and to establish improved completion as a priority at an institution.

Figure 2 illustrates the percentage of Achieving the Dream students completing the developmental math sequences to which they were referred within two years and the percentage completing gatekeeper math classes by the year they completed. The figure shows that, regardless of whether or not they were referred to developmental education, most students who completed college-level or gatekeeper math did so four years after enrolling, indicating that students may put off taking math as long as possible. The figure also shows that students who were referred to the first level of developmental education were just as likely to complete gatekeeper math over the course of four years as were students who were not referred to developmental education.

Providing such data about developmental education students to faculty and administrators can provide a more complete picture of the progress these students are making and the potential challenges the staff must address to promote success for students who are referred

Figure 3

Percentage of students persisting by enrollment in student success course by term: 2006 cohort



to developmental instruction. The data in Figure 2 might motivate colleges to find ways to encourage students to take the courses to which they are referred, and to move through their sequences in a more timely way.

Longitudinal data can be used to evaluate specific programs and to measure college-wide progress in achieving educational objectives. Figure 3 is a hypothetical example of cohort tracking applied to an evaluation of the impact of a student success course on persistence. This figure shows that students who take such a course persist at a much higher rate that those who didn't take the course. These results could be used as a starting point for institutional discussions and would ideally be supplemented with analyses and data drawn from other sources, such as interviews with instructors and student focus groups, to help determine the effect of the intervention. While longitudinal analysis can help indicate whether or not a given intervention is associated with improvement in student success, it cannot tell why this is the case and how the intervention can be improved. The "why" and the "how" can only be obtained through additional quantitative and qualitative research. Further, these data should be disaggregated by race/ethnicity, gender, and low-income status to determine if there are differences in outcomes among various student groups.

A college that has successfully used cohort tracking

DATA-BASED DECISION-MAKING: Patrick Henry Community College, Martinsville, VA

Patrick Henry Community College (PHCC) used data from pre–Achieving the Dream entering student cohorts as the foundation for its initial analysis of students' needs and challenges. The college linked Achieving the Dream student cohort data with other data, such as information from student surveys and student placement tests, to conduct a deeper analysis of patterns of student success.

With the combined database, PHCC created a "persistence/success model" which it uses to identify characteristics of students who are likely to drop out, and to develop appropriate interventions that will lead to persistence and ultimately improve the odds of success for these students. The model takes into account a combination of social (e.g., economic status), demographic (e.g., race, gender, age), psychological (e.g., student motivation and attitudinal measures), and academic (e.g., math placement scores and cumulative GPA) factors that have been shown to contribute to student achievement. These variables were used to develop a risk score for all first-time, degree-seeking students in future fall cohorts (based on all variables that could be assessed in the first semester). The outcome for the first semester of PHCC's 2007–08 Achieving the Dream cohort showed that the model was more accurate at predicting student persistence than any single piece of the information the college collects from students when they first enroll.

The model allowed the college to identify entering students who had particularly high-risk profiles and to assist these students with case management–based "intrusive mentoring" by specially trained faculty and staff. Students who participated in the mentoring program in fall 2007 showed slightly higher fall-to-spring persistence rates — about 5 percentage points better than the rates for a comparison group who did not receive the mentoring assistance. The college plans to continue monitoring the impact of the case management advising/mentoring model and make improvements accordingly.

Data that your college needs to collect

The following describes the types of data the college needs to collect for longitudinal cohort tracking.

Student characteristics data. Cohort tracking begins by assembling data elements reflecting the characteristics of students in the cohort. First, it is important to include data elements that will allow student records to be linked across terms (e.g., name, student ID). Selecting which student characteristics to include in the cohort data collection depends on data availability and what elements are important for the analysis and presentation of the results. Student characteristics could include such elements as gender, age, race/ethnicity, ESL status, previous education, purpose for enrolling, receipt of student aid in the first enrollment period, full- or part-time enrollment, dependency and marital status, income level, placement test scores, and address or ZIP code.

These student characteristics will provide the basis for disaggregating results to compare the outcomes of various student subgroups. For example, in most colleges, males are less likely to graduate than females. The data can help point to possible reasons for this — for example, males may be more likely to enroll part-time or in developmental education courses, compared to females. Research also shows that the progression patterns and outcomes of traditional-age college students (those who start before age 22) are often different from those of older, "non-traditional" students. By disaggregating the data by student age, the college can determine whether student progress differs by age group.

The more student characteristics included in the dataset, the more options will be available to disaggregate data and analyze results. An institution might not use all of the options; however, if student characteristics data are not collected, the opportunity to determine which characteristics are associated with student success, and how performance differs among the subgroups, will be lost.

Data on student progress. Once the college has decided which data elements to include when the cohort is initiated, the next step is to determine the data elements to be collected each term. Again, the decision about what to include should be guided by the questions the college is trying to answer, and the ease of collecting the data. Minimum measures here include courses attempted and courses successfully completed (for each developmental and gatekeeper course as well as any other course attempted), grades earned, term and cumulative grade point average, and participation in any special programs. It is possible to construct indicators that might help the college to understand how students move through their education. For example, a variable might be created that indicates whether a student remains full- or part-time or switches between the two statuses, and whether a student stops out and returns to the college.

For each cohort, a decision needs to be made about when to record initial student characteristics data and how long to follow that cohort. Students usually are included in a cohort if they are present in an institution's data system as of that institution's census date. A census date is typically that point in an academic term when 15 percent or 20 percent (depending on the institution) of a term has expired and at which time the student has settled her or his tuition and fees. Using the census date to determine who is in a cohort may exclude those students who dropped our earlier or who were unable to pay their tuition and fees. In either case, the institution excludes information on the experiences of some students — information that may be relevant to the student success agenda. Colleges are encouraged to follow students for at least four years, but even after six years, some students in the cohort will still be enrolled without having received a credential or having transferred.

Outcome data. Outcome data include information on any credentials the student earns at the college. Colleges may also want to know what happens to their students after they leave the institution. Information about jobs and earnings, and success at other colleges after transfer, can help a college assess the effectiveness of its own programs. Care must be taken in deciding which post-college data are meaningful. Often, community college students work while enrolled, so the key indicator is not whether they have a job, but whether they have a better job as a result of their education. Transfer to a different college is another example of post-college data that might be used, but colleges should try to learn if this change is permanent, or if it is a short-term enrollment of convenience. Trying to find and survey students who have left a college is notoriously difficult. A better way is to match student records with existing administrative data sets. Many states have state-wide unit record systems that make it possible to track students within the public higher education system, and the National Student Clearinghouse provides a national mechanism for colleges to track enrollment

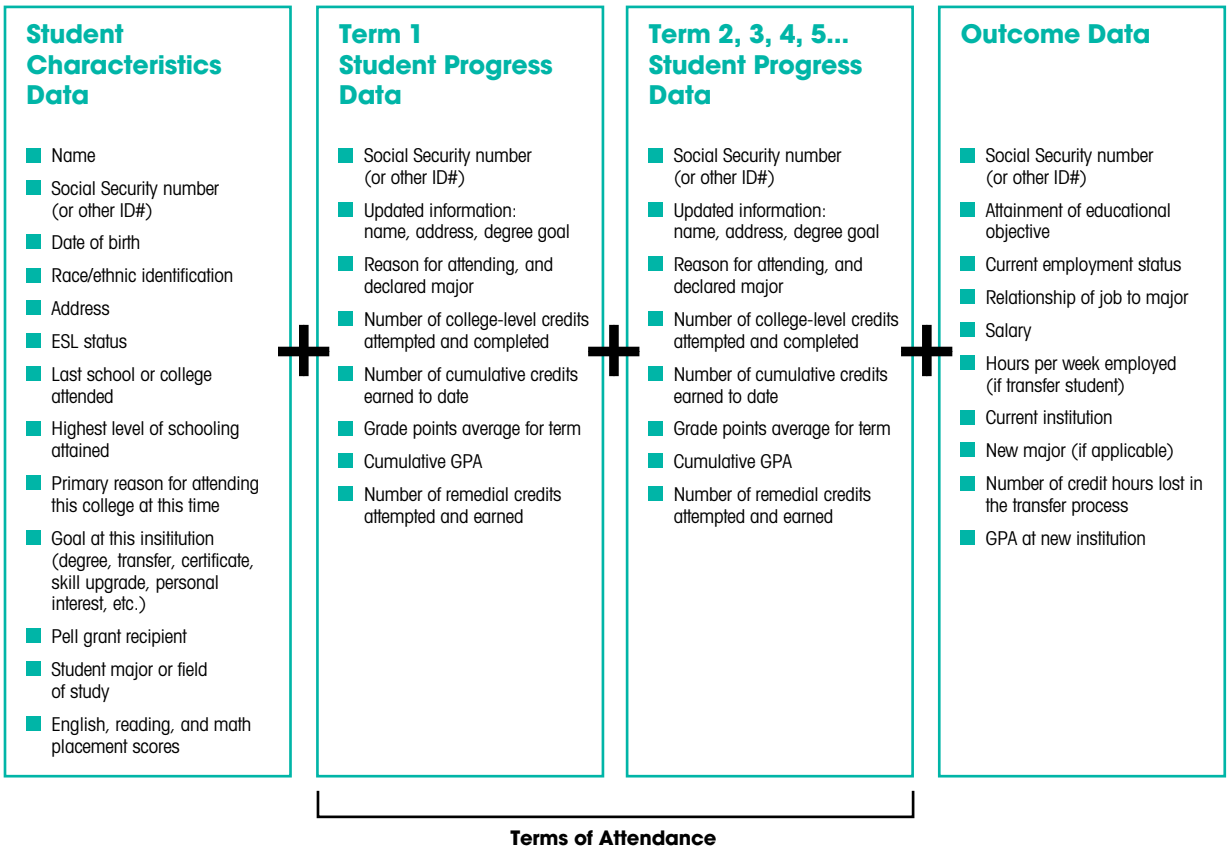
of their former students in other higher education institutions. State Unemployment Insurance (UI) wage record systems provide quarterly earnings data for individuals who are employed within the state.

Figure 4 shows how data on student characteristics, progress, and outcomes are typically organized to produce a longitudinal data file. First, the student characteristics data are gathered from data on the entering students, and assembled in a data file. Then, the data on student progress from the first term, and each term thereafter for a specified number of terms (most typically across four years), are added to the file. As term data are captured, outcome data are added.

Minimum requirements for a computerized system to manage and analyze the data include: (1) the ability to extract and download data elements already available in college records; (2) ease of modification and expansion of data elements; and (3) the preservation of student confidentiality. Colleges increasingly find it easier to compile data for cohort tracking in a "data warehouse" into which snapshot data that have been "cleaned" and put in a "flat file" format are entered. The time and attention required to clean data should not be underestimated; for example, names must be spelled correctly and consistently, student ID numbers must be accurate, indicators of student enrollment must be recorded, and updated grades should be added to students' records.

Figure 4

Components of Cohort Analysis



Tips on designing an effective cohort data analysis system

Cohort analysis can be used to help colleges understand patterns of student progression and success, but any analysis is only as good as the data that go into it. When designing longitudinal cohort tracking systems, colleges should keep in mind the following tips from other colleges that have done cohort tracking.

- **Reach consensus on technical definitions of data elements.** To ensure the institution uses consistent definitions for each data element, information technology, registration, and institutional research representatives should meet together on a regular basis to review data definitions, identify problems, and develop an approach for correcting them. For example, is enrollment captured as of the beginning of the term, the official census date, mid-term, or end of term (a student's record can look very different depending on the date used)? Is a dual credit student who comes to the college after high school a new or a continuing student? Are these students considered new students when they begin their dual credit work? Should they be included in the same new student cohort as students who first attend the college after high school? While no single "correct" answer exists, these seemingly arcane issues can have profound effects on the data used in longitudinal tracking. And if different schools use different definitions, the ability to compare institutions' performances will be severely compromised.

- **Keeping the data accurate.** Although this is the obvious first step in any kind of data analysis, some colleges may have data "siloed" in different departments. For example, academic and financial aid data may be stored separately. To build and maintain a current and relevant longitudinal database, the institutional research staff needs to have access to all data collected by various college departments. Once the database is in place, the IR staff should have established procedures for a standardized time to collect data, and for extracting, cleaning, and verifying the data, as well as a protocol for handling incorrect, missing, or updated data.

- **Develop a profile of current enrollment prior to deciding which data to disaggregate.** For example, how many students are full-time versus part-time? How many are 18–22 years old versus older adults? What are the mix of students by race/ethnicity, gender, income, and other characteristics? What percentage of entering students are referred to developmental courses? Getting a clear fix on the characteristics of entering students can help colleges decide which variables to use in disaggregating the results of longitudinal analysis.

- **Determine who is in a cohort.** Most Achieving the Dream institutions select cohorts based on the date students enter the institution. First-time, degree-seeking students enrolled in a given fall term is a typical selection criterion for creating a cohort. Colleges might also create cohorts of students who begin in other terms (since research suggests that students who enter in fall and spring terms may perform differently), or, alternatively, develop cohorts for students involved in specific programs or interventions. Achieving the Dream colleges are encouraged to look at historical trends, although retrieving older data that may have been overwritten, or are several years old, might prove too costly and time-consuming, especially if the historical data are limited or of varied quality.

- **Be judicious in the selection of variables.** Most data needed to satisfy Achieving the Dream requirements are available in existing administrative data systems and should not be complicated to retrieve and analyze. However, other data, including results of assessments, socio-economic status, marital status, and non-degree goals, are not typically found in administrative data systems. If the college decides to include data elements that are not routinely available, a collection plan and census date should be developed, along with the interface with the existing administrative data system.

- **Address student privacy concerns.** Privacy must be respected in the design and analysis of any longitudinal database. Many colleges use a unique identifier as a substitute for Social Security numbers. However, the college should be able to link these personal identifiers with Social Security numbers in order to match student data with databases outside an institution, such as other higher education institutions, state unemployment databases, teacher credentialing, and other sources that rely on Social Security numbers. This crosswalk file, as with all files that contain personally identifiable information, should be password-protected and stored in a secure location.

- **Carefully design visual presentation.** Criteria for presentation include clarity for non-technical audiences, brief analysis of the results, description of the methods used and limits of the data, and clear graphics. Care should be taken not to overwhelm audiences that are unaccustomed to data presentations with lots of quantitative graphs and charts. For these audiences, less is usually more.

Next steps after longitudinal analysis has been completed

As important as longitudinal cohort analysis is, it can only tell part of the story of student success. As mentioned, longitudinal data analysis is most useful in helping identify problems with student achievement. Additional research is usually needed to find the causes of such problems and develop possible solutions to them. For example, cohort tracking studies might show that students place into, take, and pass developmental math, and then some defer taking college-level math courses for a semester or more. If these students have a lower success rate in college-level math than do students who take it right away, then the quantitative data tell us that immediate enrollment in college-level math is associated with success. However, it does not tell us why students defer. A focus group or survey could help provide a better understanding of why these students are deferring enrollment. Ultimately, solving the problems identified through cohort analysis will require that colleges engage faculty, student services staff, and administrators to examine the data on the causes of the problems, implement and evaluate strategies to address them, and use the evaluation results to make further improvements. Thus, longitudinal cohort analysis can help to initiate what should become an ongoing process at each college of using data on student success to improve the impact of programs and services.

Rick Voorhees is principal of Voorhees Group LLC, an independent higher education consulting firm in Colorado, and has been involved in Achieving the Dream as a data facilitator and coach. He has worked with five Achieving the Dream colleges since the first round in 2004.

John B. Lee is president of JBL Associates, a consulting firm specializing in postsecondary education policy. He has dedicated his career to improving college opportunity for those who have been excluded.

Additional Resources

Bers, T. (2008). Student tracking in the community college. *New Directions for Community Colleges*, No. 143. San Francisco: Jossey-Bass, Inc.

Ewell, P.T. (2008, September). *Community College Data and Performance Measurement Toolkit.* http://www.communitycollegecentral.org/Downloads/Data_Performance_TOOLKIT.pdf

Ewell, P. T. (1995). Student tracking: New techniques, new demands. *New Directions for Institutional Research*, No. 87. San Francisco: Jossey-Bass, Inc.

Leinbach, T. & Jenkins, D. (2008, January). *Using longitudinal data to increase community college student success: A guide to measuring milestone and momentum point attainment.* CCRC Research Tools No. 2. New York: Community College Research Center, Teachers College, Columbia University. http://ccrc.tc.columbia.edu/Publication.asp?uid=570

Lillibridge, F. (2008). Retention tracking using institutional data. Student tracking in the community college, T. Bers (Ed.). *New Directions for Community Colleges*, No. 143. San Francisco: Jossey-Bass, Inc.

Tufte, E. (2002). *The Visual Display of Quantitative Information (2nd Edition)*. Cheshire, CT: Graphics Press.

Voorhees, R. A. & Hart, S. (1989). A tracking scheme for basic skills intake assessment. Using student tracking systems effectively, T. Bers (Ed.). *New Directions for Community Colleges*, No. 66. San Francisco: Jossey-Bass, Inc.

1. A "gatekeeper" course is the first college-level or degree-credit (non-developmental) course in the given subject area at a college.

Funding for this guide was provided by Lumina Foundation for Education as part of *Achieving the Dream: Community Colleges Count.* For more information, see www.achievingthedream.org. Davis Jenkins of the Community College Research Center at Teachers College, Columbia University, served as executive editor. Ellen Frishberg of JBL Associates served as managing editor. Trudy Bers, Rigoberto Rincones-Gomez of MDC, Inc, and Sue Clery and Barry Christopher of JBL Associates reviewed earlier drafts.