A guide to using state longitudinal data for applied research

Karen Levesque
Robert Fitzgerald
Jay Pfeiffer

RTI International

Levesque, K., Fitzgerald, R., and Pfeiffer, J. (2015). *A guide to using state longitudinal data for applied research* (NCEE 2015–4013). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development. Retrieved from http://ies.ed.gov/ncee/edlabs.

Contents

Overview: Seven steps in using state longitudinal data for applied research

State longitudinal data systems (SLDSs) promise a rich source of data for education research. SLDSs contain statewide student data that can be linked over time and to additional data sources for education management, reporting, improvement, and research, and ultimately for informing education policy and practice. This guide is intended for researchers who are familiar with research methods but who are new to using SLDS data, are considering conducting SLDS research in a new state environment, or are expanding into new topic areas that can be explored using SLDS data. The guide also may be useful for state staff as background for interacting with researchers and may help state staff and researchers communicate across their two cultures.

The guide highlights the opportunities and constraints that researchers may encounter and offers approaches to addressing some common problems. Readers can access the information in the guide in several ways. This overview summarizes seven steps for using SLDS data. These steps reflect sound research practice and highlight opportunities for and limitations of using SLDS data. The body of the guide discusses the steps in detail and provides examples. Checklists are provided at the end of the guide for each of the seven steps. Readers can read the entire guide; use the overview to identify sections of particular interest; or skip directly to the checklists. Readers who are familiar with SLDSs may want to focus on Steps 2, 3, 6, and 7. The seven steps could be performed in different order, and iterations may be needed as researchers become more familiar with the data. Because states volunteer to provide access to SLDS data for research, collaborating with state staff is likely to improve the quality of the resulting research.

- *Step 1. Understand the potential (and limits) of using state longitudinal data for applied research.* SLDSs contain longitudinal, universal, and comprehensive data, the combination of which provides a unique opportunity for education research. And SLDS data can often be supplemented by external data. But many of these systems are still in development, were not originally designed for research, and, as with other administrative data, may have limits when used for research. Researchers need to be familiar with data codebooks or dictionaries that describe in detail the data elements in the SLDS.
- *Step 2. Identify research questions and determine whether SLDS data are appropriate for addressing them.* SLDS data can support a broad range of descriptive and correlational analyses and may support causal analysis in some cases. As an initial step in designing a high-quality study that uses SLDS data, researchers should identify their priority research questions and determine whether SLDS data are appropriate for addressing the questions. Collaborating with state staff will ensure that the questions are of interest to the state and that researchers can better understand data strengths and weaknesses. Identifying the intended audience for an analysis is also important. Researchers should discuss with state staff how to frame the research to be beneficial both to state stakeholders and to the broader education research, policy, and practice communities.
- *Step 3. Select appropriate methods for the research.* A variety of methods for studying descriptive, correlational, and causal research questions can be supported by SLDS data. The third step in designing a high-quality study using SLDS data is for researchers to identify the most appropriate methods and the data requirements related to those methods.

- *Step 4. Confirm data availability.* A study's feasibility may depend on whether SLDS data are available for it. The study may require elements such as test scores, coursetaking, and teacher assignments for K–12 education, and links may be needed to early childhood education, postsecondary education, and workforce data. Plans for studies may need to consider the capacity to link across sectors and the specific data elements available within them. Researchers should plan to consult with state staff to determine whether the SLDS includes data needed for a planned study. Researchers also may want to consult with other researchers that have worked with a state's SLDS, which can help researchers quickly identify strengths and weaknesses in the data.
- *Step 5. Obtain the data.* In most states, researchers will have to submit a research request and complete a data-sharing agreement. Some states may require Institutional Review Board (IRB) review, and researchers affiliated with universities or other research organizations may need to have their study designs reviewed by their IRBs.
- *Step 6. Investigate data quality.* After obtaining the data, investigating its quality is a critical step in confirming the feasibility of a planned study. SLDSs are largely compiled from state and local data systems. SLDS submission processes involve checks for valid data values and formats, and some states have decision rules for reconciling data from different sources or time periods. Even so, data inconsistencies may not be revealed until researchers attempt to use SLDS data in new ways. Investigating SLDS data quality includes compiling the data, checking the data against external benchmarks, constructing derived variables, and conducting initial analyses. Even with careful planning, actually using the data can lead to surprises, and researchers may need to adjust their study plans.
- *Step 7. Communicate results.* Developing an agreement early in the research process about how results will be used and planning time for the state to review the results are crucial in the communication phase. Researchers should share results with state staff before communicating results to other audiences. State staff can help researchers identify issues that state stakeholders may have and can see problems that may threaten the validity of the results.

Researchers using SLDS data are often in new territory, particularly when studying topics that link data across education sectors and multiple data sources or when conducting SLDS research in a new state. Throughout an SLDS study, researchers should be prepared to adjust study plans as data issues and limitations become evident. Researchers can also contribute to the field about analyses that are possible with SLDS data, how SLDS data can contribute to a broad range of research questions, and how these data can inform education policy and practice. The discussions throughout this guide stress the importance of researchers collaborating with states on research design and implementation.

1. Understand the potential (and limits) of using state longitudinal data for applied research

State longitudinal data have expanded tremendously in the past decade. Since 2006, 47 states have received grants from the Institute of Education Sciences (IES) to support the design, development, implementation, and expansion of state longitudinal data systems (SLDSs).[1] All 50 states have pledged to build these systems and are in various stages of development. Information about each state's progress and plans can be found on two websites:

- SLDS Grant Program (http://nces.ed.gov/programs/slds/stateinfo.asp).
- Data Quality Campaign (http://www.dataqualitycampaign.org/your-states-progress/).

However, capacity of SLDSs varies widely between states. Many systems are still in development and differ in the sectors that they cover, the specific data sources and elements that they contain, and the number of years of data available. Researchers may want to obtain data codebooks or dictionaries from state staff or websites, which describe in detail the data elements in the SLDS (for some SLDS websites, see http://nces.ed.gov/programs/slds/stateinfo.asp).

SLDSs consolidate data from state and local systems that were typically designed to manage programs, allocate funds, and comply with federal and state accountability and reporting requirements—not to support research. While states are taking steps to make SLDS data available for research, their administrative systems can present challenges when used for that purpose. Success in designing a high-quality study that uses SLDS data may depend on understanding specific capacities and limitations of these data. These limitations are described throughout this document. Researchers should be aware that states with newer SLDSs may have more comprehensive data than states with older systems. And for some research purposes, *district* data systems may be more appropriate because they may contain additional data, such as formative assessment results and student behavior, that will not be included in state systems.

Characteristics of SLDS data

This section provides an overview of the potential of SLDS data for research and highlights how the longitudinal, universal, and comprehensive data contained in SLDSs provide a unique opportunity for education research. Individual studies may benefit from one or more of these characteristics. Often, SLDS data can also be linked to external data sources to supplement the SLDS data. Examples of studies that used SLDS data are provided to illustrate the range of possible research topics. This guide makes no claims, however, about the quality or validity of the studies cited. Often, previous research can be improved by using the more comprehensive SLDS data now available.

Longitudinal data

The typical SLDS warehouse contains annual or biannual data submissions of K–12 student records that can be linked over time and to other data sources to create student longitudinal records.[2] As researchers familiar with longitudinal analysis know, longitudinal data make it possible to measure change over time for individual students, schools, and districts. Depending on the capacity of an SLDS, it may be possible to

describe different paths that students take through education and into the workforce and to examine outcomes associated with different educational experiences. The following examples highlight studies that used state longitudinal data:

- Description of the relationship of school readiness to prior early care experiences using the Maryland Model for School Readiness kindergarten assessment connected to SLDS data (Maryland State Department of Education, 2012).
- Examination of whether teachers switching schools was associated with a change in teacher effectiveness as measured by student achievement gains: the study examined whether high-performing teachers in high-performing schools who were moved to low-performing schools maintained their effectiveness (Xu, Özek, & Corritore, 2012).
- Calculation of 4-year and 5-year longitudinal graduation rates: the study followed individual students in a high school cohort over time (Sass & Cartwright, 2008).
- Description of postsecondary education outcomes for high school students: the study analyzed immediate and delayed postsecondary enrollment and postsecondary completion within different timeframes for cohorts of high school graduates and examined postsecondary enrollment within four years following ninth-grade fall enrollment for cohorts of first-time ninth graders, using National Student Clearinghouse data on private and out-of-state postsecondary institutions connected to SLDS data on state public secondary and postsecondary education (Levesque & Fitzgerald, 2013).
- Examination of the paths that postsecondary students took toward a bachelor's degree: the study examined how enrollment in more than one institution was related to degree completion and subsequent earnings (Andrews, Li, & Lovenheim, 2012).

Some SLDSs contain annual data that extend back 5 to 10 years or longer; newer systems may contain data only for recent years. Having a longer span of data makes it possible to examine outcomes over a longer time period, such as examining postsecondary remediation rates associated with students' eighth-grade state test scores, which requires at least 6 years of student records. A longer span of data also makes it possible to examine trends in longitudinal and cross-sectional (point-in-time) measures.

The comparability and availability of data elements in SLDSs may vary over time, however, which may affect both longitudinal and cross-sectional trend analysess. When using data from multiple years, it is important to investigate whether definitions of key data elements have changed.[3] The following studies used state longitudinal data to explore trends over 10 or more years:

- Measurement of changes in school and classroom segregation as race-conscious school assignment policies were banned: the authors extended analyses of classroom segregation from 1994 through 2006 (Clotfelter, Ladd, & Vigdor, 2008).
- Examination of trends in teacher bonuses: the author examined changes in student achievement associated with the introduction of teacher bonuses between 1997/98 and 2007/08 (Vigdor, 2008).

Universal data

SLDSs typically encompass all districts, schools, and students.[4] The universality of SLDS data—at least in K–12 public education—makes it possible to answer questions about specific schools, districts, and states and their policies, practices, conditions, and outcomes. It is also possible to examine groups or programs that are small and may be underrepresented in sample surveys, construct comparison groups for descriptive or correlational analysis or for planning causal studies, and investigate the contribution of each education level (classroom, school, district, and state) to outcomes. And universal SLDS data make it possible to replicate previous research in new settings—for example, by extending research about urban areas into rural settings. The following studies used the universal data in an SLDS:

- Projection of county and regional trends in teacher demand in California (White & Fong, 2008).
- Comparison of the achievement gains of students taught by traditional and Teach for America teachers by using data from districts with concentrations of Teach for America teachers (Xu, Hannaway, & Taylor, 2007).
- Comparison of the achievement of charter school students with a matched comparison group of noncharter students: the comparisons included replicating a randomized controlled trial with propensity score matching to determine how closely the two methods produced similar estimates of student achievement (Furgeson et al., 2012).
- Examination of classroom peer effects and student achievement by using the Florida longitudinal data system to identify students' classroom peers and teachers in the elementary, middle, and high school grades (Burke & Sass, 2008).

Comprehensive data

SLDSs typically are comprehensive and contain many different types of data. Data in an SLDS often include the following data elements:

- Student, school, and district identifiers.[5]
- Student demographics, accountability subgroups, and participation in federal programs (such as special education status, migrant status, and eligibility for the National School Lunch Program).
- Enrollment and attendance rate.
- Completion[6] and withdrawal.
- Behavioral data, such as suspensions and expulsions.
- State and possibly local assessment data.
- Coursetaking, grades, and credits earned.
- College Board and ACT test data.[7]
- Staff data, including teacher data.

The content of specific SLDSs varies, however, as does the capacity to link across different K–12 data sources. Researchers should investigate the specific capacity of an SLDS. For example, common types of teacher data in SLDSs include demographics, certification and licensure, years of experience, and course assignments. Less common data may include compensation, professional development activities, and course

sections taught. Some SLDSs may not support linking of teachers with their students and classrooms, which is an important limitation for researchers who are considering value-added analyses concerning teachers.

Since 2009, states have begun expanding their SLDSs to encompass early childhood education, K–12 education, postsecondary education, and workforce data, which have been labeled P-20W systems (see http://nces.ed.gov/programs/slds/about_SLDS.asp). The Data Quality Campaign reports that as of 2014, 18 states plus the District of Columbia (DC) link data systems across the P-20W spectrum, although specific linking capacity still varies.[8] Many states link K–12 education data annually with at least one other system:

- Early childhood education (43 states).[9]
- Postsecondary education (43 states).
- Workforce (19 states).

In addition, some states have links to or are planning to establish links to data on career and technical education, adult education, and social services. Most states maintain these data separately from K–12 education, early childhood education, postsecondary education, and workforce data. As SLDSs become more comprehensive, researchers will be better able to examine relationships among postsecondary and workforce outcomes and preK–12 education. Examples from several studies suggest the usefulness of being able to link data across sectors:

- Examination of the association between early childhood initiatives and grade 3 outcomes, linking North Carolina birth records and administrative records for early childhood programs with K–12 assessment data (Ladd, Muschkin, & Dodge, 2012).
- Comparison of the effectiveness of graduates of different teacher preparation programs, as measured by achievement gains of their students, linking information from Missouri's teacher certification system with teacher and student assessment data in grades 4–6 (Koedel, Parsons, Podgursky, & Ehlert, 2012).
- Comparison of the variety of pathways through public postsecondary institutions in Texas and linking those pathways to postcollege earnings by using state unemployment insurance records (Andrews, Li, & Lovenheim, 2012).

Additional information about SLDS data

Linking to external data sources

SLDSs typically consist of administrative data compiled from state and local data systems and contain data sources and elements common across the state,[10] while data sources unique to particular districts or schools are maintained locally. Depending on the focus of a study, researchers may need to supplement SLDS data with data that are not in the SLDS, such as teacher certifications, enrollment at out-of-state postsecondary institutions, or district data on formative assessments. Most states assign a unique student ID when a student enters the K–12 system. A few states use the same ID for postsecondary education, which facilitates links between K–12 and postsecondary data. In some states, SLDSs can be linked to external data sources by using Social Security numbers (SSNs), names, and dates of birth. SLDS data can also be linked to data available from federal and national sources, such as the Common Core of Data (CCD), the Integrated

Postsecondary Education Data System (IPEDS), and Census data. The following studies merged SLDS data with data from external sources:

- For a descriptive study of the postsecondary education outcomes of public high school students in three states (Florida, Georgia, and Texas), researchers supplemented K–12 and postsecondary education data from SLDSs with National Student Clearinghouse (NSC) data on student-level enrollments and completions at private and out-of-state postsecondary institutions; with CCD data on school and district characteristics; and with IPEDS data on postsecondary institution characteristics (Levesque & Fitzgerald, 2013).
- For a study of teacher preparation programs, researchers used other state sources to supplement data on students and teachers from the Florida longitudinal data system with data on teachers' certification and their institution of preparation (Mihaly, McCaffrey, Sass, & Lockwood, 2012).
- For a study of the relationship between teacher credentials and student achievement at high schools in North Carolina, researchers supplemented student and teacher data from the SLDS with data in school-provided activity reports to link students and teachers to classrooms and to each other (Clotfelder, Ladd, & Vigdor, 2007b).

Understanding SLDS governance and structure

State education agencies typically are responsible for the K–12 SLDS. The structure and governance of larger P-20W systems varies, however. Some states have centralized P-20W data warehouses. Others have data-sharing agreements among agencies and compile data on an ad hoc basis to respond to research requests. In some states, state agencies are the governing agency for the P-20W data. Other states have created or designated an independent agency to govern P-20W systems. For example, in Texas, the Texas Education Agency is responsible for the K–12 longitudinal data system, whereas centers at three state universities compile and manage its P-20W data. Recently, Washington established the Education Research & Data Center in the state's Office of Financial Management, which is responsible for compiling data from source agencies to respond to requests. Additional information on the structure and governance of P-20W data efforts can be found in the SLDS Grant Program (2013).

Collaborating with state staff

Generally, it is mutually beneficial for researchers to engage in an ongoing dialogue with state staff about SLDS interests and concerns. State staff can help researchers understand the capacities and limitations of an SLDS. The resulting research can inform state policy and advance practice. Researchers can provide feedback on data inconsistencies or decision rules or other data strategies developed through the research, which can improve SLDS quality. Researchers can also communicate their research results to policymakers and other state leaders—at state or national conferences or panels, or as a part of legislative testimony, budget justifications, or policy evaluations. In Steps 2–7, which follow, it is important for researchers to consult with state staff to ensure that the planned study using SLDS data is mutually beneficial.

Researchers should realize, however, that responding to research questions about, and requests for, SLDS data can take substantial time and resources on the part of state staff, who often must respond first to state priorities. Keeping these state priorities in mind can help establish a respectful and collaborative

relationship. The Forum Guide (National Forum on Education Statistics, 2012) offers additional guidance to both researchers and state staff for partnering productively.

Taking advantage of SLDS resources

The Institute of Education Sciences SLDS Grant Program offers a variety of support resources on its website, http://nces.ed.gov/programs/slds/resource_type.asp. The National Forum on Education Statistics produces helpful publications and resources related to SLDSs and data systems in general on its website, http://nces.ed.gov/forum/index.asp. The Forum is a voluntary and cooperative federal-state-local body with a mission to develop and recommend strategies for building education data systems that support local, state, and national efforts to improve public and private education. Links to websites are inserted in the text when they provide access to multiple resources or to online information that is not available in a downloadable document; other citations are provided in the References section.

2. Identify research questions and determine whether SLDS data are appropriate for addressing them

A second step in designing a high-quality research study that uses SLDS data is to identify the research questions and determine whether SLDS data are appropriate for addressing them. Identifying the intended audience is also important and can help focus the research questions on stakeholder needs and interests.

Developing a research partnership

Consulting with state staff may yield research questions that benefit state stakeholders and the researcher. Researchers should try to understand state goals and align their research requests with these goals, where possible. Making the effort to understand the state context can build trust in the researcher and help identify research opportunities. For example, a change in state policy could create the opportunity for a natural experiment. *SLDS Data Use Issue Brief 2* offers additional guidance on forming research partnerships (SLDS Grant Program, 2012a).

Situating SLDS research within an overall research agenda

Researchers may find it helpful to situate SLDS research within their overall research agenda, with SLDS data used to explore some aspects of the agenda and different data used to explore other aspects. Situating SLDS research within a research agenda can clarify aspects of the agenda that can be addressed using SLDS data, set priorities among research questions and analyses, and identify gaps in the research that may need to be filled with other strategies. An agenda is straightforward to describe, and it may be easier for state staff to develop an ongoing partnership based on a research agenda than for state staff to respond to ad hoc data requests. For example, a research agenda on college readiness might comprise the following questions:

1. How do faculty at state postsecondary institutions define college readiness?
2. What proportion of students are college ready when they graduate from high school, as measured by proficiency on state achievement tests, scores on the SAT or ACT exams, and no need of remedial college coursework?
3. How is high school achievement—measured by state test proficiency, high school coursetaking, and grades—related to the need for remediation?
4. Does attaining college-ready benchmarks in high school (such as completing higher-level mathematics courses or attaining specified performance levels on state and national tests) reduce the need for remediation?

Researchers should look for areas of intersection with state research concerns. Some states publish research agendas on their websites (for some SLDS websites, see http://nces.ed.gov/programs/slds/stateinfo.asp).

Analysis of SLDS data could address questions 2 and 3, while question 1 may require a survey of faculty. Question 4 may require a causal research design, although SLDS data can still play a role in the study. For questions 2 and 3, the data on postsecondary remediation that are typically available in an SLDS may call for reframing the research question or incorporating additional data sources.[11]

In defining research questions for a study, researchers should consult with state staff to clarify the study population as carefully as possible. Who is included or excluded from a study can affect the credibility of the resulting research and how state stakeholders perceive the findings. For example, researchers should consider types of schools and students to include, given the state context, such as whether charter schools or summer graduates should be in the analysis.[12] SLDS staff can clarify whether the study population of interest is contained in specific SLDS data sources or whether additional data sources or files must be requested for full coverage.

For each of Steps 2–6 in this guide, boxed examples focus on college readiness and success, drawing from a recent three-state pilot project. The pilot is used to provide illustrative, concrete examples based on the authors' own experience. See Example 1 for a description of the pilot. Examples from other studies are also used throughout the guide. The guide makes no claims, however, about the quality or validity of the examples and studies cited.

Addressing descriptive research questions

SLDSs have the capacity to address a broad range of descriptive research questions, including What? Who? How many? How often? At what rate? Descriptive questions can be the starting point when first analyzing a state's data because they offer opportunities to investigate the availability, completeness, and appropriateness of SLDS data for specific research topics. The results of descriptive analyses can offer states, districts, and schools useful information that they may not already have compiled on their own.

Example 1. Pilot tracking postsecondary outcomes for high schools

The Tracking Postsecondary Outcomes for High Schools pilot project collected SLDS data on all public school students in grades 8 through 12 in three states (Georgia, Florida, and Texas) in each academic year 2000/01 through 2009/10, where available. The pilot also collected SLDS data on enrollment, remediation, persistence, and completion at state public postsecondary institutions over the same period. Ultimately, the pilot consolidated data from 300 million education records—from dozens of SLDS data sources over 10 academic years in the three states—to create 4 million unique student longitudinal records and more than 100 measures describing students' high school and college experiences. These state data were supplemented with NSC data on postsecondary enrollments and completions at private and out-of-state institutions, CCD data on school and district characteristics, and IPEDS data on institutional characteristics. Results of the analyses were made available during the pilot to state, district, and school staff through descriptive reports and a Web tool. Use of pilot examples is not intended as an endorsement of the pilot.

Three broad categories of descriptive research questions are (i) describing status and trends, (ii) describing populations of interest, and (iii) comparing schools and districts. These questions take advantage of the longitudinal, universal, and comprehensive nature of SLDS data.

Describing status and trends

These questions seek to describe the status and trends in educational conditions and outcomes for specific states, districts, schools, or other organizational units. Answers can help identify educational conditions and outcomes that warrant additional attention. The following examples show descriptive questions studied with SLDS data:

- What proportion of entering kindergarteners is ready to perform kindergarten work, and what is the trend in school readiness over time (Maryland State Department of Education, 2012)?
- What is the state's graduation rate, and what is the trend of this rate (Sass & Cartwright, 2008)?
- What is the state's need for new teachers in the coming decade, and how does this need vary by county and region (White & Fong, 2008)?
- How does career and technical education participation vary across high schools and regions, and how good a fit is there between program participation and labor market demands statewide and for each region (Mokher, 2011)?

Describing populations of interest

Most descriptive questions that use SLDS data can be examined for specific populations of interest, such as gender, race/ethnicity, poverty or economic disadvantage status, English learner status, disability status, and migrant status. For example:

- How do rates of school readiness vary for low-income versus non-low-income children in the state (Maryland State Department of Education, 2012)?
- How does the state's longitudinal high school graduation rate vary by race/ethnicity of ninth graders (Sass & Cartwright, 2008)?

Comparing schools and districts

Most descriptive analyses that use SLDS data can be examined by school, district, or other organizational unit, such as a county or region. Results can help create comparison benchmarks and identify higher- and lower-performing schools and districts in the state.

Example 2. Organizing descriptive reports in an online library

The Tracking Postsecondary Outcomes for High Schools pilot project sought to develop high-quality, actionable data that linked K–12 and postsecondary education for use by schools, districts, and states to improve the college readiness and success of their students. These data were disseminated through the pilot's Web tool, Advance, which offered an online library of descriptive reports organized by question, as well as a build-your-own table tool. The following high-level questions were included in the reports library:

- What are the postsecondary experiences of our high school students (including postsecondary enrollment, remediation, persistence, and completion)?
- How does performance on standardized tests relate to our students' postsecondary success?
- How does high school coursetaking relate to our students' postsecondary success?
- How does grade 9 performance relate to our students' postsecondary success?

Standard reports included breakdowns by gender, race/ethnicity, and economic disadvantage status. Reports showed results for the most recent year available, trends over four years, and comparisons with similar schools and districts.

Visit the Reports Library and table tool prototype at http://review.mprinc.com/advance/.[13]

Examining correlational and causal research questions

Researchers, policymakers, practitioners, and other education stakeholders want to understand the factors and conditions associated with education outcomes and the kinds of interventions that improve these outcomes. For example, many studies have used SLDS data to examine how characteristics of teachers and students correlate with student test scores (Xu, Hannaway, & Taylor, 2007; Goldhaber, Gross, & Player, 2007; Clotfelter, Ladd, & Vigdor, 2007a, 2007b, 2009; Vigdor, 2008; Master, Loeb, Whitney, & Wyckoff, 2012; Xu, Özek, & Corritore, 2012; Koedel et al., 2012; Fuller & Ladd, 2013). Researchers have also used SLDS data to investigate other correlational and causal research questions:

- How curriculum and instructional practices are associated with student achievement in high school (Cavalluzzo, Lowther, Mokher, & Fan, 2012; Clotfelter, Ladd, & Vigdor, 2012).
- How student mobility is associated with student achievement (Xu, Hannaway, & D'Souza, 2009).
- How participation in early childhood education programs is associated with child outcomes (Peisner-Feinberg & Schaaf, 2010; Ladd, Muschkin, & Dodge, 2012; Barnett, Jung, Youn, & Frede, 2013).
- How education policies and practices are associated with early postsecondary education outcomes (Cornwell & Mustard, 2002; Fong, Huang, & Goel, 2008).
- How student and school characteristics are associated with postsecondary and labor market success (Calcagno & Long, 2008, 2009; Andrews, Li, & Lovenheim, 2012; Clotfelter, Ladd, Muschkin, & Vigdor, 2012; Dadgar & Weiss, 2012).

The National Center for Analysis of Longitudinal Data in Education Research (CALDER) has a long history of research using SLDS data. CALDER's website (www.caldercenter.org) describes its research focus and analytical methods and includes a listing of publications on a variety of topics. The National Center for Education Evaluation and Regional Assistance (NCEE) website (http://ies.ed.gov/ncee/edlabs/projects/) provides links to reports prepared by the Regional Education Laboratories, some of which used SLDS data.

3. Select appropriate methods for the research

The next step in designing a high-quality study is to select appropriate methods for addressing the research questions. Both the capacities and constraints of SLDSs should be considered when selecting an appropriate research method. And researchers may need to revisit the initial research method after examining the data.

This section describes methods for addressing the descriptive, correlational, and causal research questions described in Section 2 and explains how SLDS data support the methods. The section also describes possible uses of SLDS data in exploratory analyses, randomized studies, and natural experiments. Examples of studies are provided to illustrate the range of possible analyses that may be conducted. The guide makes no claims, however, about the quality or validity of the studies cited. As mentioned earlier, previous research can often be improved by using the more comprehensive SLDS data that are now available.

Addressing descriptive research questions

Methods for analyzing descriptive questions include:

- Cohort analysis.
- Trend analysis.
- Gap analysis.
- Tests of difference.
- Banding of similar schools.
- Residuals analysis.

Example 3. Conducting simple but informative descriptive analysis

The sample high school feedback report from the Tracking Postsecondary Outcomes for High Schools pilot project in Appendix A displays simple but informative descriptive analyses related to college transition:

- Percentage of high school graduates who enrolled immediately in postsecondary education, with breakouts for the characteristics of postsecondary institutions attended.
- Subgroup breakouts on the main outcome by race/ethnicity, gender, and income status, compared with the state average.
- Persistence and continuous enrollment, by level and control of postsecondary institution.
- A 4-year trend for the main outcome.
- Subgroup breakouts by proficiency on the high school achievement test.
- Breakouts for proficiency on the high school achievement test, by proficiency level on the grade 8 achievement test.
- Placement in a statewide histogram.
- Scatterplots of all high schools in the state with schools' postsecondary enrollment rates on the y-axis and schools' grade 8 proficiency rates for their graduates on the x-axis, identifying schools with the largest positive residuals in each grade 8 achievement band.

Reports were also produced at the district and state levels. These reports were used by district and school staff to identify strengths and weaknesses related to the college outcomes of their students, conduct root cause analysis, and identify high-performing peer schools.

When presenting results of descriptive analyses, it helps to include key definitions, clear labels, text that helps to interpret charts and graphics, and questions that prompt further discussion and reflection—and to point users to additional technical information. For illustrations of supporting information for practitioner audiences, see the sample high school feedback reports in Appendix A. Presentation of results matters for policy and practitioner audiences, and it is advisable to share draft presentations with stakeholders before presenting final results. Researchers should not aim to gain stakeholder approval of the research findings, but rather to engage in a collaborative conversation that can help build relationships and deeper understanding of the findings and ensure that the results are presented in a useful and understandable way.

Several cautions should be noted in using SLDS data for descriptive analysis. One caution for descriptive research is in interpreting results for small populations, particularly at the school level. To facilitate interpretation, it is important to provide the population numbers underlying each reported result. For example, postsecondary outcomes for a high school graduating class of 50 students should be interpreted with greater caution than for a graduating class of 500 students. Researchers may also want to consider not reporting some results for small populations (in the Tracking Postsecondary Outcomes for High Schools pilot project, researchers and stakeholders agreed not to report school results for high schools with fewer than 30 graduates). Typically, state education agencies require that results for very small numbers (usually cell sizes of 10 or fewer) be suppressed for confidentiality. See SLDS Technical Brief 3 (Seastrom, 2010) for a discussion of methods to protect personally identifiable information in aggregate reporting.

Another caution is the possibility of outliers among schools. SLDSs may contain data from magnet schools, selective schools, correctional facilities, and medical facilities, and the different settings may generate data that appear anomalous. Researchers should consider whether they will include or exclude these schools from analysis.

A third caution for descriptive research is handling missing data. Counts of students or graduates that do not match expectations can erode confidence in the data and results. Providing explanations for missing data is important, especially when missing data are nonrandom—for example, when a large group is missing (such as, in the Tracking Postsecondary Outcomes for High Schools pilot project, missing enrollment data for an entire postsecondary institution or campus). Useful resources for handling bias due to missing data are provided in Section 6 under "Initial analyses to investigate data quality."

Addressing correlational and causal research questions

Researchers may want to use statistical analysis to estimate relationships between outcomes and interventions or characteristics. The following studies use SLDS data in different research designs:

- *Regression discontinuity,* examining effects of college scholarships (Cornwell & Mustard, 2002), school vouchers and accountability (Rouse, Hannaway, Goldhaber, & Figlio, 2007; Özek, 2012), and teacher bonuses (Vigdor, 2008).
- *Difference-in-differences*, examining teacher mobility (Feng, Figlio, & Sass, 2010) and school principals (Branch, Hanushek, & Rivkin, 2012).

- *Fixed effects*, examining teachers (Xu, Hannaway, & Taylor, 2007; Harris & Sass, 2007a,b; Goldhaber, Gross, & Player, 2007), curricula (Clotfelter, Ladd, & Vigdor, 2007a, 2009), peer effects (Burke & Sass, 2008), and school mobility (Xu, Hannaway, & D'Souza, 2009).
- *Hazard analysis*, examining teacher career transitions (Goldhaber, Gross, & Player, 2007) and teacher mobility (Feng & Sass, 2011).
- *Propensity score matching*, examining charter schools (Witte, Wolf, Dean, & Carlson, 2010; Fortson, Verbitsky-Savitz, Kopa, & Gleason, 2012).

Other studies based on SLDS data used interrupted time series designs for examining effects of teacher bonuses (Vigdor, 2008) and benchmark assessments (Henderson, Petrosino, Guckenberg, & Hamilton, 2008); multilevel modeling for examining National Board for Professional Teacher Standards–certified teachers (Harris & Sass, 2007b) and classroom peer effects (Burke & Sass, 2008); and instrumental variables for examining public school choice (Nichols & Özek, 2010) and school restructuring (Özek, Hansen, & Gonzalez, 2012). Discussions of various methods and papers that used SLDS data can be found on the CALDER (www.caldercenter.org/methods.cfm) and NCEE (http://ies.ed.gov/ncee/pubs/) websites. CALDER and NCEE paper numbers are included in citations in the References section of this guide and can be used to locate the papers on these websites.

The potential and limitations of using SLDS data with some common designs also emerge from previous studies:[14]

Regression discontinuity

- *Potential:* SLDS data may provide a sufficient number of cases both above and below a cutpoint, contributing to precise estimates of effects.[15]
- *Limitations:* This method requires an assignment or 'forcing' variable with appropriate cutpoints, which may not be available in the SLDS. For example, students may be eligible for programs based on family income or other continuous measures of socioeconomic status, which often are not available in an SLDS. Typically, SLDSs contain only dichotomous measures relates to student poverty, such as free or reduced-price lunch status or economic disadvantage status.

Difference-in-differences

- *Potential:* The annual and longitudinal nature of SLDS data makes it possible to measure the same students (or other units) both before and after the intervention of interest.[16]

- *Limitations:* Years of available data may be insufficient.

Fixed effects

- *Potential:* The annual and longitudinal SLDS data lend themselves to this method, which relies on multiple observations of cases over time, including before and after the intervention of interest.[17]
- *Limitations:* Years of available data in the SLDS may be insufficient, or data may not be comparable over time (such as noncomparable test results). For school effects, student mobility

rates across schools within the school year may vary widely and be nonrandom. Researchers must carefully consider decision rules for assigning students to schools. With regard to teacher effects, researchers should investigate which adults contribute to instruction in the classroom. For example, two teachers may be assigned to a larger class rather than one teacher assigned to a smaller class. In addition, nonteachers may be doing teaching-type tasks, and students may interact more with this person. Attributes of classrooms may be nonrandom and correlated with outcomes. For studies that examine effects of teacher preparation programs, sufficient years of data may not be available to avoid regional clustering of graduates, which tends to lessen over time (Mihaly et al., 2012).

Hazard analysis (or event history analysis)

- *Potential:* Because student data in SLDSs tend to be recorded in academic year intervals, discrete-time methods may be used.[18]

- *Limitations:* Years of available data may be insufficient for this method. Measures may be too infrequent.

Propensity score matching

- *Potential:* Again, the annual and longitudinal SLDS data make it possible to measure students (or other units) before and after the intervention.[19]

- *Limitations*: The SLDS may not contain an adequate number of appropriate regressors to estimate the propensity scores. Pretest scores are critical. Without them, course grades, grade point average, or interim assessments could be used as achievement measures. Without sufficient matching traits, the fit of the probability model that produces the propensity scores may be poor, and propensity score matching may not represent an improvement over regression.

SLDS data can be used with other matching methods (Fortson et al., 2012). Whether any of the above methods can be used depends on whether the underlying assumptions of the method are met and the data required by the analysis are available.

<u>Using SLDS data in exploratory analyses</u>

SLDS data can also be used in exploratory data analyses, which may use descriptive or correlational methods. Exploratory analysis can be useful to states and preliminary to further research. Such analyses can identify predictive or at-risk factors that help state staff administer programs, allocate resources, and refine policies. For example, identifying at-risk factors could help state efforts to reduce dropout rates. And research methods for defining comparable schools could help state accountability efforts. Researchers should discuss with state staff the exploratory analyses that could be useful and should plan to share the results of these analyses ahead of final study results.

Using SLDS data in randomized studies and natural experiments

SLDS data can support randomized studies. SLDS data can provide outcome data for a randomized trial, for example, providing state achievement test scores. An SLDS can support randomized trials by enabling researchers to identify students or schools that could be randomized. SLDSs can also support analyzing actual or "natural" experiments that arise, for example, from charter school lotteries, program oversubscription or wait lists, or changes in state policy. And SLDSs can supplement and possibly fill gaps in local education data. For example, studies that seek to track longer-term outcomes for participants can use the data in an SLDS to locate students who have left the school or district where an intervention occurred but are still enrolled in the state. SLDS data can help confirm the withdrawal status of students and whether students persisted, graduated elsewhere, or dropped out.

Applying general cautions when analyzing SLDS data

A general caution when using the preceding methods is that practices for collecting data submitted to the SLDS may vary because of differences in policy or procedures. These variations may not be evident in the SLDS data or documentation. For example, districts that have implemented universal free lunch may report that all of their students are eligible for free or reduced-price lunch, while other districts may report individual student eligibility for free or reduced-price lunch.

A second caution is attrition of the study sample. When data are used from multiple years and data sources within the SLDS, cases with valid values on all key variables may be significantly smaller than the population of interest. It is important to compare the characteristics of the final study sample with the relevant population and to conduct an attrition bias test, if possible.

A third caution is that older SLDS data may be less complete or accurate than more recent data. To improve the usefulness of older data, researchers may want to use newer data to validate or impute older data. For a discussion of these and other analysis issues when using administrative data, see SLDS Data Use Issue Brief 4 (SLDS Grant Program, 2012b).

A fourth caution when using data from multiple sources is that the timing and frequency of data collection for common or important data elements may vary, and the least frequent periodicity should possibly be used (Holt, White, & Lichtenberger, 2013).

4. Confirm data availability

The next step in designing a high-quality study is to determine whether the SLDS includes the data needed for the planned study. It is important to make this determination before taking formal steps to obtain the data. Confirming data availability involves confirming the content areas contained in an SLDS and the capacity to link across these areas, as well as confirming the availability of specific data elements. It is also important to clarify the state's policy for accessing the SLDS.

Confirm SLDS content and capacity

SLDSs are still under development in many states, and few may support the full range of possible content described in Step 1:

- K–12 education data, including

  - Student, school, and district identifiers.
  - Student demographics, accountability subgroups, and participation in federal programs (such as special education status, migrant status, and eligibility for the National School Lunch Program).
  - Enrollment and attendance.
  - Completion and withdrawal.
  - Behavioral data, such as data on suspensions and expulsions.
  - State and possibly local assessment data.
  - Coursetaking, grades, and credits earned.
  - College Board and ACT test data.
  - Staff data, including teacher data.

- Early childhood education data.
- Postsecondary education data.
- Workforce data.
- Career and technical education data.
- Adult education data.
- Social services data.

The capacity of SLDSs to link data across content areas varies. According to the Data Quality Campaign, 18 states and the District of Columbia have comprehensive P-20/W systems (including linked early learning, K–12, postsecondary, and workforce data) as of 2014.[20] The Data Quality Campaign publishes state-by-state information on SLDS capacity, including the ability to link across sectors (see www.dataqualitycampaign.org/). In addition, the National Center for Education Statistics (NCES) publishes a variety of resources on its SLDS Grant Program website (http://nces.ed.gov/programs/slds/), including a table that describes the characteristics of statewide student data systems as of 2009/10 (see Appendix B). The capacities reported by these sources may mask considerable variation in state practice, however, and additional investigation is necessary to understand a state's operational capacity.

Review SLDS documentation and talk to state staff

Researchers should investigate the capacities of an SLDS by reviewing its documentation (which may include data handbooks, data dictionaries, user guides, manuals, and decision rules) and by discussing specific data needs identified in Step 3 with SLDS staff. Links to relevant websites for most states can be found at http://nces.ed.gov/programs/slds/stateinfo.asp. Some states do not have data dictionaries for their SLDSs or source systems, however, and some documentation may not include all available data elements. Additional state data sources, such as item-level test data, may not be included in the SLDS but could be made available and linked to the SLDS data for research purposes.

A caution is that not all states correct historical data after problems have been found. Such errors may be posted as errata on the state website but not always. Researchers should clarify state practices for correcting and documenting errors and where to find errata.

Confirm content and coverage of specific data sources

Most SLDSs contain data from multiple sources—such as the K–12 and postsecondary sectors—and it is important to clarify elements from each data source. For example, assessment data for K–12 education are typically maintained separately from enrollment, attendance, disciplinary, and course records. The assessment data files may contain basic demographics and school of enrollment at time of testing but may be missing exit or withdrawal status (including graduation and dropout status), suspensions, and coursetaking, among other data. Details are not always documented, and it may be necessary to talk with state staff. Ultimately, the researcher may need to request data from multiple data sources within the SLDS in order to construct the complete picture of a study population. In some cases, it may be necessary to redefine a study population on the basis of available data. For additional examples, see SLDS Data Use Issue Brief 4 (SLDS Grant Program, 2012b).

Understand SLDS linking practices

Understanding a state's data linking practices is critical for understanding what SLDS links are possible as well as the quality of the resulting data. The typical SLDS warehouse contains annual data submissions of K–12 student records that can be linked over time and to other data. Common student identifiers are used to link across K–12 records. While a few states use SSNs as the common identifier, most states assign a unique student ID when a student enters the K–12 system. A few states use the same ID for both K–12 and postsecondary education, which can facilitate links between the two sectors. In the few cases where a state uses SSNs as the K–12 identifier, linking to other data sources such as postsecondary education is straightforward, although there are always duplicate or invalid SSNs for which rules must be developed. Where a state does not use SSNs, linking to other data sources can be done by using probability methods. Matching is done by using a combination of names, date of birth, gender, and race/ethnicity. SLDS rules may define a "match" when the probability is high (for example, greater than 90 percent). Researchers should request information on the state's match rate to understand how well the linked data reflect the population and to identify possible biases if some groups are less well-matched.

<u>Confirm linking capacity</u>

Most SLDSs contain K–12 student data but vary in their capacity to link them with other data sources. For example, an SLDS may contain teacher data, but it may not have information to link these data to specific students or classrooms. Some SLDSs may contain postsecondary education data but not have information to link these data to K–12 student data.

Linking K–12 teachers to students.

Teacher effectiveness has been the focus of many studies. The growth of value-added assessments of teachers and the use of value-added measurements in areas such as teacher pay, tenure, and dismissal are important drivers for incorporating teacher data into SLDSs. Such analyses require links among data about assessments, courses, students, and teachers. States vary in their capacity to link student and teacher data. Where teachers cannot be linked to students, researchers may need to develop workaround procedures:

- Clotfelter, Ladd, and Vigdor (2007b) examined the relationship between teacher characteristics and student scores on reading and math tests. Records reported the name of the staff member who administered the test, but it was not always the case that this person was the student's classroom teacher. To create a more accurate link, researchers created a restriction that a proctor whose name appeared on a course record had to teach a valid math or reading course during the period covered by the test.
- Xu, Hannaway, and Taylor (2007) used SLDS data from North Carolina to compare the effectiveness of Teach for America teachers and regular teachers. They developed procedures to improve the accuracy of the match between assessed students and their teachers by matching test classrooms to instruction classrooms based on student demographic compositions, and then matching to the classroom instructors.

Linking K–12 and postsecondary education data

Currently, all 50 states as well as DC and Puerto Rico use a unique student identifier (ID) for K–12 education (Data Quality Campaign, 2014). In some states (including Florida, Maryland, and Texas), the unique identifier is the SSN. Other states have or are implementing a K–20 ID process to assign students an ID at the point and time of entry into the education system. When a unique ID is not available, researchers can match K–12 and postsecondary students by using a combination of names and date of birth and possibly high school of graduation and graduation year. As of 2014, 42 states plus DC report the ability to match student records annually between K–12 and higher education systems.

A state may report that postsecondary education data are in its SLDS, but not all state postsecondary systems may be represented. Public 4-year or public 2-year institutions are more likely to be linked to K–12 data, and public less-than-2-year institutions are less likely.[21] Moreover, not all states have centralized data stores for each postsecondary system, and individual institutions or campuses may be responsible for submitting data, resulting in uneven representation in the SLDS. National Student Clearinghouse (NSC) data can fill gaps in coverage of postsecondary institutions. In fact, some states use the NSC as their primary data source on postsecondary enrollments and completions.

The Data Quality Campaign (2014) reports that 43 states annually link K–12 data with early childhood education data, but most SLDSs do not fully cover early childhood programs. Many states can link K–12 data systems to early special education (44) and state prekindergarten programs (42), but fewer can link to subsidized child care (12) or Head Start–Early Head Start programs (23). Thirteen states are members of the Early Childhood Data Collaborative and report making progress with their early childhood education data (see www.ecedata.org/state-success-stories). North Carolina has been a leader in linking early childhood education and elementary education data (Peisner-Feinberg & Schaaf, 2010; Ladd, Muschkin, & Dodge, 2012).

Confirm the availability of specific data elements

After confirming whether an SLDS contains the content areas and the capacity to link across areas, it is important to confirm the availability of specific data elements. The number of years of data and specific data elements in SLDSs vary widely. In some states, such as Florida, researchers are asked to justify each data element they request, so it may be necessary to understand the SLDS data elements thoroughly to prepare the data request.

Where analyses require specific data, it is important for researchers to investigate the capacity of an SLDS to provide the needed data. Beginning with data elements needed for the planned study, researchers should first review documentation, such as data codebooks or dictionaries, that describe the data elements in the SLDS and are published on the state website (for some SLDS websites, see http://nces.ed.gov/programs/slds/stateinfo.asp). Researchers should then clarify the availability of critical data elements with SLDS staff, where possible. In cases where SLDS documentation is not published online, researchers should ask SLDS staff whether such documentation is available and can be shared with researchers. Important considerations include the availability and comparability of critical data elements over time and the level and distribution of missing data about these critical elements. In addition, some data elements that may be readily available in prepared research datasets, such as survey data available from the National Center for Education Statistics (NCES), may simply not be available in SLDSs.

Temporal limitations

Researchers should check what years of data are available in the SLDS. Some analyses may require more years of data than others, and some SLDSs may not contain enough years. Researchers should also be aware that time lapses between when data are collected and when they subsequently become available in the SLDS can limit available data. In some states, data are collected at key dates (such as once in the fall and once in the spring). The lapses between when agencies submit data and when it is posted to the SLDS can vary widely. In some states, data may be stored temporarily in an operational data store until the data are finalized and certified, but access to these temporary data varies, as does their quality.

Test information

Scores on state achievement tests are common data elements in SLDSs. In the NCES table in Appendix B, all but one state (West Virginia) reported maintaining yearly records of test scores in their SLDS. To meet federal accountability requirements, most states now assess students annually in English language arts and mathematics in grades 3 through 8 and at least once in high school. Some states also require passing exit exams for high school graduation. These are usually given in grade 10 or 11, though students may take the test as many times as they need to pass. Other states administer end-of-course assessments in high school and sometimes in middle school.

Test scores are a common outcome measure for many education interventions. But the state test scores in SLDSs may be limited in their usefulness for a specific research study. The annual nature of most state tests limits their usefulness for interventions that may be shorter. State tests may be given only in English language arts and mathematics. Test content may not reflect the focus of a specific curricular or instructional intervention. Test scores are typically not available before the third grade and may be available only for grade 10 or 11 in high school. And state tests may not be "vertically equated," which means scores from year to year do not directly measure student learning growth. Over time, tests can change in content, cutscores, and grades tested.

Where state tests are not well suited to be an outcome of interest due to their annual nature or generic content, researchers may want to investigate alternative assessments administered at the district level. The Northwest Evaluation Association's Measures of Academic Progress assessment may be used to address both timing and content problems. Measures of Academic Progress is administered by a number of districts across the country. It can be administered regularly as either a formative or summative assessment tool so that change can be measured over shorter periods of time. It also produces student growth scores—overall and for subscales—for each grade K through 12, with end-of-course tests in algebra 1 and higher-level mathematics. Some districts may use other commercial formative assessment instruments or may develop homegrown assessments, whose results could be merged with other SLDS data for a study.

Socioeconomic status

Some variables available in federal datasets may not be available in SLDSs. For example, common variables used to create socioeconomic status, such as parental education and family income, generally are not available.[22] SLDSs typically contain dichotomous measures that relate to student poverty, such as free or reduced-price lunch status or economic disadvantage status, or both. Analyses that require a continuous student-level socioeconomic status variable may not be possible in most states. The lack of a continuous socioeconomic measure could affect the feasibility of regression discontinuity designs. For other analyses, researchers will need to use categorical poverty variables. Continuous variables at the classroom or school level can generally be constructed.

The National School Lunch Program, administered by the U.S. Department of Agriculture, provides free and reduced-price lunch for eligible students. The relevant law strictly limits how eligibility and program participation information may be used. Student-level eligibility status may not be shared in most cases. States may, however, share student-level economic disadvantage status, where economic disadvantage is

determined by at least one criterion in addition to eligibility for free or reduced-price lunch. The No Child Left Behind Act of 2001 mandates that performance on state standardized tests for K–12 education include disaggregated results for economically disadvantaged students. In many states, eligibility for free or reduced-price lunch is used as the No Child Left Behind indicator for economic disadvantage. Other states, such as Texas and California, supplement eligibility with additional information for their definitions of economic disadvantage. Researchers wanting a student-level measure of poverty should clarify whether the state's economic disadvantage variable incorporates information in addition to eligibility for free or reduced-price lunch and can be shared for research purposes. See Cheung, Clements, and Pechman (1997) and visit the U.S. Department of Education's Privacy Technical Assistance Center (http://ptac.ed.gov/) for additional guidance.

Participation in the National School Lunch Program tends to fall as age increases, particularly among high school students, because of real and perceived stigma (Ralston, Newman, Clauson, Guthrie, & Buzby, 2008). This declining participation may result in underreporting of student poverty at the high school level. For high school analyses, capturing information about students' eligibility for free or reduced-price lunch or economic disadvantage status in grade 8 can improve the identification of economically disadvantaged high school students and improve the reliability of the measure (see Example 4). Eligibility for free or reduced-price lunch can change over time, however, as families move in and out of poverty.

Example 4. Defining economic disadvantage status

In the Tracking Postsecondary Outcomes for High Schools pilot project, the measure, "ever economically disadvantaged in grades 8–12," was created and defined as a high school student's economic disadvantage status in any grade 8 through 12. Economic disadvantage status was determined each school year, and the composite measure was constructed from multiple years of data.

Teaching experience

The measure of years of teaching experience, used in teacher value-added and other analyses, may not increase in annual increments as expected. Often this occurs when teachers are moved into a higher salary grade that requires a minimum number of years of teaching experience. Thus, actual years spent teaching may not align with the number of years of teaching experience recorded in state systems. For example, recorded values may be 1, 2, 3, 5, with the latter coinciding with an increase in salary grade for a teacher, while actual years of teaching experience may be 1, 2, 3, 4. Multiple years of data are needed to detect the problem and construct an alternate variable.

Clarify SLDS data access policy and state privacy law

Researchers should clarify the state's policy for access to SLDS data. The federal Family Educational Rights and Privacy Act (FERPA) prohibits disclosure of personally identifiable information from education records without written consent, except under the "studies" and "audit or evaluation" exceptions (see Appendix C). States interpret FERPA in different ways, however, and have their own privacy laws. Researchers should become familiar with privacy laws of states in which they are considering research. Some states publish their relevant laws on the state education agency or SLDS website.

Some states will release identifiable student-level data if data security and confidentiality protections are sufficient. Other states (such as Washington and Tennessee) will not release data elements considered to be personally identifiable information (such as student ID, name, and date of birth) but will release de-identified student data. A few states (such as Texas) will not release student data under any circumstances, but they allow researchers to work with the de-identified data on site in a central secured location.

If a state will not release personally identifiable information, researchers may need to explore whether the state can merge external data with SLDS data. Researchers should also clarify whether the analysis file will provide school and district identifiers. These identifiers make it possible to merge SLDS data with other data, such as CCD and IPEDS data.

Consider hardware and software needs

Depending on the size of the state and the nature of a data request, analysis may not be possible on a personal computer. The computer must have sufficient storage and sufficient memory to manipulate the data. Memory considerations are both a hardware and a software issue. Some popular analysis programs, such as Stata, store data in random-access memory, so the size of the data sets that can be analyzed is limited by the amount of machine memory. Other programs, such as SAS, use hard drive space as a temporary work area, but relying on standard hard drives for work space can slow processing considerably. Use of shared computer resources, such as a multiuser server, can compromise data security. Data security may require purchasing specific computer equipment or obtaining a locked area with limited access where the data will be housed and used.

5. Obtain the data

After confirming the feasibility of the planned study, researchers can take formal steps to obtain the data, including submitting a research request to and developing a data-sharing agreement with the state. Some states may require Institutional Review Board (IRB) review, and researchers who are affiliated with universities or other research organizations may also need to seek review by their IRB before research can be conducted.

Setting appropriate expectations

On average, data requests are filled in 4–8 months, but the process has been known to take 1–2 years. State staff may prioritize research requests that clearly address state research needs or are easier to fulfill. For example, preparing a custom data extract takes more time than preparing a "data dump." It is also easier to provide student data than to provide school data because of the structure of the SLDS data.

States are not required to share data (see Appendix C). Sharing SLDS data is voluntary, and fulfilling data requests is at the state's discretion. States that received American Recovery and Reinvestment Act funding for database development are committed to supporting research, but states are cautious about providing data for research on sensitive topics (such as vouchers or charters), especially when they are not certain how the researcher will handle the topic. Politically sensitive topics may go through a more involved review process and take longer.

Approaching state staff

Investing time in developing a partnership with state staff will facilitate responses to data requests. Demonstrating a thorough understanding of the SLDS data shows that researchers have done their homework, which will increase the researchers' credibility with state staff. If the researchers are not known to state staff, or if the research request does not align clearly with state goals, it may be helpful for researchers to find a state staff person who can advocate for them. One place to start is identifying state staff who have presented on SLDS research at NCES data conferences (see http://ies.ed.gov/whatsnew/conferences/).

Researchers could consider forming research collaborations, consortia, and alliances that create economies of scale and streamline data requests. For example, a single data extract could support multiple research efforts. While FERPA requires a data destruction date, data-sharing agreements can be amended if data are still being used in an active study.

Submitting a research request

Increasingly, states are using formal research request processes. There may be some back and forth with state staff to ensure that research resulting from the request is mutually useful. Some states are now charging fees to compensate for staff time to respond to data requests, which, depending on the magnitude of the data request, can run into tens of thousands of dollars. If researchers help address state data or research questions, however, they may not be charged.

Researchers should review the *Forum Guide to Support Data Access for Researchers* (National Forum on Education Statistics, 2012), which provides an overview of the data request process and includes sample templates for research requests and data-sharing agreements, among other useful information. A research request may include:

- Description of the proposed research.
- Description of the requested data.
- Benefit to the state and alignment with state goals.
- Research timeline.
- Requestor information, including a summary of qualifications.

After a preliminary request is submitted, researchers should discuss the feasibility of the proposed study with state staff, adjust the study parameters as necessary, and submit a final request. The subsequent state review process, which may involve an Institutional Review Board IRB), can vary widely from expedited to full review, depending on the nature of the data request and the sensitivity of the research topic.

Developing a data-sharing agreement

Once the final research request is approved by the state, a data-sharing agreement or Memorandum of Understanding (MOU) can be executed. Typically, state staff will initiate the process. While FERPA establishes minimum requirements for the disclosure of personally identifiable information (PII) from education records (see Appendix C), state requirements may be stricter. Regardless of whether a state releases PII, releases de-identified student data, or requires that the researcher work on site with de-identified data, the state will require a data-sharing agreement to protect the confidentiality of SLDS data.

Ultimately, states determine the requirements of agreements pertaining to use of data from their SLDSs, and legal staff may review agreements. Increasingly, states have a standard agreement template. Elements of a data-sharing agreement may differ between the "studies" and "audit or evaluation" exceptions to disclosing PII in FERPA (see Appendix C). The following list provides common elements of SLDS data-sharing agreements (see also National Forum on Education Statistics, 2012):

- Parties to the agreement or understanding.
- Purpose of the research.
- Definition of terms.
- Roles of the parties, including responsibility for fixing data errors. Some errors in source data, such as incorrect or missing IDs or other key data elements, will need to be fixed by state staff. Other missing data elements, such as composite or derived variables, may be constructed by the researcher. The data-sharing agreement should clarify expectations for data revisions and any costs associated with revising data files.
- Scope of work, including the data elements and years to be provided.
- Data use restrictions and ownership: state agencies own their data and have both the right and responsibility to know in advance how the data will be used in research and how results will be disseminated.

- Reporting requirements: for example, reporting results in the aggregate so that no students are identified; possibly also not identifying any schools or districts in results; possibly suppressing aggregate results based on small numbers.
- Data transfer and storage.
- Confidentiality and data security.
- Term of the agreement or understanding.
- Term or date of data destruction.
- Indemnification.
- Default and termination.
- Signatures.

Becoming familiar with Institutional Review Board reviews

As mentioned above, some states may require an IRB review of the proposed research. Research institutions may have their own requirements, and researchers affiliated with universities or research organizations may need additional review by their internal IRBs. The purpose of IRB review is to minimize potential harm to individuals participating in the research, such as harm that may result from disclosure of individually identifiable information. The state's review process and university IRB process may differ, and researchers will want to familiarize themselves with the various requirements. IRB review of subsequent amendments to the research is typically faster than initial review.

6. Investigate data quality

Investigating the quality of the received SLDS data is the next step in confirming the feasibility of a planned study. The quality investigation involves checking for valid values and variables in received files, compiling the data, performing external data checks, constructing derived variables, and conducting initial analyses. While these activities reflect general research practice, this section focuses on specific issues that may be encountered when using SLDS data. The process is likely to be iterative as researchers learn more about the data. Throughout, it is good practice to record the many data issues encountered and the decision rules developed to handle them. For additional tips for preparing the data, see SLDS Data Use Issue Briefs 3 and 4 (SLDS Grant Program, 2012b,c). Despite careful planning, data often bring surprises, and researchers may need to adjust their study plans after reviewing the data.

Documenting the decision rules

Researchers need to develop many business and decision rules to handle data conflicts and other SLDS data issues, and it is good practice for them to keep a record of these rules. It may be beneficial for researchers and state staff to establish a means of sharing these decisions, such as developing a research data dictionary that maps to the state data dictionary or creating forums or other means of sharing data information and code. Because staff may have limited time for providing iterations of the data, it is also good practice to make a plan with state staff for how data questions and problems will be handled.

Checking for valid values and variables

As a first step, researchers should check that the files received from state staff contain the expected variables and that these variables contain valid values. Problems can arise from miscommunication, human error, or data corruption. In some cases, variable coding practices may have changed from one year to the next. Early data validation can identify issues to be clarified with state staff and avoid later rework.

Compiling the data

Generally, SLDSs do not provide research-ready analysis files. An SLDS warehouse may contain annual data submissions that can be linked over time by common identifiers or other personally identifiable information (such as names and date of birth) but may not contain ready-made, student-level longitudinal records. The warehouse may contain separate stores for different types of data (such as data on student enrollment, demographics, attendance, assessment, coursetaking, completion and withdrawal, teachers, and separate state postsecondary education systems), often obtained from different source systems. Typically, states will provide researchers with separate data files for each year and data source in the SLDS that contains requested data.

A first step in analyzing student-level SLDS data is to construct one or more longitudinal, student-level analysis files from the annual records received from different sources in the SLDS.[23] Some data elements may not be organized on a student-by-year basis, however. Student discipline records, for example, may be organized by incident, which requires aggregating the incidents by student and year. Researchers may need to devote substantial time to creating the comprehensive longitudinal or panel analysis files needed for a

study from the multiple data extracts provided by state staff. Compiling these data often reveals data issues that must be resolved before proceeding, such as reconciling data that vary across time, data sources, and multiple records.

Researchers may want to confer with SLDS staff about any data issues encountered and then document decisions for handling the issues. Researchers may also need to merge external data with the SLDS data to create the data files needed for the planned analysis.

Reconciling data across time, sources, and multiple records

SLDS submission processes involve checks for valid data values and formats. Some states also have decision rules for reconciling data from different sources or time periods. Even so, some data inconsistencies may not be revealed until researchers attempt to combine and use SLDS data in new and different ways. Often, the farther back in time one goes, the more data issues arise. Some states (such as Colorado and Virginia) have "federated" warehousing structures, in which different source data systems may be linked only temporarily for specified purposes or requests.[24] Performing consistency checks on SLDS data values that appear for the same student in different source files or records can ensure accuracy and validity. In consultation with SLDS staff, researchers will likely need to develop decision rules for handling data values that vary across time and data sources and for handling multiple records. Examples of the types of decisions that researchers may need to make are described in the following sections.

Variability in data elements for the same students over time

Some data elements can be expected to vary over time for a student, such as eligibility for free or reduced-price lunch, English learner status, or other program participation status. In such cases, the research question may determine how the variability should be handled. For some studies, particularly correlational studies, program status at a precise time (such as the preintervention and postintervention periods) may be critical. For other studies, including descriptive studies, a student's ever having had a particular status, such as ever being eligible for free or reduced-price lunch or ever being an English learner, may be more important.

In contrast, some data elements may be expected to remain fairly constant at the student level, such as gender, race/ethnicity, and date of birth. In these cases, variability may be due to data errors or, possibly, actual change. It is also possible for variable coding to change over time, such as the coding of race/ethnicity categories. In these cases, state education agencies typically develop a crosswalk between old and new categories, where possible. Researchers will need to develop decision rules appropriate to the study (see Example 5).

Example 5. Handling expected and unexpected variability over time

Policymakers and practitioners often want to know outcomes for students who were ever classified as an English learner, and that classification can change over time as students acquire proficiency. For the Tracking Postsecondary Outcomes for High Schools pilot project, the composite measure, "ever an English language learner in grades 9–12," was created.

For the pilot project, values for gender, race/ethnicity, and date of birth varied over time in 1 percent or less of cases. For these students, the variable was assigned the modal value if one existed and the most recent value otherwise.

Inconsistency in data elements for the same students across data sources

Data elements may vary for the same students across data sources. For example, enrollment and assessment data files may indicate different schools of record for a student, possibly because of the different timing of source data collection. Depending on the purpose of a study, researchers may want to either assign students to a single school or retain the variability in school assignment (see Example 6). Other possible inconsistencies across data sources include different grade-level designations in enrollment and coursetaking files. In this case, it may be possible for a student to take a course that is primarily designated for students in a different grade. If so, the grade appearing in the enrollment file may be considered more reliable than that in the coursetaking file.

Example 6. Assigning students to a single school each year

For the Tracking Postsecondary Outcomes for High Schools pilot project, students were assigned to a single school in each academic year. If a student had enrollment records in more than one school in a given year but assessment (test) records for exactly one school in that year, the student was assigned to the school associated with the assessment record for that year.[25]

In a different study, researchers might want to identify all schools in which a student was enrolled in a given year. As stated in the preceding section, they should develop decision rules appropriate to the study.

Multiple records per student

SLDS data files may contain multiple legitimate records per student, different from duplicate records. Students who transfer between schools, for example, may legitimately have multiple enrollment records for a single year. Assessments may be administered more than once, especially if a student did not pass the test or attain proficiency on the first try, and assessment data files may include more than one record per student—one for each test administration where a student took the test. In such cases, researchers must develop rules to determine which of the multiple records to use (see Example 7).

Example 7. Handling multiple assessment records

In the Tracking Postsecondary Outcomes for High Schools pilot project, when multiple assessment records were encountered, the first record for each test was retained (that is, the result associated with the first time the student took the test was retained). Because the high school test was an exit exam, students were allowed to take it until they passed. Almost all high school graduates eventually scored proficient on the state test, and final proficiency attainment was associated with little variability in postsecondary education outcomes. In contrast, the first test result was associated with significant variability in outcomes for graduates and indicated a gap in postsecondary outcomes associated with early high school achievement. Although students may have eventually scored proficient, those who initially scored below proficient exhibited significantly lower postsecondary enrollment and persistence rates, even among students who scored similarly on grade 8 achievement tests. This finding generated discussion among stakeholders about factors contributing to initial tested achievement and how to improve it.

For correlational analyses, the test results most immediately preceding and following the intervention of interest may be of greater interest than the initial tested achievement that was a focus for the Tracking Postsecondary Outcomes for High Schools pilot project.

Merging data from external sources

Depending on the focus of a study, researchers may need to merge SLDS data with data from external sources to create the comprehensive data files needed, including data from teacher personnel or certification systems, locally developed assessments in grades not tested by the state (particularly grades K–2), college readiness exams (such as AP, SAT, and ACT exams, though SLDSs increasingly contain these data), and NSC data from private and out-of-state postsecondary institutions. To merge individual-level data from external sources, the state must either allow the release of individually identifiable records to the researcher, or state staff must be willing and able to conduct the merge on behalf of the researcher.

Researchers may also find it convenient to merge federal CCD and IPEDS data to obtain standard school and institutional characteristics. CCD is an annual data collection administered by NCES, which collects fiscal and nonfiscal data about all public schools, public school districts, and state education agencies in the country. It provides aggregate school-, district-, and state-level estimates of student demographic characteristics (such as percent minority and percent eligible for free or reduced-price lunch) as well as school size (enrollments) and urbanicity or locale codes. But those data may lag one or more years behind state data sources. To be able to merge the SLDS data with CCD data, it is important to request NCES school IDs from the state, in addition to the state's school IDs. Alternatively, school- and district-level demographic characteristics can be calculated directly from the student-level data.

IPEDS is an annual NCES data collection of every postsecondary institution in the country that participates in federal student aid programs. IPEDS surveys collect information about institutional characteristics such as location, level (4-year, 2-year, less than 2-year), control (public, private not-for-profit, private for-profit), price of attendance, enrollment, persistence, degrees awarded, and financial aid, as well as information about faculty, credit hours, and other aspects of postsecondary education. Note, however, that IPEDS classifications of postsecondary institutions may not match a state's classification of its institutions, and some adjustments may be needed to reflect state practice.

Conducting external data checks

Once an initial analysis file has been created, it is prudent to conduct external data checks to ensure that the compiled SLDS data align with published figures on basic measures. Descriptive research may receive attention from a broad audience, and school, district, and state staff may expect to see findings similar to those in other published data. Counts of students or graduates that do not match expectations can erode confidence in the data and results. States often publish, on state education agency websites, counts of enrolled students and graduates and rates of student proficiency on state tests. In addition, some public postsecondary systems and institutions publish data on overall enrollment, remediation, and completion. These public data can provide a useful external check for SLDS data (see Example 8).

Example 8. Performing external checks

External checks were performed for the Tracking Postsecondary Outcomes for High Schools pilot project against state-published figures and included

- Number of students enrolled in the state, by grade, in each year:

  - Overall.
  - By gender.
  - By race/ethnicity.
  - By eligibility for free or reduced-price lunch or economic disadvantage status.

- Number of high school graduates in the state in each year.
- Performance on state achievement tests by grade, including the percentage of students at each proficiency level:

  - Overall.
  - By race/ethnicity.

- Number of postsecondary students enrolled in state public institutions, by institution.
- Number of first-year postsecondary students enrolled in remedial math and English courses.
- Number of associate's and bachelor's degrees awarded, if using completion data.

Spot checks of published data for specific districts or schools were also conducted. Differences of 5 percent or larger were discussed with state agency personnel to determine possible sources of any differences and to confirm that the data received from the SLDS were complete and researchers were constructing the measures appropriately.

For additional examples of external data checks, see SLDS Data Use Issue Brief 4 (SLDS Grant Program, 2012b).

Constructing derived variables

After compiling the SLDS and external data into comprehensive longitudinal analysis files and conducting checks on these data against public sources, researchers typically construct derived variables from the administrative data (see Example 9). For example, researchers may need to

- Create variables for values that are captured only in certain grades (such as parental education, which may be obtained only at the point of entry into the education system).
- Create composite variables that summarize information from annual records (such as a student's ever being economically disadvantaged in grades 8–12).
- Transform variable values and categories (such as collapsing many race/ethnicity categories into a few standard categories).
- Create indicator variables (such as identifying high school graduates based on multiple withdrawal codes and creating graduate cohort flags).
- Assign "missing value" codes.

Performing initial analyses to investigate data quality

A next step is performing initial analyses to examine the completeness of important variables and patterns of missing data. Results of these analyses may lead to adjustments in the source and constructed variables planned for a study.

Completeness of the data

It is prudent to look at important variables—overall and by year—and intended subsets of observations (such as targeted cohorts, intervention and comparison groups, and subgroups) to check the data's completeness. Significant amounts of invalid values or missing data on important variables could jeopardize the validity and reliability of a study or create small study samples that add to sampling variance and limit the generalizability of findings (see Example 10).

Correlational and causal methods discussed earlier require reasonably complete data about key variables in pre- and post-intervention years. Regression discontinuity requires completeness on the assignment variable and on the characteristics of students above and below the cutpoint. Where important variables have significant amounts of invalid or incomplete data on any of the explanatory variables, results may be biased. The NCES 70-percent standard for response rates may be too low to produce reliable results for some correlational analyses (Seastrom, 2002).

Distribution of missing data

To identify nonrandom missing data, it can help to run frequency distributions on critical variables for each year and cohort, overall and by school district or school. Missing data are more common for SLDS data elements that are not required for federal or state reporting or accountability, and the amount of missing data may vary by school district and year (see Example 10). When large amounts of data are missing (such as for an entire school or district, postsecondary institution, or academic year), it is critical to report it and potential biases in order to facilitate appropriate interpretation and build confidence in the analysis. For

descriptive analyses, researchers may want to follow NCES recommendations to conduct a nonresponse bias analysis when a unit or item response rate is less than 85 percent (Seastrom, 2012).

Example 10. Defining a complete transcript

A critical step in any high school coursetaking analysis is to define a complete transcript, typically on the basis of state-specific high school graduation requirements. For the Tracking Postsecondary Outcomes for High Schools pilot project, high school transcripts were compiled from annual course records in the SLDS. Because of large amounts of missing data in noncore courses, coursetaking measures were constructed for all high school students with complete transcript records in core subjects only (English, mathematics, science, and social studies). Missing data tended to be nonrandom and were concentrated in earlier years and certain school districts because of varying course data submission practices over time and across districts. Where missing data exceeded 30 percent statewide in a given year, the entire academic year's course records were suppressed statewide. This affected transcript records for cohorts for which one or more of their four high school years contained a suppressed year, and it resulted in suppression of the entire transcript record for earlier cohorts. In addition, any school district with missing data exceeding 30 percent in a given year had that year's data suppressed, affecting multiple cohorts with one or more of their four high school years containing a suppressed year.

Check for sample attrition

When data are used from multiple years and data sources within the SLDS, the sample that has valid values on all key variables may be significantly smaller than the population of interest. It is important to compare the characteristics of the final study sample with the relevant population. If possible, conduct an attrition bias test. See SLDS Data Use issue Brief 3 (SLDS Grant Program, 2012c).

Addressing bias due to missing data

Missing data may be problematic in descriptive analyses, where the data can significantly affect school or district means. For correlational and causal analyses, missing data are problematic when they affect the treatment and control groups differentially. A number of treatments may be used to address bias due to incomplete or missing data. Useful resources include Puma, Olsen, Bell, & Price (2009); Enders (2010); Graham (2012); Little, D'Agostino, Cohen, Dikersin, & Emerson (2012); van Buuren (2012); and Little (2013).

Confirming SLDS study feasibility

After conducting initial analyses on the constructed files, it is important to decide whether the SLDS data are of sufficient quality to support the planned study. If SLDS data quality is problematic, researchers may want to modify their research questions or analysis methods. It may be necessary to scale back a correlational analysis to focus on descriptive analysis only. Or it may be necessary to narrow a descriptive analysis to a smaller set of observations or variables that have complete and valid data. After assessing whether the planned study is feasible, it is important to update the analysis plan to describe any data limitations revealed and to clarify the appropriate inferences that can be drawn from the research. Researchers may want to share their revised plans with state staff and discuss what they have learned about the SLDS data.

7. Communicate results

After planning, refining, and carrying out a study that uses SLDS data, the final step is to communicate results. A theme throughout this guide is the importance of building relationships between researchers and state staff around using SLDS data. How researchers handle the communication of study results can help build the trust of state staff and increase the face validity of the research among stakeholders. Consideration of the intended audience is also important. Recommendations for how to present analysis findings in a clear and transparent way are provided in Step 3.

State agencies own their data and have both the right and responsibility to know in advance how the data will be used and how results will be disseminated. States take a risk in releasing data to researchers without knowing with certainty what analyses can be performed or what modifications researchers will need to make to their analysis plans because of data limitations. Researchers will need to (a) balance concerns about academic freedom with state interests when developing an agreement early in the research process on how results will be used and (b) plan time for state review before the results are published. While the federal Freedom of Information Act applies only to data collected by the federal government, many states have passed their own open records laws that may apply to SLDS research. Researchers should be aware of state open records laws and the implications for public inspection of research results. They should listen to staff concerns before communicating results to other audiences, such as the media. State staff can help researchers identify questions that state stakeholders may have about the methods and results and help catch problems that may undermine the validity of the research. Transparency at every step of the research is best so that state staff and other stakeholders will trust the results.

8. Checklists for the seven steps in using SLDS data

Researchers who use SLDS data are often exploring new territory, particularly when studying topics that link data across education sectors and multiple data sources or when conducting SLDS research in a new state. While most research studies may need periodic course corrections, studies involving SLDS data may be more likely to require such modifications. Throughout a study using SLDS data, researchers should be prepared to adjust their study plans as new data surprises and limitations arise. This section provides checklists for each of the seven steps for using SLDS data for applied research. These steps reflect good general research practice and also highlight the specific opportunities and limitations of using SLDS data. Reviewing the seven steps before undertaking SLDS research can alert researchers to potential pitfalls and help avoid some surprises.

Step 1. Understand the potential (and limits) of using state longitudinal data for applied research. SLDSs contain longitudinal, universal, and comprehensive data, the combination of which provides a unique opportunity for education research. And SLDS data can often be linked to external data sources to supplement the SLDS data. But many of these systems are still in development, and the administrative data were not originally designed for research. Researchers may want to obtain data codebooks or dictionaries from state staff or websites, which describe in detail the data elements in the SLDS (to access SLDS websites for states, see http://nces.ed.gov/programs/slds/stateinfo.asp).

1.A. Check the state's progress and plans.

- Visit the SLDS Grant Program website: http://nces.ed.gov/programs/slds/stateinfo.asp.
- Visit the Data Quality Campaign website: http://www.dataqualitycampaign.org/your-states-progress/.

1.B. Consider the features of SLDS data.

- Longitudinal.
- Universal.
- Comprehensive.

1.C. Consider the possibilities for linking to external sources.

- Data sources maintained locally.
- Out-of-state data.
- Federal and national data.

1.D. Plan to collaborate with state staff.

- Read the Forum Guide (National Forum on Education Statistics, 2012) sections on productive data partnerships between researchers and state staff.
- Plan to consult with state staff at each Step 2–7, as appropriate.
- Identify busy times for state staff during the year.

- Consider ways to help the state, such as providing feedback on the SLDS data or communicating about results.

1.E. Take advantage of SLDS resources and publications.

- http://nces.ed.gov/programs/slds/resource_type.asp.
- http://nces.ed.gov/programs/slds/publications.asp.
- http://nces.ed.gov/forum/index.asp.

Step 2. Identify research questions and determine whether SLDS data are appropriate for addressing them.
SLDS data can support a broad range of descriptive, correlational, and causal research questions. So to design a high-quality research study that uses SLDS data, researchers should determine whether SLDS data are appropriate for addressing the research questions identified by researchers and state staff.

2.A. Develop a research partnership with state staff.

2.B. Start with a research agenda.

- Select a high-priority research topic.
- Discuss with state staff how to frame the research to be beneficial both to state stakeholders and the researcher.
- Develop a set of research questions that cover the full range of research concerns related to the topic.
- Identify questions that require original data collection.
- Identify questions that require a causal research design. Consider whether SLDS data can be used to address related descriptive or correlational questions as a first step before attempting a causal design, or whether SLDS data can supplement such a design.
- Identify any additional research questions that may be appropriate for using SLDS data.
- Add to the research agenda any additional questions identified.

2.C. Identify high-priority research questions that may use SLDS data.

- Based on the full set of research questions developed in Step 2.B, identify high-priority descriptive questions that:

  ○ Describe status or trends in educational conditions or outcomes.
  ○ Describe populations of interest.
  ○ Compare schools, districts, or other organizational units.

2.D. On the basis of the full set of research questions developed in Step 2.A, identify high-priority correlational and causal questions that seek to clarify or determine the relationship among outcomes and contributing factors.

2.E. Make an initial determination of whether SLDS data are an appropriate data source for addressing the research questions.

- Are longitudinal data that track changes over time for individual students or teachers needed to address the high-priority research questions?
- Are trend data—particularly annual trend data—needed to describe educational conditions or outcomes or track multiple cohorts?
- Are statewide data needed to:

  ○ Describe particular policies, practices, or outcomes of interest to the state?
  ○ Analyze all students, schools, districts, or other entities in the state?
  ○ Draw sufficient samples of populations or programs of interest?
  ○ Construct comparison groups?
  ○ Investigate the contribution of different education levels (such as classroom, school, and district) to outcomes?

2.F. Specify initial data needs and questions.

- Make an initial determination of the types of data that are needed to address the high-priority research questions. Be as specific as possible. The following types of data may be contained in SLDSs:

  ○ Unique student identifier, for purposes of linking data over time or across data sources.
  ○ K–12 student data (such as student demographics and other characteristics, enrollments, attendance, completion and withdrawal, state test scores, college readiness exams, coursetaking, grades, credits, and so on).
  ○ K–12 school, district, or other aggregate characteristics.
  ○ Teacher data (such as demographics, certification and licensure, years of experience, course and section assignments, compensation, and professional development activities, among other data).
  ○ Early childhood education data (student and program).
  ○ Postsecondary education data (student and institution).
  ○ Workforce data.
  ○ Other sectors and types of data.

- Indicate where high-priority research questions require linking data across the data sources listed above.
- List specific SLDS data elements that are considered critical for addressing the high-priority research questions.
- Identify important derived variables to be constructed from source data elements; verify that all component data elements are on the list of critical SLDS data elements.
- Specify the number of years or timing of data needed for each critical data element.
- Are critical data elements likely to be common statewide elements that are included in the SLDS? Are they unique to a locale or program? Do they require supplementing SLDS data with other data sources?

- List any initial questions that you may have about the availability or characteristics of needed data sources and elements in the SLDS.

Step 3. Select appropriate methods for the research. SLDS data can support a variety of analytic methods for addressing descriptive, correlational, and causal research questions and can be used to support exploratory analyses, randomized studies and natural experiments. Researchers should select the most appropriate methods for addressing their research questions and then specify any data assumptions and requirements related to the methods.

3.A. Addressing descriptive research questions.

- For each high-priority descriptive research question, identify the methods that you intend to use to address the question, such as:

  - Cohort analysis.
  - Trend analysis.
  - Gap analysis.
  - Tests of difference.
  - Banding of similar schools.
  - Residuals analysis.
  - Other analysis.

- Specify at what level of aggregation the results will be reported, such as at the state, district, or school level. Consider whether any similar results may have been previously published and identify possible external sources against which you can check your results.
- Determine state requirements or practices for suppressing results on the basis of small numbers to protect confidentiality—for example, state practices that typically require minimum cell sizes of 5 or 10 students. See SLDS Technical Brief 3 (Seastrom, 2010) for a discussion of methods to protect personally identifiable information in aggregate reporting.
- Identify the intended audiences for the analysis results. Consider whether to engage these audiences in a discussion of how best to present results, including:

  - Whether to set higher minimum cells sizes for reporting results for schools, districts, or subgroups.
  - How to handle missing data (see section on missing data below).
  - How to present results to make them most useful for the intended audiences, including types of presentations (tables, graphs) and supporting information (such as Ns, definitions, labels, text interpretations of tables and graphs, prompting or discussion questions, and detailed technical information).
  - Whether and how to engage stakeholders in a review of research plans and initial results.
  - How to disseminate final results to the intended audiences.

- Review the list of critical SLDS data elements and identify any additional data elements needed for the selected analyses.

- Indicate where analyses may require linking data across different sources or sectors contained in the SLDS.
- Consider how missing data on the critical elements may affect the planned analyses, including:

  - At what level does random missing data become problematic? The NCES 70-percent standard for response rates (Seastrom, 2002) may allow for too much missing data for some audiences and purposes. If high stakes are associated with analysis results, such as publishing results for identified schools or districts or basing policymaking or funding decisions on the results, then the bar for missing data may need to be set higher.
  - How should nonrandom missing data be handled? Make a plan for investigating the distribution of missing data on key data elements or variables, for example:

    - Investigate missing data by reporting source. For example, while test results are typically provided by the state education agency, coursetaking may be reported by districts to the state, and postsecondary data may be reported by system or institution.
    - Investigate missing data for the level of aggregation at which results will be reported.
    - Identify derived variables that must be constructed from multiple years or data sources and investigating the level of missing data in each component year or source.

- Identify any SLDS data elements for which availability or quality determines the feasibility of the planned approach.
- Update your list of questions about the availability or characteristics of SLDS data elements needed for the selected descriptive analyses.

3.B. Addressing correlational and causal research questions.

- For each high-priority research question, identify the methods that you intend to use to address the question, such as:

  - Regression discontinuity.
  - Difference-in-differences.
  - Fixed effects.
  - Hazard analysis.
  - Propensity score matching.
  - Other matching methods.
  - Interrupted time series.
  - Multilevel modeling.
  - Instrumental variables.
  - Other method.

- Review the list of critical SLDS data elements and identify any additional data elements needed for the selected approach.
- Indicate where analyses may require linking data across different sources or sectors contained in the SLDS.

- For the critical data elements identified, consider the following questions:

  ○ How many years of data are needed for the selected approach?
  ○ How many years of comparable data are needed, particularly for outcome variables?
  ○ What periodicity is needed? Are annual observations sufficient?
  ○ What timing is needed for the different data elements, including the timing of pre- and postintervention measures (and lapsed time between the two)?
  ○ What data elements must be continuous versus categorical?
  ○ What levels or distribution of missing data can be tolerated (see 3.A above)?

- Determine the optimal size of any intervention and comparison groups needed to produce reliable results; estimate the size of such groups that is likely to be available in the SLDS data, given what is known about the state education system and population.
- Identify any other specific data assumptions for the selected method.
- Identify any SLDS data elements for which availability or quality determines the feasibility of the planned approach.
- Update your list of questions about the availability or characteristics of SLDS data elements needed for the selected correlational analyses.

3.C. Using SLDS data in exploratory analyses. Determine what exploratory data analyses, which may use descriptive or correlational methods, can be useful to states and preliminary to further research.

3.D. Using SLDS data in randomized controlled trials and natural experiments. Determine how SLDS data can be used to set up these studies or supplement the data that are collected for them.

3.E. Applying general cautions when analyzing SLDS data. General cautions when using the preceding methods include variability in data collection practices, attrition of the study sample due to missing or invalid data, incompleteness or inaccuracy of older data, and variability in the timing and frequency of data collection for common or important data elements.

Step 4. Confirm data availability. The feasibility of a planned study depends in part on the availability of the SLDS data needed for the study, including the availability of broad content areas (listed in 4.A), the capacity to link across these areas, and the specific data elements. Because the data in an SLDS cannot support all planned studies, researchers should make an initial determination that the data needed for the planned study are available in the SLDS.

4.A. Confirm SLDS content and capacity.

- Review the types of data needed to address high-priority research questions, including:

  ○ Unique student identifier.
  ○ K–12 student data.
  ○ K–12 school, district, or other aggregate characteristics.
  ○ Teacher data.

- ○ Early childhood education data.
- ○ Postsecondary education data.
- ○ Workforce data.
- ○ Other sectors and types of data.

- Review any initial questions you had about the availability or characteristics of needed data sources in the SLDS.
- Review where priority research questions and selected methods require linking across SLDS content areas.
- To determine the specific content and capacity for linking across content areas of an SLDS, visit the Data Quality Campaign (http://www.dataqualitycampaign.org/your-states-progress/ and http://nces.ed.gov/programs/slds/) websites as starting points. See also Appendix B. Remember that the information provided on these sites may mask significant variation in state practice. Additional investigation is necessary to understand actual state capacity.
- Review available SLDS documentation on the content and capacity to link across content areas; follow with conversations with SLDS staff, where possible.
- Consider the following questions, as appropriate to the planned study:

  - ○ Does the state use a unique P-20/W student identifier? How reliable is it? For how many academic years is this available in the SLDS? If not available, how are student-level data linked across data sources?
  - ○ If the answer to the preceding question is no, does the state use a unique K–12 student identifier? How reliable is it? For how many academic years is this identifier available in the SLDS? If not available, how are K–12 student-level data linked over time?
  - ○ Can K–12 teacher data be linked directly to students and classrooms? If not, how else can these data be linked?
  - ○ Can K–12 student data be linked directly to postsecondary education data? If not, how else can these data be linked?
  - ○ Are all state postsecondary systems and institutions represented in the SLDS? Are private and out-of-state institutions represented? Can these data be linked to K–12 student data?
  - ○ Can early childhood education data be linked directly to K–12 student data? If not, how else can these data be linked?
  - ○ Are all early childhood programs in the state represented in the SLDS? If not, what programs are represented?
  - ○ Can other specific links that are needed for the planned study be made?
  - ○ Are other data sources needed for the planned study available in the SLDS?

4.B. Confirm availability of specific data elements.

- Review the list of critical data elements needed to address priority research questions and selected methods.
- Review any initial questions you had about the availability or characteristics of these data elements.

- Review available SLDS documentation about the data elements; follow with conversations with SLDS staff, where possible.
- Consider the following questions related to critical data elements needed for the planned study:

  - Does the SLDS contain the critical data elements needed for the study?
  - For what academic years are the critical data elements available?
  - Are the critical data elements comparable over time?
  - Are any of the data available more frequently than once a year?
  - What proportion of students has valid values on the critical data elements? How does this vary over time?
  - What is the level and distribution in the state of missing data on the critical data elements? Do missing data vary by district or school? How do missing data vary over time? Are any variables particularly problematic?
  - When do SLDS data for the most recent academic year become available for research?

4.C. Confirm study feasibility.

- Make an initial determination of whether the available SLDS data are sufficient to carry out the planned research.
- Revise the research questions and analytic methods as necessary, on the basis of the available SLDS data.
- Identify any important gaps in SLDS data sources or elements for conducting the planned study that could be supplemented with external data sources.

4.D. Clarify SLDS data access policy and state privacy law.

- Does the state release identifiable student-level data? If yes, under what FERPA exception(s) does the state release these data?
- If not, does the state release de-identified student-level data?
- If not, under what conditions does the state allow researchers to analyze the SLDS data?
- If the state does not release identifiable student-level data, can state staff merge external student-level data with the SLDS data before making the data available for research? Under what circumstances is this possible?

Step 5. Obtain the data. After identifying research questions and appropriate methods that use SLDS data and after making an initial determination that available SLDS data would support the planned research, researchers are ready to take formal steps to obtain the data. During the research request process, which can take months, researchers may further refine their study design.

5.A. Submit a research request.

- Does the state have a formal research request process?
- If yes, does the state have a standard form for submitting a research request to use SLDS data? How can the form be obtained? To whom should it be submitted?

- If no, what process is used for approving research requests?
- How long does the process usually take?
- Are there any times of year when it is better or worse to submit a research request?

5.B. Execute a data-sharing agreement or Memorandum of Understanding (MOU).

- Does the state have a standard data-sharing agreement for research that uses SLDS data?
- If so, is a data-sharing template or sample agreement available?
- How is the data-sharing agreement process initiated?
- Who must be a party to the data-sharing agreement?
- How long does the process usually take?

5.C. Obtain an Institutional Review Board (IRB) review.

- Does the state require IRB review of the proposed research?
- Does the researcher's organization require IRB review?
- How do the requirements differ?
- What is the anticipated timeline for review?

5.D. Clarify data formats.

- In what format will released data be provided to the researcher?
- Are student-level longitudinal records available, or will different academic years be provided in separate files?
- Will data from different sources be provided in separate files? If yes, what are some examples of the different sources and files?
- Will state staff merge any data sources or years before providing the data to the researcher?

Step 6. Investigate data quality. After obtaining the data, investigating their quality is a critical last step in confirming the continuing feasibility of a planned study that uses SLDS data. Investigating SLDS data quality includes compiling the data, performing external data checks, constructing derived variables, and conducting initial analyses. While these activities reflect general research practice, this guide focuses on specific issues that may be encountered when using SLDS data. The process is likely to become iterative as researchers learn more about the data. Throughout, it is good practice to record the many data issues encountered and decision rules developed to handle them. Despite careful planning, data often bring surprises, and researchers may need to adjust their study plans after reviewing the received data.

6.A. Check for valid variables and values. Check that the files received from state staff contain the expected variables and that these variables contain valid values.

6.B. Compile SLDS data.

- If longitudinal student-level data are not provided by the SLDS and researchers must construct the longitudinal analysis file(s), begin by creating an inventory of the different SLDS data files provided

for each academic year (such as student enrollment, demographics, attendance, assessment, coursetaking, completion and withdrawal, teacher assignments, teacher credentials, and separate state postsecondary education system files).

- Compiling SLDS data into longitudinal analysis files requires establishing rules for making decisions. Where possible, researchers should consult with state staff about any data inconsistencies that are found, the potential sources of these inconsistencies, and the best approaches to handling them.

- Creating a longitudinal student-level analysis file requires linking data across time and sources:

    ○ Determine what student and other unit-level identifiers are available in each data file (such as student ID, name, and date of birth).

    ○ Where IDs are available, verify the quality of linked IDs by checking the name and birthdate associated with the ID from each year and source. Where an ID is associated with more than one student, consider selecting the modal student's data (or the most recent or earliest student, depending on research interest) and set other data associated with the ID to missing.

    ○ Where names are not available (possibly for older data), verify the quality of linked IDs by checking the birthdate associated with the ID from each year and source. Where an ID is associated with more than one birthdate, consider selecting the modal student's data (or the most recent or earliest student, depending on research interest) and set other data associated with the ID to missing.

    ○ Some states may not provide personally identifiable information, and researchers will need to use the study IDs provided for linking over time and across source without the ability to check associated names and birthdates.

    ○ Where data sources use different ID systems or where a large number of students cannot be linked (such as across K–12 and postsecondary education data sources), use names, birthdates, gender, and race/ethnicity to link across the data files. For students linked this way, consider creating an ID crosswalk for possible future use.

- Handling variability in characteristics for the same students over time:

    ○ For characteristics that may be expected to vary over time (such as eligibility for free or reduced-price lunch, English learner status, or other program participation status), researchers should determine whether they need to retain the variability or create one or more composite variables to address their specific research question. For some studies, the precise timing of a status may be important, while for other studies, having ever held a particular status may be important.

    ○ For characteristics that may be expected to remain fairly constant over time at the student level (such as gender, race/ethnicity, and date of birth), variability may be due to data entry errors or, possibly, to actual change. It is also possible for variable coding to change over time. In these cases, researchers could assign the modal value (if one exists) or the most recent or earliest value otherwise, depending on their specific research question.

- Handling inconsistency in data elements for the same students across sources:

- ○ Some studies may require assigning students to schools and districts. Data sources for the same academic year, such as enrollment and assessment files, however, may indicate different schools of record for a student, possibly due to the different timing of source data collections. Researchers should determine whether they need to retain this inconsistency (or change) or select the school associated with a particular point in time in order to address their specific research question.
  - ○ For each inconsistency found across data sources (such as different grade-level designations in enrollment and coursetaking files), researchers will need to determine whether to retain the inconsistency or select a value based on timing or the reliability of the data source.

- • SLDS data files may contain multiple legitimate records per student, because of, for example, student mobility or multiple test administrations. As in the preceding items, researchers should determine whether they need to retain this variability or create one or more composite variables to address their specific research question.
- • Merging data from external sources:

  - ○ Researchers may need to merge SLDS data with data from external sources to fill gaps in data needed for a particular study.
  - ○ Review the list of external data sources identified in Step 1.C.
  - ○ See guidelines for linking student- and other unit-level data in Step 6.B.
  - ○ If an SLDS will not release personally identifiable information, the SLDS staff may need to merge unit-level data from external sources with the SLDS data to assign a study ID before releasing the data to researchers. This role for SLDS staff should be specified in the data-sharing agreement.
  - ○ Researchers may also want to merge school- or district-level data or other aggregate data from external sources, such as CCD or IPEDs. For example, the CCD often contains the state-assigned school and district IDs, and some state data files contain the NCES school and district IDs, making it fairly easy to merge these data. It is important to request NCES school IDs from the state, in addition to the state's school IDs. Researchers should verify the quality of linked IDs, however, by checking the school names and possibly other information (such as school type, enrollment or size, and grades served) associated with the ID from each year and source. School and district data may vary over time as some become authorized, de-authorized, or combined with other entities. In addition, states may have a list of the IPEDS IDs for each state postsecondary institution.

6.C. Conduct external data checks.

- • After compiling the longitudinal analysis file in Step 6.B, researchers should conduct initial consistency checks for common measures against state-published sources. Common measures include:

  - ○ Number of students enrolled in the state, by year, grade, gender, race/ethnicity, and eligibility for free or reduced-price lunch or economic disadvantage status.

- ○ Performance on state achievement tests, including the percentage of students at each proficiency level, by subject, year, grade, and race/ethnicity.
  - ○ Depending on the focus of the SLDS study, researchers may want to check other important cohorts or measures as well (such as counts of high school graduates or postsecondary education students).

- State-published data may be found on state websites or postsecondary system or institutional websites.
- Before the analysis proceeds any further, finding data discrepancies can help identify missing data or data-handling procedures that were inconsistent with state practices. Researchers may want to investigate differences of about 5 percent or greater and discuss any such discrepancies with SLDS staff or source system staff.

6.D. Construct derived variables.

- To investigate the quality of the received SLDS data, it is critical to construct all important derived variables needed for the planned study from the available SLDS data elements to ensure that the quality of these variables will support the study.
- Review the list of critical data elements and derived variables needed for your high-priority research questions that you identified in Step 2.D.
- Review the list of critical data elements and derived variables needed for the research methods that you identified for descriptive and correlational analyses in Step 3.
- Construct the necessary derived variables. Refer to the suggestions in Step 6.B. for handling data variability and inconsistency over time and across sources in constructing the variables.

6.E. Perform analyses to investigate data quality.

- Examine the completeness of critical variables and patterns of missing data.
- Review the data issues and questions that you identified in the checklists in Steps 2–4.
- Consider the following questions related to completeness of the variables identified in Steps 2–4 as critical to the study and their patterns of missing data:

  - ○ Are a sufficient number of years available for the critical variables?
  - ○ Are these variables comparable over time?
  - ○ Is the timing and periodicity of the variables appropriate? Do you have sufficient pre- and postintervention measures, as applicable?
  - ○ Are needed continuous and categorical variables available?
  - ○ Is the size of needed cohorts, subgroups, or intervention and comparison groups sufficient?
  - ○ What proportion of students has valid values on the critical data elements? How does this vary over time?
  - ○ What is the level and distribution in the state of missing data for the critical data elements? How do missing data vary over time? Do missing data vary by district or school? Are there any variables that are particularly problematic with regard to missing data? A number of treatments may be used to address bias that is due to incomplete or missing data.

- - Do the critical variables meet all other data needs or assumptions that you identified in Steps 2–4?

- Review the important data elements and variables that must be linked for the study, as identified in Steps 2–4:

  - Perform any important preliminary linkages, such as linking teachers to specific students or classrooms or both.
  - Cross-tabulate important related variables.
  - Examine the level of complete or missing data or both on these combined variables.
  - Is the size of groups with complete data about important combinations of variables sufficient?
  - What are the level and distribution of missing data on these important combinations of variables?

6.F. Confirm SLDS study feasibility.

- Decide whether the received SLDS data and resulting variables are of sufficient quality to support the planned study.
- If SLDS data quality is problematic, consider modifying the research questions or methods to align with the quality of the available data. Consider scaling back a correlational analysis to focus on descriptive methods only or narrowing a descriptive analysis to a smaller set of observations or variables that have complete and valid data.
- Update the analysis plan to describe any data limitations that were revealed and to clarify the appropriate inferences that can be drawn from the research.
- Consider sharing the revised plan with SLDS staff and discuss lessons learned about the SLDS data so far.
- As the study proceeds, be prepared to continually refine the analytic approach to reflect any data constraints that are revealed.

Step 7. Communicate results. After planning, refining, and carrying out a study that uses SLDS data, the final step is communicating and reviewing results with state staff.

- How will results of the proposed research be used?
- Who are the intended audiences for the research?
- When and how will state staff review the results? How can the transparency of the research be ensured?
- What state open records laws apply to SLDS research?
- What concerns do state staff have about the results?
- What questions do state staff think stakeholders may have?
- What problems may threaten the face validity of the research among stakeholders?
- What issues does the researcher need to investigate before releasing results?

Notes

1. The three states not receiving grants are Alabama, New Mexico, and Wyoming.

2. Generally, SLDSs do not provide research-ready analysis files. Researchers will have to construct the analysis file needed for a particular study. This topic is discussed in greater detail in Step 6.

3. For example, race/ethnicity categories may change. State tests may change in terms of content, proficiency cutpoints, and grade levels tested. In addition, some data elements may be added or dropped over time.

4. Coverage of individual public schools and public school districts may vary over time as some become authorized, de-authorized, or combined with other entities. A few SLDSs also include private school data.

5. Student identifiers are unique within K–12 education but may not be unique across other data sources. SLDS linking practices and use of identifiers are discussed in detail in Step 4.

6. Students may complete school by graduating with different diploma types, receiving a certificate of completion, or attaining another completion status. Depending on the state, completion and withdrawal codes can be numerous.

7. These data at the student level are becoming more common.

8. The 18 states are Alaska, Arkansas, Delaware, Florida, Indiana, Kansas, Kentucky, Maine, Maryland, Missouri, North Carolina, North Dakota, Rhode Island, Tennessee, Texas, Utah, Virginia, and Washington. For up-to-date information, see http://www.dataqualitycampaign.org/your-states-progress/10-state-actions?action=one.

9. While many states can link K–12 data systems to early special education and state prekindergarten programs, far fewer can link to other early childhood education programs. This linking capacity is discussed further in step 4.

10. While a data element (such a student's race/ethnicity or date of birth) may be common across the state, its specific format and values may vary by school district, and the SLDS may need to standardize the data elements obtained from local student information systems.

11. Some SLDSs contain data on remedial courses taken but not on the need for remediation. The research question may need to be reframed in terms of coursework taken or researchers may need to obtain institution-level placement data to discern need.

12. Some states and school districts may include, for example, charter schools or summer graduates in addition to regular schools and graduates in their public or accountability data. Excluding them from an SLDS study could produce different results from published data and undermine confidence in the analysis.

13. Some government networks may not allow access to this site.

14. The Institute of Education Sciences' What Works Clearinghouse (WWC) publishes guidelines for determining whether causal studies meet scientific evidence standards. For a detailed discussion, see National Center for Education Evaluation (2014).

15. For background on regression discontinuity, see Imbens and Lemieux (2008).

16. For background on difference-in-differences methods, see Angrist and Pischke (2008).

17. For background on fixed effects method, see Allison (2009).

18. For background on hazard analysis, see Allison (2003).

19. For background on propensity score matching, see Cook and Wong (2008).

20. See http://www.dataqualitycampaign.org/your-states-progress/10-state-actions?action=one.

21. See http://www.dataqualitycampaign.org/your-states-progress/10-state-actions?action=one.

22. NCES longitudinal surveys—for example, the High School Longitudinal Study, the Education Longitudinal Study, the National Education Longitudinal Study, and High School and Beyond—include these variables. SLDSs typically do not.

23. Researchers may need to create different analysis files to address different research questions. In addition, some analyses may be performed at the teacher or school level.

24. If these links are frequently needed, a special purpose "data mart" or operational data store may be developed.

25. In a number of states, student test results will be reported for a school only if the student attended the school for a minimum number of days.

References

ACT, Inc. (2006). *Reading between the lines: What the ACT tells us about college readiness in reading.* Iowa City, IA: Author

Adelman, C. (1998). *The kiss of death? An alternative view of college remediation.* National CrossTalk. Retrieved April 11, 2013, from http://www.highereducation.org/crosstalk/ct0798/voices0798-adelman.shtml

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Alliance for Excellent Education. (2006). *Paying double: Inadequate high schools and community college remediation* (Issue Brief). Washington, DC: Author.

Allison, P. (2003). *Survival analysis using SAS.* Cary, NC: SAS Institute Inc.

Allison, P. (2009). *Fixed effects regression models* (Quantitative Applications in the Social Sciences No. 160). Thousand Oaks, CA: Sage Publications.

Andrews, R., Li, J., & Lovenheim, M. (2012). *Heterogeneous paths through college: Detailed patterns and relationships with graduation and earnings* (Working Paper No. 83). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Attewell, P., Lavin, D., Domina, T. & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education, 77*(5), 887–924.

Bailey, T. (2008). *Challenge and opportunity: Rethinking the role and function of developmental education in community college* (CCRC Working Paper No. 14). New York: Columbia University, Teachers College, Community College Research Center.

Barnett, W. S., Jung, K., Youn, M.-J., & Frede, E. C. (2013). *Abbott Preschool Program longitudinal effects study: Fifth grade follow-up.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

Bifulco, R. (2012). Can nonexperimental estimates replicate estimates based on random assignment in evaluations of school choice? A within-study comparison. *Journal of Policy Analysis and Management, 31*(3), 729–751.

Bornfreund, L., & Severns, M. (2010). *Many missing pieces: The difficult task of linking early childhood data and school-based data systems.* Washington, DC: New America Foundation.

Boylan, H. R., & Bonham, B. (1992). The impact of developmental education programs. *Research in Developmental Education, 9*(5), 1–4.

Braley, R., & Ogden, W. (1997). When failing indicates higher graduation potential. *College Student Journal, 31*(2), 243–250.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (Working Paper No. 66). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Burke, M. A., & Sass, T. R. (2008). *Classroom peer effects and student achievement* (Working Paper No. 18). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Calcagno, J. C. (2007). *Evaluating the impact of developmental education in community colleges: A quasi-experimental regression-discontinuity design.* Unpublished doctoral dissertation, Columbia University, New York.

Calcagno, J. C., & Long, B. T. (2008). *The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance* (NBER Working Paper No. 14194). Cambridge, MA: National Bureau of Economic Research.

Calcagno, J. C., & Long, B. T. (2009). *Evaluating the impact of remedial education in Florida community colleges: A quasi-experimental regression discontinuity design* (NCPR Brief). New York: Columbia University, Teachers College, National Center for Postsecondary Research.

Campbell, J. W., & Blakey, L. S. (1996). *Assessing the impact of early remediation in the persistence and performance of underprepared community college students.* Paper presented at the 36th annual forum of the Association for Institutional Research, Albuquerque, NM.

Cavalluzzo, L., Lowther, D. L., Mokher, C., & Fan, X. (2012). *Effects of the Kentucky Virtual Schools' hybrid program for algebra I on grade 9 student math achievement* (NCEE No. 2012-4020). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Chen, X., Wu, J., & Tasoff, S. (2010). *Academic preparation for college in the high school senior class of 2003–04* (NCES No. 2010-169). National Center for Education Statistics Working Paper. Washington, DC: U.S. Department of Education.

Cheung, O., Clements, B., & Pechman, E. (1997). *Protecting the privacy of student records* (NCES No. 97-527). National Center for Education Statistics Working Paper. Washington, DC: U.S. Department of Education.

Clotfelter, C. T., Ladd, H. F., Muschkin, C. G., & Vigdor, J. L. (2012). *Success in community college: Do institutions differ?* (Working Paper No. 74). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. I. (2007a). *How and why do teacher credentials matter for student achievement?* (Working Paper No. 2). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. I. (2007b). *Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects* (Working Paper No. 11). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. I. (2008). *School segregation under color-blind jurisprudence: The case of North Carolina* (Working Paper No. 16). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J.I. (2009). *Are teacher absences worth worrying about in the U.S.?* (Working Paper No. 24). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2012). *The aftermath of accelerating algebra: Evidence from a district policy initiative* (Working Paper No. 69). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Code of Federal Regulations (2009). *Protection of human subjects* (4CFR 97). Washington, DC: U.S. Department of Education.

Cook, T. D., & Wong, V. C. (2008). Better quasi-experimental practice. In P. Alasuutar, J. Brannen, & L. Bickman (Eds.), *The handbook of social research* (pp. 134–165). Thousand Oaks, CA: Sage Publications.

Cornwell, C., & Mustard, D. B. (2002). *HOPE, the brain drain, and diversity: The impact of the scholarship on high achievers and African Americans* (Public Policy Research Series Policy Notes, Vol. 3, No. 4). Athens, GA: University of Georgia, Carl Vinson Institute of Government.

Dadgar, M., & Weiss, M. J. (2012). *Labor market returns to sub-baccalaureate credentials: How much does a community college degree or certificate pay?* New York: Columbia University, Teachers College, Community College Research Center.

Data Quality Campaign. (2014). *Data for action 2014: Paving the path to success.* Washington, DC: Author.

Early Childhood Data Collaborative. (2010). *Building and using coordinated state early care and education data systems: A framework for state policymakers.* Washington, DC: Data Quality Campaign.

Easterling, D., Patten, J., & Krile, D. (1995). *The impact of developmental education on student progress: A three-year longitudinal analysis.* Paper presented at the 35th annual forum of the Association for Institutional Research, Boston.

Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: Guilford.

Eyster, L., Nightingale, D. S., Barnow, B. S., O'Brien, C. T., Trutko, J., & Kuehn, D. (2010). *Implementation and early training outcomes of the high growth job training initiative: Final report.* Washington, DC: The Urban Institute Center on Labor, Human Services, and Population.

Feng, L., Figlio, D., & Sass, T. (2010). *School accountability and teacher mobility* (Working Paper 47). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Feng, L., & Sass, T. (2011). *Teacher quality and teacher mobility* (Working Paper 57). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Fong, A. B., Huang, M., & Goel, A. M. (2008). *Examining the links between grade 12 mathematics coursework and mathematics remediation in Nevada public colleges and universities* (Issues & Answers Report, REL 2008–058). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West.

Fortson, K., Verbitsky-Savitz, N., Kopa, E., & Gleason, P. (2012). *Using an experimental evaluation of charter schools to test whether nonexperimental comparison group methods can replicate experimental impact estimates* (NCEE No. 2012-4019). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Fuller, S. C., & Ladd, H. F. (2013). *School-based accountability and the distribution of teacher quality across grades in elementary schools* (Working Paper No. 75). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Furgeson, J., Gill, B., Haimson, J., Killewald, A., McCullough, M., Nichols-Barrer, I., et al. (2012). *Charter-school management organizations: Diverse strategies and diverse student impacts.* Report prepared by Mathematica Policy Research and the University of Washington's Center on Reinventing Public Education. Princeton, NJ: Mathematica Policy Research.

Geiser, K. E., Horwitz, I. M., Gerstein, A., Butlet, B., Mallonee, N., McLaughlin, M., et al. (2012). *Early childhood education and community schools linkages project: Implementation study.* Palo Alto, CA: John W. Gardner Center for Youth and Their Communities.

Goldhaber, D., Gross, B., & Player, D. (2007). *Are public schools really losing their "best"? Assessing the career transitions of teachers and their implications for the quality of the teacher workforce* (Working Paper No. 12). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Goldhaber, D., Grout, C., Pennucci, A., & Bignell, W. (2012). *Teacher pension choice: Surveying the landscape in Washington State* (Working Paper 81). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2006). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41*(6), 872–884.

Graham, J. W. (2012). *Missing data: Analysis and design.* New York: Springer.

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications.* New York: Sage.

Guryan, J. (2001). *Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts* (NBER Working Paper No. 8269). Cambridge, MA: National Bureau of Economic Research.

Harris, D., & Sass, T. (2007a). *Teacher training, teacher quality, and student achievement* (Working Paper No. 3). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Harris, D., & Sass, T. (2007b). *The effects of NBPTS-certified teachers on student achievement* (Working Paper No. 4). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Henderson, S., Petrosino, A., Guckenberg, S., & Hamilton, S. (2008). *A second follow-up year for measuring how benchmark assessments affect student achievement* (REL Technical Brief, REL 2008–002). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.

Hernandez, D. J. (2012). *PreK–3rd: Next steps for state longitudinal data systems* (PreK-3rd Policy to Action Brief No. 8). New York: Foundation for Child Development.

Holzer, H. J., & Nightingale, D. S. (2009). *Strong students, strong workers: Models for student success through workforce development and community college partnerships.* Washington, DC: Center for American Progress.

Holt, J., White, B., & Lichtenberger, E. (2013). *Developing indicators of progress toward a college degree.* Edwardsville, IL: Illinois Education Research Council.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615–635.

Institute of Education Sciences. (2009). *Grants for statewide, longitudinal data systems under the American Recovery and Reinvestment Act of 2009* (CFDA number 84.384). Washington, DC: U.S. Department of Education.

Jacob, B. J., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226–244.

Jacobson, L. (2011a). *Improving community college outcome measures using Florida longitudinal schooling and earnings data* (Policy Brief). Research supported by the Bill & Melinda Gates Foundation. McLean, VA: New Horizons Economic Research and CNA.

Jacobson, L. (2011b). *Strengthening community colleges' influence on economic mobility.* Paper presented at the Federal Reserve Bank of San Francisco Conference, Creating and Employing a Skilled Work Force: Challenges and Opportunities, San Francisco.

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2012). *Teacher preparation programs and teacher quality: Are there real differences across programs?* (Working Paper No. 79). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Ladd, H. F., Muschkin, C. G., & Dodge, K. (2012). *From birth to school: Early childhood initiatives and third grade outcomes in North Carolina.* Durham, NC: Duke University, Sanford School of Public Policy.

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature, 48*(2), 281–355.

Levesque, K. (2013). *Using data to improve college readiness and success: Lessons from a three-state pilot* (Research brief supported by the Bill & Melinda Gates Foundation). Seattle, WA: Bill & Melinda Gates Foundation.

Levesque, K., & Fitzgerald, R. (2013). *Developing high-quality data to improve college readiness and success: Lessons from a three-state pilot* (Research brief supported by the Bill & Melinda Gates Foundation). Seattle, WA: Bill & Melinda Gates Foundation.

Little, R. J., D'Agostino, R., Cohen, M. L., Dikersin, K., & Emerson, S. S. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine, 367,* 1555-1360.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: Guilford.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics, 122*(1), 159–208.

Martorell, P., & McFarlin, I. 2007. *Help or hindrance? The effects of college remediation on academic and labor market outcomes.* Unpublished manuscript.

Maryland State Department of Education. (2012). *Children entering school ready to learn.* Baltimore, MD: Author.

Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2012). *Different skills: Identifying differentially effective teachers of English language learners* (Working Paper No. 68). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: RAND Corporation.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*(2), 698–714.

Mihaly, K., McCaffrey, D., Sass, T.R., & Lockwood, J. R. (2012). *Where you come from or where you go? Distinguishing between school quality and the effectiveness of teacher preparation program graduates* (Research Paper Series No. 12-12). Atlanta, GA: Georgia State University, Andrew Young School of Policy Studies.

Mokher, C. (2011). *Aligning career and technical education with high-wage and high-demand occupations in Tennessee* (Issues & Answers Report, REL 2011–111). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia.

National Center for Education Evaluation. (2014). *What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0.* Retrieved April 6, 2014, from http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19.

National Forum on Education Statistics. (2012). *Forum guide to supporting data access for researchers: A state education agency perspective* (NFES No. 2012-809). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Retrieved April 6, 2014, from http://nces.ed.gov/programs/slds/publications.asp.

Nichols, A., & Özek, U. (2010). *Public school choice and student achievement in the District of Columbia* (Working Paper No. 53). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Ohio Board of Regents. (2006). *Costs and consequences of remedial course enrollment in Ohio public higher education: Six-year outcomes for fall 1998 cohort.* Columbus, OH: Author.

Özek, U. (2012). *One day too late? Mobile students in an era of accountability* (Working Paper No. 82). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Özek, U., Hansen, M., & Gonzalez, T. (2012). *A leg up or a boot out? Student achievement and mobility under school restructuring* (Working Paper No. 78). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Peisner-Feinberg, E. S., & Schaaf, J. M. (2010). *Long-term effects of the North Carolina More at Four pre-kindergarten program: Children's reading and math skills at third grade.* Chapel Hill, NC: University of North Carolina, Frank Porter Graham Child Development Institute.

Pittard, W. B., III, Hulsey, T. C., Laditka, J. N., & Laditka, S. B. (2012). School readiness among children insured by Medicaid, South Carolina. *Preventing Chronic Disease, 9*(110333).

Puma, M. J., Olsen, R. B., Bell, S. H., and Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Radwin, D., & Hensley, E. (2012). *Using cross-segmental data effectively to support alignment.* Background paper prepared for the College Board's Advocacy and Policy Center Affinity Network. New York: The College Board.

Ralston, K., Newman, C., Clauson, A. Guthrie, J., & Buzby, J. (2008). *The National School Lunch Program: Background, trends, and issues* (Economic Research Report No. 61). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2012). *How teacher turnover harms student achievement* (Working Paper 70). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician, 39*(1), 33–38.

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2007). *Feeling the Florida Heat? How low-performing schools respond to voucher and accountability pressure* (Working Paper 13). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Sass, T., & Cartwright, S. (2008). *High school diploma and GED attainment in Florida* (Research Note 1). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Schweinhart, L. J., Xiang, Z., Echols, M. D., Browning, K., & Wakabayashi, T. (2012). *Michigan Great Start Readiness program evaluation 2012: High school graduation and grade retention findings.* Ypsilanti, MI: HighScope Educational Research Foundation.

Seastrom, M. M. (2010). *Statistical methods for protecting personally identifiable information in aggregate reporting* (NCES No. 2011-603). National Center for Education Statistics SLDS Technical Brief 3. Washington, DC: U.S. Department of Education. Retrieved June 6, 2014, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011603.

Seastrom, M. M. (2012). *NCES statistical standards* (NCES No. 2014-097). National Center for Education Statistics Working Paper. Washington, DC: U.S. Department of Education. Retrieved June 6, 2014, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014097.

SLDS (Statewide Longitudinal Data Systems) Grant Program. (2012a). *Forming research partnerships with state and local education agencies* (SLDS Data Use Issue Brief 2). Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved April 6, 2014, from http://nces.ed.gov/programs/slds/publications.asp.

SLDS (Statewide Longitudinal Data Systems) Grant Program. (2012b). *Techniques for analyzing longitudinal administrative data* (SLDS Data Use Issue Brief 4). Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved April 6, 2014, from http://nces.ed.gov/programs/slds/publications.asp.

SLDS (Statewide Longitudinal Data Systems) Grant Program. (2012c). *Turning administrative data into research-ready longitudinal datasets* (SLDS Data Use Issue Brief 3). Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved April 6, 2014, from http://nces.ed.gov/programs/slds/publications.asp.

SLDS (Statewide Longitudinal Data Systems) Grant Program. (2013). *Building a centralized P20W data warehouse* (SLDS Issue Brief). Washington, DC: U.S. Department of Education Sciences, Institute of Education Sciences. Retrieved April 6, 2014, from http://nces.ed.gov/programs/slds/publications.asp.

van Buuren, S. (2012). *Flexible imputation of missing data.* Boca Raton, FL: CRC Press.

Vigdor, J. L. (2008). *Teacher salary bonuses in North Carolina* (Working Paper 15). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Weissman, J., Silke, E., & Bulakowski, C. (1997). Assessing developmental education policies. *Research in Higher Education, 38*(2), 187–200.

White, M. E., & Fong, A. B. (2008). *Trends in California teacher demand: A county and regional perspective* (Issues & Answers Report, REL 2008–057). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West.

Wilde, E. T., & Hollister, R. (2007). How close is close enough? Evaluating propensity score matching using data from a class size reduction experiment. *Journal of Policy Analysis and Management, 26*(3), 455–477.

Witte, J. F., Wolf, P. J., Dean, A., & Carlson, D. (2010). *Milwaukee Independent Charter Schools study: Report on one year of student growth* (School Choice Demonstration Project Milwaukee Evaluation Report 21—Version 1.1). Little Rock, AR: University of Arkansas.

Xu, Z., Hannaway, J., & D'Souza, S. (2009). *Student transience in North Carolina: The effect of school mobility on student outcomes using longitudinal data* (Working Paper 22). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Xu, Z., Hannaway, J., & Taylor, C. (2007). *Making a difference? The effects of Teach for America in high school* (Working Paper 17). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Xu, Z., Özek, U., & Corritore, M. (2012). *Portability of teacher effectiveness across school settings* (Working Paper 77). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Appendix A. Sample High School Feedback Reports

1. NSC Pilot: Tracking Postsecondary Outcomes for High Schools
   Source: http://collegeready.gatesfoundation.org/Learning/NationalStudentClearinghousePilot
2. Kentucky College and Career Readiness Lafayette High School High School Feedback Report
   Source: http://dataportal.cpe.ky.gov/hsfr/2008/publichs/Public_LafayetteHighSchool.pdf
3. Florida High School Feedback Report
   Source: http://data.fldoe.org/readiness/

National Student Clearinghouse Pilot:
Tracking Postsecondary Outcomes for High Schools

How many of our public high school graduates enroll in postsecondary education?

Percentage of 2009 graduates who enrolled in postsecondary education by fall 2009

27%
No confirmed enrollment by fall 2009

73%
Enrolled by fall 2009

9%
91%

11% 2%
14%
73%

10%
90%

Location
- In state
- Out of state

Type
- 4-year public
- 4-year private
- 2-year public
- 2-year private

Status
- Full-time
- Half-time
- Less than half-time
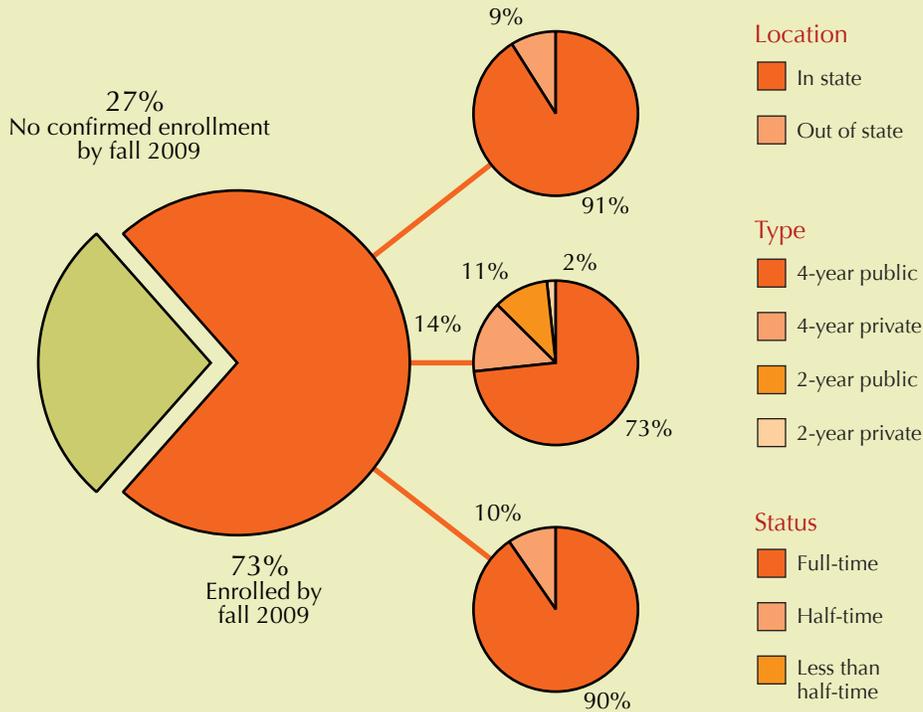
About This Report

This report is the first product of a Pilot project to improve how postsecondary outcomes are reported to and used by the K-12 community. It includes outcomes based on data reported to the National Student Clearinghouse by institutions serving 92% of postsecondary students nationwide. More information and detail, including state-provided postsecondary data and a variety of additional measures, will be available in a Pilot web tool in fall 2010. This report includes information on about 350 of the school's 2009 graduates. See page 4, Understanding the Data, for more detail.

Key Findings and Questions

◄ 73% of graduates enrolled in a postsecondary institution by the fall following graduation.

↓ The enrollment rate for non-low-income students was 15 percentage points higher than the rate for low-income students. What can be done to increase the college-going rate of low-income students in particular?

How does postsecondary enrollment vary by student demographic subgroups?

Percentage of 2009 graduates who enrolled in postsecondary education by fall 2009

Race					Gender		Income	
Asian/Pacific Islander	Black	Hispanic	White	Other	Male	Female	Eligible for free/reduced lunch	Never eligible for free/reduced lunch
91%	75%	44%	76%	70%	72%	76%	64%	79%

All students in state 65%
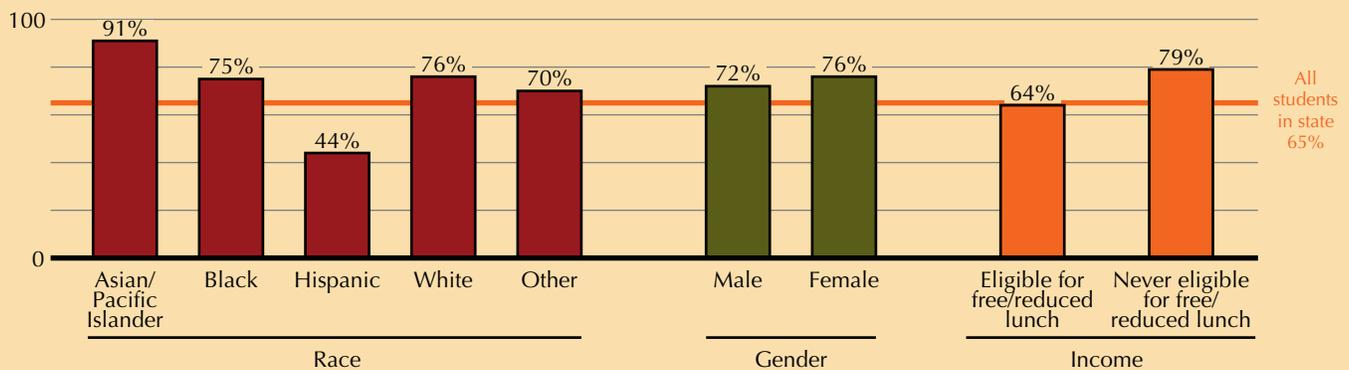
*Among the 2009 graduates included in this report, 5% were Asian/Pacific Islander, 34% were Black, 9% were Hispanic, 50% were White, and 2% were of another ethnicity.

How many of our public high school graduates persist in postsecondary education?

Persistence into second year

Percentage of 2008 graduates who enrolled in the first year after high school

13%
Enrolled but did not return in second year

87%
Enrolled and returned in second year
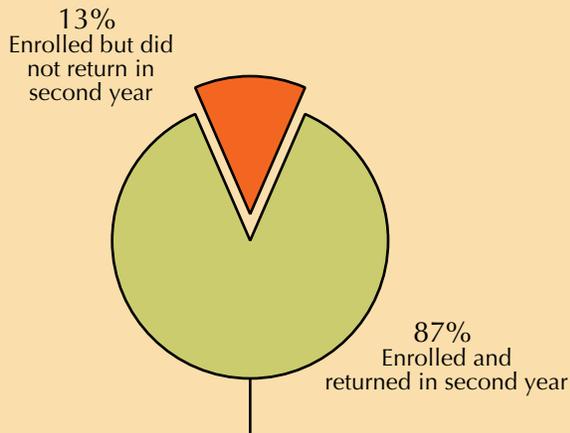
Rates of persistence into second year, by type of institution first attended

91%	89%	73%	68%

Public	Private	Public	Private
4 year		2 year	
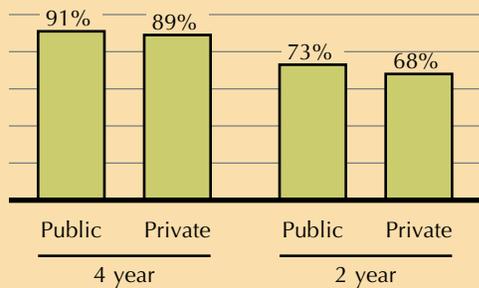
Continuous enrollment over 4 years

Percentage of 2005 graduates who enrolled in the first year after high school in a 4-year institution

31%
Enrolled for fewer than 4 years

69%
Enrolled continuously for 4 years
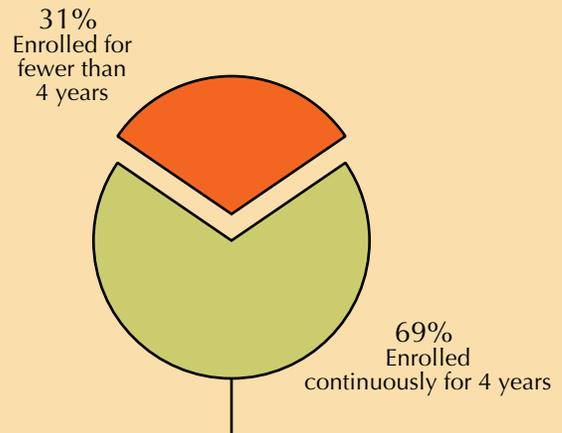
Rates of continuous enrollment over 4 years, by type of 4 year institution first attended

69%	59%

Public	Private
4 year	
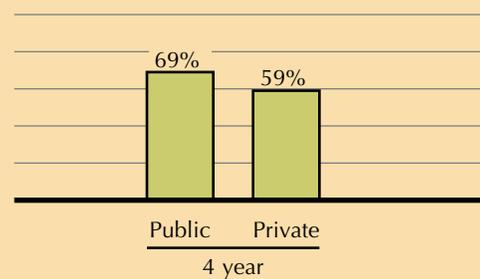
How has postsecondary enrollment changed over time?

Percentage of graduates who enrolled by the fall following graduation

100

69%   74%   67%   73%

0

| 2006 | 2007 | 2008 | 2009 |
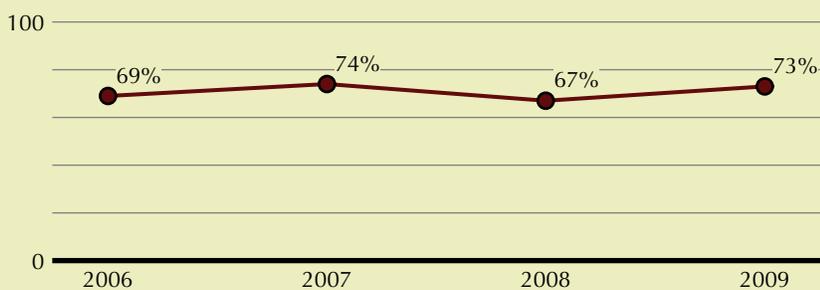
Key Findings and Questions

⬆ Of those who enrolled in college within one year following graduation, 87% persisted into their second year. What factors may contribute to persistence? What can be done to improve persistence rates?

⬅ Enrollment rates varied over time. What can be done to help more graduates both enroll and persist in college?

How does postsecondary enrollment differ for students with similar high school achievement levels but different demographic characteristics?

Percentage of 2009 graduates who enrolled in postsecondary education by fall 2009, by proficiency level on grade 11 GHSGT mathematics assessment[1]

. . . by race/ethnicity[2]

Key Findings and Questions

← As performance on the high school test improved, enrollment rates increased. What can be done to help students attain higher performance levels on the state test?

. . . by eligibility for free or reduced lunch

← The gap in college enrollment rates between low-income and non-low-income students ranged from 9 to 24 percentage points, depending on students' performance level on the high school test. What can be done to reduce these college-going gaps?

[1]Results are based on first test administration.

[2]In these charts, there were too few students to report the "other" race/ethnicity category.

Key Findings and Questions

➡ Among students who were proficient in 8th grade, those who attained advanced proficiency later in high school enrolled in college at higher rates than those who remained proficient. What can be done to help more students make progress on these benchmarks?

Percentage of 2009 graduates who enrolled in postsecondary education by fall 2009, by proficiency level on grade 8 CRCT mathematics assessment and grade 11 GHSGT mathematics assessment
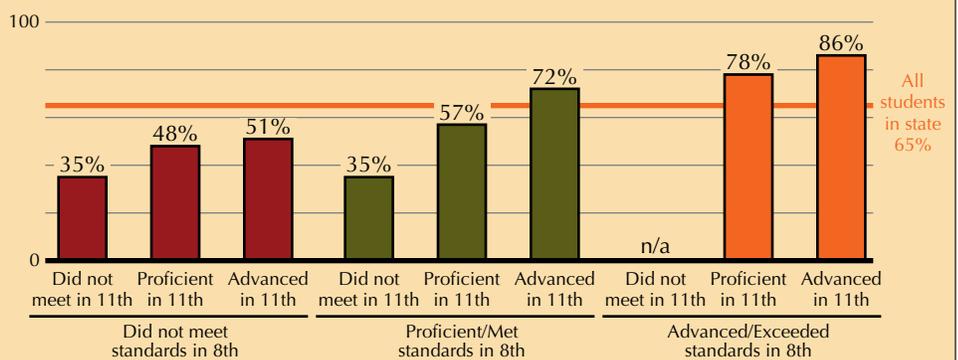
How do public high schools in the state compare?

Distribution of high schools in the state by the percentage of 2009 graduates who enrolled in postsecondary education by fall 2009

Percentage of school's graduates who enrolled by fall 2009

← Only 26 schools saw 80% or more of their graduates enroll immediately in college. What policies and practices might help more schools move into this range?

↓ There is wide variation in the college enrollment rates of high schools serving students with similar achievement rates. What could high schools learn from others like them who are seeing more graduates enroll in college?

ABOUT THIS CHART

Each bar on this chart represents a group of schools whose graduating classes enrolled in college at similar rates. Schools are separated into 20 categories, according to the percentage of their 2009 graduating class who enrolled immediately in college. For example, 67 schools exhibited college enrollment rates of 60 to 65%, the most common outcome among schools in the state. Your school's results are included in the red highlighted column.

BEATING THE ODDS

1. Sample School 1, Sample County
2. Sample School 2, Sample County
3. Sample School 3, Sample County
4. Sample School 4, Sample County
5. Sample School 5, Sample County
6. Sample School 6, Sample County
7. Sample School 7, Sample County
8. Sample School 8, Sample County
9. Sample School 9, Sample County
10. Sample School 10, Sample County

Scatterplot of high schools in the state by percentage of 2009 graduates who enrolled in postsecondary education by fall 2009 and percentage of graduates who met or exceeded standards on grade 8 CRCT mathematics assessment

Percentage of 2009 graduates who met or exceeded standards on the grade 8 CRCT mathematics assessment
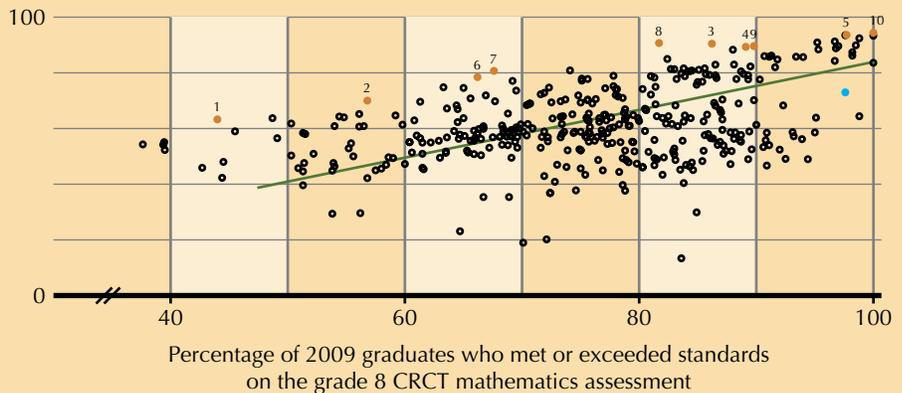
ABOUT THIS CHART

This chart can help identify out-performers among schools with similar student populations based on achievement levels when students entered high school. Each circle on this chart represents a high school. Schools are arrayed from those with the lowest achievement levels on the left to those with the highest achievement levels on the right. Schools that fall above the line exhibit higher-than-expected college enrollment rates, based on the achievement level of their graduates when they entered high school. The top 10 out-performing schools are highlighted in orange and listed at left. Your school is highlighted in blue.

Understanding the Data: This report presents information on graduates of public high schools with at least 30 graduates in the years reported. Students who did not graduate, about 21% of students statewide, are not included in this report. Due to privacy concerns, results are suppressed when there are fewer than 10 students in a group. Suppressed results are indicated by n/a. Additional detail on the data and results in this report including numbers (Ns) for all graphs are available at www.NSCPilotInfo.org, access code ▮▮▮

The Kentucky College and Career Readiness High School Feedback Report is collaboratively produced by Kentucky's Council on Postsecondary Education (CPE), the Kentucky Department of Education (KDE), the Kentucky Higher Education Assistance Authority (KHEAA), and the P-20 Data Collaborative. Its purpose is to provide information about this school's 2008 class of high school seniors including the number who graduated and subsequently enrolled in postsecondary institutions and their level of college and career readiness as compared to their peers from the district and the state as a whole. All college-going information in this report is for 2008 high school graduates who entered postsecondary education in the summer or fall of 2008 and not later.

High School Performance

These numbers provide an overview of this school's graduating class of 2008. District numbers may include alternative high schools. Refer to the Technical Notes for an explanation of dashes.

	School	District	Kentucky
1. Number of high school graduates	394	1,904	43,362
2. Average high school GPA	2.84	2.87	2.94
3. Average ACT scores for this class			
English	22.3	21.8	20.2
Mathematics	21.7	22.0	20.0
Reading	23.6	23.1	21.3
Science	22.4	22.0	20.6
Composite	22.6	22.3	20.7
4. Number of Advanced Placement (AP) tests taken by members of this class	287	1,013	10,557
5. Percentage of Advanced Placement (AP) tests with scores of 3 or higher (the minimum necessary to receive college credit)	60.6%	64.2%	48.8%
6. Average Kentucky Educational Excellence Scholarship (KEES) award earned by members of this class	$1,168	$1,124	$1,097
7. High school graduation rate	84.0%	80.6%	84.5%
8. In-state college-going rate	57.1%	60.6%	56.8%
9. Out-of-state college-going rate	9.6%	11.1%	6.2%
10. Overall college-going rate	66.8%	71.7%	63.0%

Important Statistics

High school graduation rate	84.0%
College-going rate	66.8%
Percentage with developmental needs in one or more subjects	32.6%
Percentage with developmental needs in English	20.7%
Percentage with developmental needs in mathematics	22.2%

Average ACT Scores

	English	Mathematics	Reading	Science	Composite
School	22.3	21.7	23.6	22.4	22.6
District	21.8	22.0	23.1	22.0	22.3
Kentucky	20.2	20.0	21.3	20.6	20.7
National	20.6	21.0	21.4	20.8	21.1
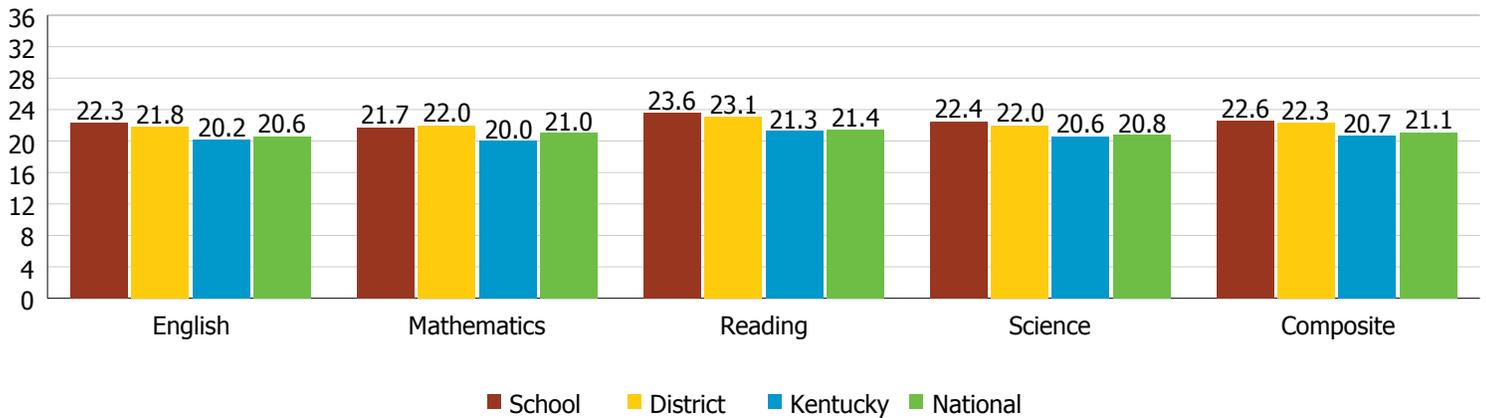
Note: The 2008 Kentucky ACT benchmarks for college and career readiness were scores of 18 in English, Mathematics, and Reading.

For more information, including technical notes, please visit http://cpe.ky.gov/info/hsfr/ or e-mail cpe.hsfr@ky.gov.

A-6

In-State Postsecondary Enrollment Characteristics

These numbers show the types of postsecondary institutions and programs selected by the graduating class of 2008 and make the link between high school performance and postsecondary institution type. The information in item 1.c is only available for public colleges and universities in Kentucky and Association of Independent Kentucky Colleges and Universities (AIKCU) institutions. District numbers may include alternative high schools. Refer to the Technical Notes for an explanation of dashes.

	School	District	Kentucky
1. College enrollment of 2008 graduates entering college in-state	225	1,145	23,286
a. Type of college or university attended			
Four-year public university	132	677	12,713
	58.7%	59.1%	54.6%
Two-year public community or technical college (KCTCS)	63	309	6,501
	28.0%	27.0%	27.9%
Association of Independent Kentucky Colleges and Universities (AIKCU) member institution	26	122	3,372
	11.6%	10.7%	14.5%
Other in-state college or university	4	37	700
	1.8%	3.2%	3.0%
b. Started college full-time	202	1,037	21,115
	89.8%	90.6%	90.7%
c. Degrees sought			
Bachelor's degree	129	674	13,603
	58.4%	60.8%	60.2%
Associate degree	29	176	4,586
	13.1%	15.9%	20.3%
Certificate or diploma	6	20	536
	2.7%	1.8%	2.4%
Undeclared or nondegree	57	238	3,860
	25.8%	21.5%	17.1%
2. Average high school GPA of 2008 graduates entering college in-state			
Overall mean for students entering any Kentucky postsecondary institution	2.92	2.94	3.19
Four-year public university	2.98	3.13	3.41
Two-year public community or technical college (KCTCS)	2.54	2.53	2.80
Association of Independent Kentucky Colleges and Universities (AIKCU) member institution	3.45	3.14	3.38
Other in-state college or university	-	2.05	2.71

Average High School GPA by Postsecondary Institution Type
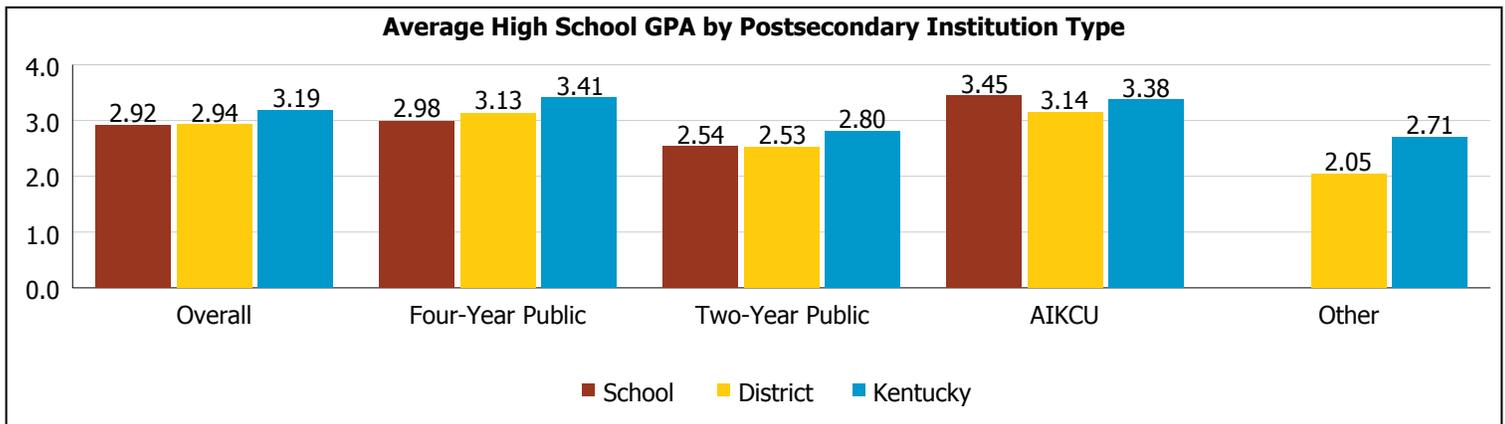
College Readiness Details

These numbers explore the link between college readiness and type of postsecondary institution attended and show college readiness levels. These data can help students, parents, and educators understand the typical ACT scores of students accepted for admission at different types of postsecondary institutions. This information is only available for students who entered public colleges and universities in Kentucky or Association of Independent Kentucky Colleges and Universities (AIKCU) institutions that submitted entrance exam data. District numbers may include alternative high schools. Refer to the Technical Notes for an explanation of dashes.

	School	District	Kentucky
1. Average ACT scores of 2008 graduates entering college in-state			
English	22.3	21.9	21.6
Mathematics	21.6	21.9	21.0
Reading	23.8	23.3	22.8
Science	22.3	22.2	21.6
Composite	22.5	22.3	21.7
2. Average ACT scores of 2008 graduates entering college in-state, by institution type			
a. Four-year public university			
English	23.3	22.9	22.7
Mathematics	22.5	22.9	22.0
Reading	24.5	24.3	23.9
Science	23.1	23.2	21.6
Composite	23.4	23.2	22.7
b. Two-year public community or technical college (KCTCS)			
English	19.5	18.7	18.4
Mathematics	18.8	18.8	18.4
Reading	22.1	20.4	20.0
Science	20.3	19.7	19.5
Composite	20.2	19.5	19.0
c. Association of Independent Kentucky Colleges and Universities (AIKCU) member institution			
English	20.0	20.5	21.6
Mathematics	20.4	20.8	20.7
Reading	21.1	22.5	22.8
Science	19.9	20.9	21.7
Composite	20.4	21.2	21.6
3. Percentage not ready for college, 2008 graduates entering college in-state (received an ACT subject score of less than 18 or an equivalent score on an alternative test)			
English	20.7%	21.0%	21.8%
Mathematics	22.2%	22.6%	28.2%
Reading	12.9%	16.4%	16.6%
Not college-ready in one or more subjects	32.6%	33.2%	38.2%

Percentage Not Ready for College by Subject

For more information, including technical notes, please visit http://cpe.ky.gov/info/hsfr/ or e-mail cpe.hsfr@ky.gov.
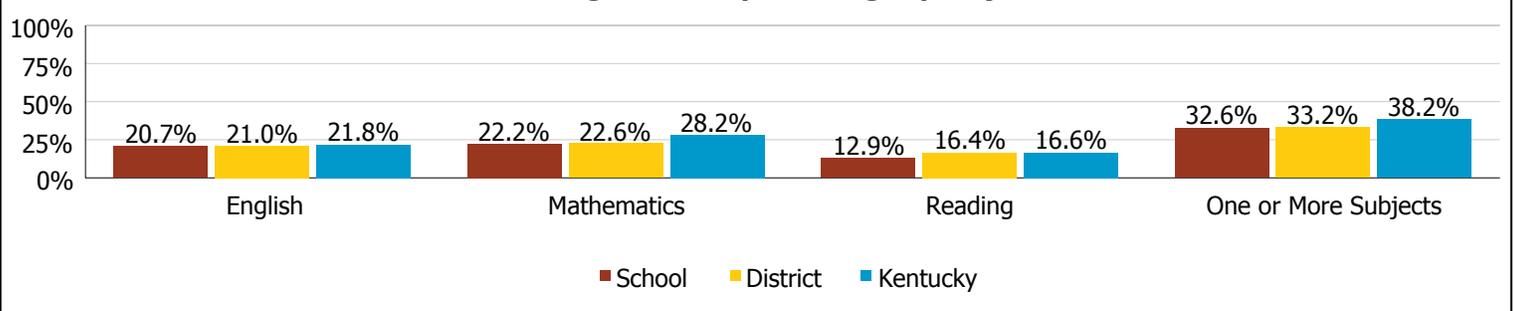
ACT Score Distributions of College-Going Students

Each graph depicts the distribution of ACT scores among college entrants from the class of 2008 and compares high school and state distributions. Kentucky ACT college readiness benchmarks are minimum scores at which students place out of non-credit-bearing developmental or remedial education courses in a given subject. Red bars indicate the 2008 benchmarks and blue bars indicate the revised benchmarks for 2010 and later. No benchmark score has been established for science. This information is only available for students who entered public colleges and universities in Kentucky or Association of Independent Kentucky Colleges and Universities (AIKCU) institutions that submitted entrance exam data.

Legend: ◼ School Score — State Score ▬ Benchmark Score 2008 ▬ Benchmark Score 2010

English
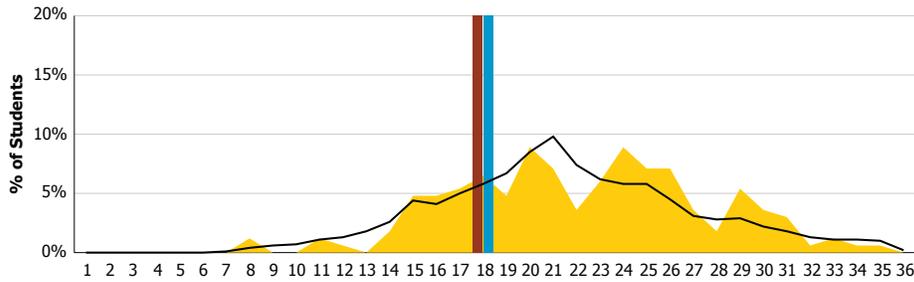
Note: The 2008 Kentucky ACT benchmark for college readiness in English was a score of 18 and is unchanged under the new 2010 benchmarks. Eighteen is also the score ACT, Inc., recommends for college readiness.

Mathematics
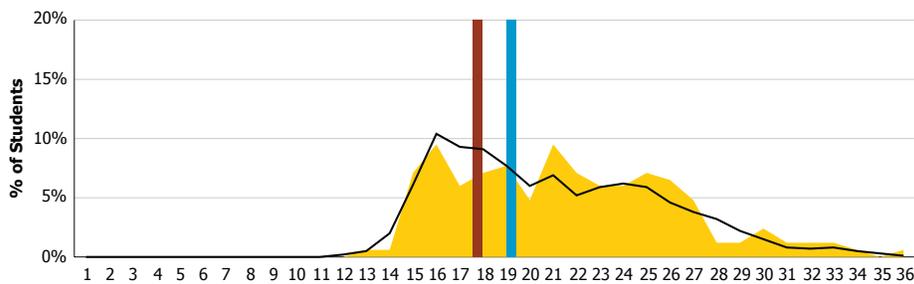
Note: The 2008 Kentucky ACT benchmark for college readiness in mathematics was 18, and changes to 19 for the entering class of 2010. This is the minimum score needed for placement in credit-bearing mathematics and does not guarantee entrance into all majors, especially majors in the STEM fields. The state benchmark for placement into college algebra is 22 and 27 for calculus. The score recommended by ACT, Inc., as a benchmark for college readiness in mathematics is 22.

Reading
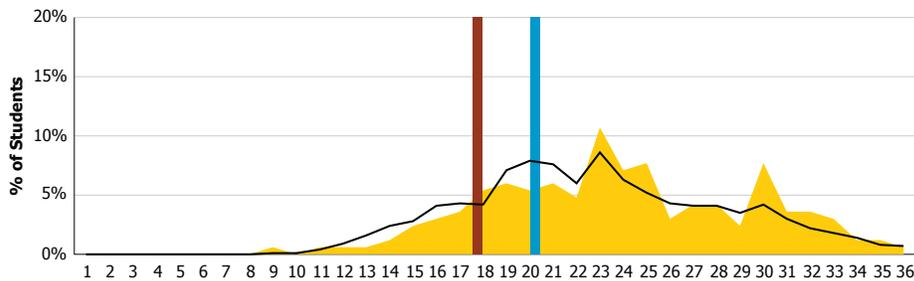
Note: The 2008 Kentucky ACT benchmark for college readiness in reading was 18, and will rise to 20 for the entering class of 2010. The score recommended by ACT, Inc., as a benchmark for college readiness in reading is 21.

Science
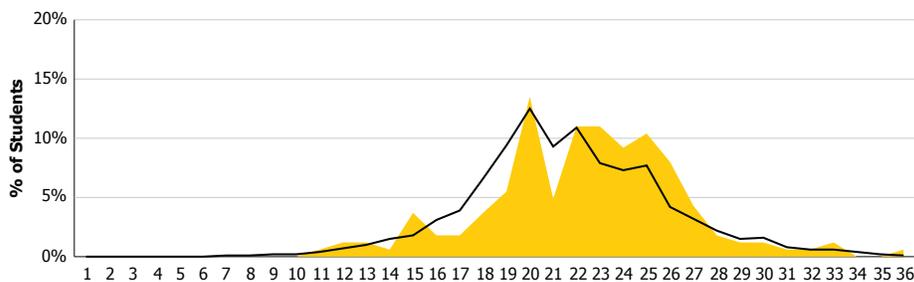
Note: A Kentucky ACT benchmark for college readiness in science has not been established. The score recommended by ACT, Inc., as a benchmark for college readiness in science is 24.

Composite
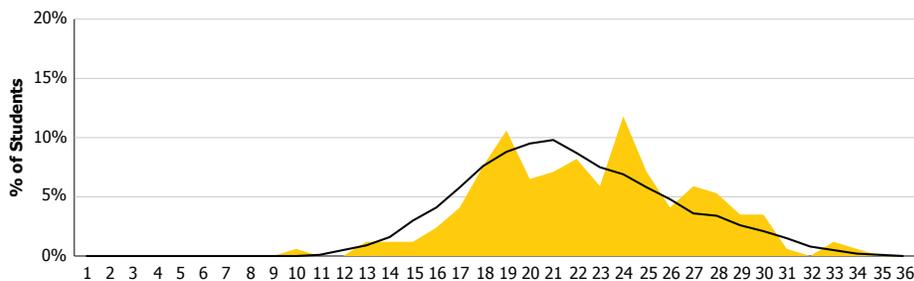
Note: No ACT composite score benchmarks for college readiness have been established by the state or recommended by ACT, Inc.

For more information, including technical notes, please visit http://cpe.ky.gov/info/hsfr/ or e-mail cpe.hsfr@ky.gov.

Colleges and Universities Attended and College Majors

These numbers show the Kentucky colleges and universities in which the class of 2008 enrolled and the majors that they declared. College majors are only available for students who entered public colleges and universities in Kentucky or Association of Independent Kentucky Colleges and Universities (AIKCU) institutions. Majors were not reported for some students.

Colleges and Universities Attended	Students
Four-year public university	
University of Kentucky	69
Eastern Kentucky University	34
Northern Kentucky University	11
Kentucky State University	6
University of Louisville	6
Western Kentucky University	4
Morehead State University	1
Murray State University	1
Two-year public community or technical college (KCTCS)	
Bluegrass CTC	62
Madisonville CC	1
Association of Independent Kentucky Colleges and Universities member institution	
Transylvania University	13
Centre College	4
Lindsey Wilson College	3
Campbellsville University	2
Berea College	1
Georgetown College	1
Union College	1
University of the Cumberlands	1
Other in-state college or university	
Sullivan University and College of Technology and Design	2
Out-of-state	38

College Majors	Students
Major Undeclared	94
Education	15
Liberal Arts and Humanities	14
Health Professions	11
Visual and Performing Arts	11
Business, Management and Marketing	10
Biological and Biomedical Sciences	9
Psychology	8
Engineering	7
Security and Protective Services	6
Agriculture	5
Communications and Journalism	5
Computer and Information Sciences	5
Mechanic and Repair Technologies	3
Social Sciences	3
Family and Consumer Sciences	2
History	2
Mathematics and Statistics	2
Personal and Culinary Services	2
Physical Sciences	2
Engineering Technologies	1
English Language and Literature	1
Multi/Interdisciplinary Studies	1
Natural Resources and Conservation	1
Public Administration	1

DISTRICT: CITRUS

SCHOOL: CITRUS HIGH SCHOOL	SCHOOL	DISTRICT	STATE
1 Number of 2009 high school graduates with standard diploma or GED:	357	1,125	149,459

PRE-GRADUATION INDICATORS	SCHOOL	DISTRICT	STATE
2 Percent of 2009 graduates who scored at level 3 or better on the 10th grade FCAT in:			
Math.....................................	77.1%	75.3%	78.4%
Reading................................	47.6%	44.7%	46.8%
Both Reading and Math.........	43.3%	40.9%	44.3%
3 Percent of 2009 graduates who completed a college prep curriculum:	50.1%	43.3%	59.8%
4 Percent of 2009 graduates who were eligible for the maximum Bright Futures award:			
FL Academic Scholars..........	6.16%	4.35%	7.21%
FL Medallion Scholars............	32.7%	26.1%	25.4%
FL Gold Seal Vocational........	2.24%	1.95%	1.08%
5 Percent of 2009 graduates who completed at least one AP, IB, AICE or Dual Enrollment course:	30.5%	27.4%	42.4%
6 Percent of 2009 graduates enrolled in Algebra I or equivalent in a FL public school prior to 9th grade:	17.9%	18.4%	29.6%
7 Percent of 2009 graduates who completed at least one level 3 high school math course:	27.1%	22.1%	47.2%
8 Percent of 2009 graduates who completed at least one dual enrollment math course:	8.40%	5.06%	6.06%
9 Percent of 2009 graduates who completed at least one level 3 high school science course:	33.6%	28.3%	53.9%
10 Percent of 2009 graduates who completed at least one dual enrollment science course:	3.36%	3.2%	2.83%
11 Percent of students who took PSAT or PLAN two years prior to graduation year:			
PSAT....................................	81.7%	80.5%	69.4%
PLAN....................................	NA	NA	11.8%
12 Percent of 2009 graduates who took the SAT, ACT, CPT or PERT:			
SAT......................................	21.0%	23.1%	48.0%
ACT......................................	61.6%	47.0%	51.7%
CPT......................................	41.7%	26.7%	32.6%
PERT....................................	NA	NA	NA
13 Percent of 2009 graduates with standard high school diploma who took the SAT / ACT / CPT / PERT, entered a public college or university in Florida in the year following graduation, and scored at or above college-level cut scores:	Link to PCPT Report Page		
Math.....................................	75.0%	73.8%	68.4%
Reading................................	82.3%	82.3%	78.5%
Writing.................................	79.9%	81.8%	80.4%
All Three Subjects................	66.0%	66.7%	62.1%

POST-GRADUATION INDICATORS		SCHOOL	DISTRICT	STATE
14	Percent of 2009 graduates enrolled in a Florida public postsecondary institution in Fall 2009:	53.5%	41.5%	53.0%
15	Percent of 2009 graduates found enrolled in Independent Colleges and Universities of Florida (ICUF) in Fall 2009:	1.12%	2.04%	3.02%
16	Percent of 2009 graduates found enrolled in an out-of-state public or private institution in Fall 2009:	NA	NA	NA
17	Percent of 2009 graduates:			
	at a community college in Florida during Fall 2009................	38.3%	29.1%	34.6%
	at a state university in Florida during Fall 2009.....................	12.6%	9.95%	17.7%
	at a technical education center in Florida during Fall 2009.....	3.36%	3.28%	1.22%
18	Percent of 2009 graduates enrolled in college credit courses in Fall 2009 at a FL public postsecondary institution earning a GPA above 2.0:	74.2%	73.9%	76.0%
19	Percent of graduates enrolled in college credit courses at Independent Colleges and Universities of Florida (ICUF) earning a GPA above 2.0:	100%	91.6%	81.1%
20	Of the graduates enrolled in a Math course in Florida in Fall, the percent who successfully completed the course:			
	Remedial Math (non-college credit)	62.0%	66.6%	60.2%
	Intermediate Algebra (for elective credit only)	80.6%	73.8%	61.9%
	Entry-level Math (for Math credit)	41.1%	69.2%	67.7%
	Advanced Math	73.8%	71.8%	59.5%
21	Of the graduates enrolled in an English course in Florida in Fall, the percent who successfully completed the course:			
	Remedial Reading or Writing.................................	70.5%	75.3%	81.0%
	Freshman Comp I or II...	79.3%	80.9%	79.9%
	Other College-level English...................................	79.3%	80.9%	80.5%

Appendix B. Characteristics of statewide student data systems, by state: 2009–2010

Characteristics of Statewide Student Data Systems, by state: 2009-10

| State | PK-12-Postsecondary | | | | | PK-12 | | | | | Postsecondary | | Total |
	1) Uses Unique Student Identifier	2) Contains student-level enrollment, demographic, and program participation information	3) Contains exit, drop out, transfer, and completion information of P-16 programs	4) Communicates with higher education	5) Assesses Data Quality, Validity, Reliability	1) Contains Yearly Test Records	2) Contains information on students not tested	3) Uses teacher Identification systems and can match them with students	4) Contains student-level transcript information	5) Contains college readiness test scores	1) Contains information on students' transition from high school to postsecondary institutions	2) Contains other information to determine alignment and preparedness for success in postsecondary education	
United States	43	45	36	33	48	49	49	30	28	40	28	29	
Alabama	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	9
Alaska	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	6
Arizona	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	7
Arkansas	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
California	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	9
Colorado	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	11
Connecticut	No	No	No	No	Yes	Yes	Yes	No	No	Yes	No	Yes	5
Delaware	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
District of Columbia	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	No	No	7
Florida	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
Georgia	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	8
Hawaii	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
Idaho	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	9
Illinois	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	10
Indiana	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	9
Iowa	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	9
Kansas	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	10
Kentucky	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	10

State	PK-12-Postsecondary					PK-12					Postsecondary		Total
	1) Uses Unique Student Identifier	2) Contains student-level enrollment, demographic, and program participation information	3) Contains exit, drop out, transfer, and completion information of P-16 programs	4) Communicates with higher education	5) Assesses Data Quality, Validity, Reliability	1) Contains Yearly Test Records	2) Contains information on students not tested	3) Uses teacher Identification systems and can match them with students	4) Contains student-level transcript information	5) Contains college readiness test scores	1) Contains information on students' transition from high school to postsecondary institutions	2) Contains other information to determine alignment and preparedness for success in postsecondary education	
Louisiana	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
Maine	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	No	No	6
Maryland	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	No	No	7
Massachusetts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
Michigan	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Yes	No	No	6
Minnesota	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes	No	No	6
Mississippi	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes	8
Missouri	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	11
Montana	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Yes	8
Nebraska	No	No	No	No	Yes	Yes	Yes	No	No	Yes	No	No	4
Nevada	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	10
New Hampshire	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	4
New jersey	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes	No	No	6
New Mexico	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
New York	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
North Carolina	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	8
North Dakota	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	No	8
Ohio	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	8
Oklahoma	Yes	Yes	No	No	Yes	No	No	Yes	No	No	Yes	Yes	6
Oregon	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	8
Pennsylvania	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12

| State | PK-12-Postsecondary | | | | | PK-12 | | | | | Postsecondary | | Total |
	1) Uses Unique Student Identifier	2) Contains student-level enrollment, demographic, and program participation information	3) Contains exit, drop out, transfer, and completion information of P-16 programs	4) Communicates with higher education	5) Assesses Data Quality, Validity, Reliability	1) Contains Yearly Test Records	2) Contains information on students not tested	3) Uses teacher Identification systems and can match them with students	4) Contains student-level transcript information	5) Contains college readiness test scores	1) Contains information on students' transition from high school to postsecondary institutions	2) Contains other information to determine alignment and preparedness for success in postsecondary education	
Rhode Island	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	11
South Carolina	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
South Dakota	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	10
Tennessee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
Texas	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	No	8
Utah	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	9
Vermont	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	9
Virginia	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	10
Washington	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
West Virginia	No	Yes	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	7
Wisconsin	No	No	No	No	Yes	Yes	Yes	No	No	Yes	No	Yes	5
Wyoming	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12

Appendix C. The Family Education Rights and Privacy Act— guidance for reasonable methods and written agreements

Source: https://www2.ed.gov/policy/gen/guid/fpco/pdf/reasonablemtd_agreement.pdf.

What is the Family Educational Rights and Privacy Act?

The Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. §1232g, is a Federal privacy law administered by the Family Policy Compliance Office (FPCO or Office) in the U.S. Department of Education (Department or we). FERPA and its implementing regulations in 34 CFR part 99 protect the privacy of students' education records and afford parents and eligible students (i.e., students who are 18 years of age or older or attend an institution of postsecondary education) certain rights to inspect and review education records, to seek to amend these records, and to consent to the disclosure of personally identifiable information from education records (PII from education records).

The general rule under FERPA is that PII from education records cannot be disclosed without written consent. However, FERPA includes several exceptions that permit the disclosure of PII from education records without consent. Two of these exceptions are discussed in this document—the studies exception and the audit or evaluation exception. The two exceptions contain specific, and slightly different, requirements, described more fully in the implementing regulations (34 CFR Part 99).

What is the purpose of this document?

The audience for this document includes schools, school districts (also referred to as local educational agencies [LEAs]), postsecondary institutions, and State educational authorities (such as State educational agencies [SEAs]) that may disclose PII from education records. Our intent is to provide these entities with information about requirements and best practices for data disclosures under the studies exception and the audit or evaluation exception.

What is the Studies Exception? (see 20 U.S.C. §1232g(b)(1)(F) and §99.31(a)(6))

The studies exception allows for the disclosure of PII from education records without consent to organizations conducting studies for, or on behalf of, schools, school districts, or postsecondary institutions. Studies can be for the purpose of developing, validating, or administering predictive tests; administering student aid programs; or improving instruction.

> Example: An SEA may disclose PII from education records without consent to an organization for the purpose of conducting a study that compares program outcomes across school districts to further assess what programs provide the best instruction and then duplicate those results in other districts.

What is the Audit or Evaluation Exception? (see 20 U.S.C. 1232g(b)(1)(C), (b)(3), and (b)(5) and §§99.31(a)(3) and 99.35)

The audit or evaluation exception allows for the disclosure of PII from education records without consent to authorized representatives of the Comptroller General of the U.S., the Attorney General, the Secretary of Education, and State or local educational authorities (FERPA-permitted entities). Under this exception, PII from education records must be used to audit or evaluate a Federal- or State-supported education program, or to enforce or comply with Federal legal requirements that relate to those education programs (audit, evaluation, or enforcement or compliance activity). The entity disclosing the PII from education records is specifically required to use reasonable methods to ensure to the greatest extent practicable that its designated authorized representative complies with FERPA and its regulations.

> Example: An LEA could designate a university as an authorized representative in order to disclose, without consent, PII from education records on its former students to the university. The university then may disclose, without consent, transcript data on these former students to the LEA to permit the LEA to evaluate how effectively the LEA prepared its students for success in postsecondary education.

How do you define education program?

"Education program" is an important term under the audit or evaluation exception because PII from education records can only be disclosed to audit or evaluate a Federal- or State-supported "education program," or to enforce or to comply with Federal legal requirements related to an education program. As specified in the FERPA regulations, §99.3, an education program must be principally engaged in the provision of education, including, but not limited to, early childhood education, elementary and secondary education, postsecondary education, special education, job training, career and technical education, and adult education, and any program that is administered by an educational agency or institution. For a definition of "early childhood program" please refer to §99.3 of the FERPA regulations.

Do we need to have a written agreement to disclose PII from education records without consent?

Yes. Both the studies exception and the audit or evaluation exception specifically require that the parties execute a written agreement when disclosing PII from education records without consent. The mandatory elements of that agreement vary slightly between the two exceptions.

Are there mandatory provisions for written agreements under the studies exception?

Yes. Written agreements under the studies exception must in accordance with the requirements in §99.31(a)(6)(iii)(C):

1. Specify the purpose, scope, and duration of the study and the information to be disclosed. Your agreement must specify the purpose of the study, describe its scope and its duration, and identify the information being disclosed.

2. Require the organization to use PII from education records only to meet the purpose or purposes of the study as stated in the written agreement. Your agreement must specify that the PII from education records must only be used for the study identified in the agreement.

3. Require the organization to conduct the study in a manner that does not permit the personal identification of parents and students by anyone other than representatives of the organization with legitimate interests. Your agreement must require the organization to conduct the study so as not to identify students or their parents. This typically means that the organization should allow internal access to PII from education records only to individuals with a need to know, and that the organization should take steps to maintain the confidentiality of the PII from education records at all stages of the study, including within the final report, by using appropriate disclosure avoidance techniques.

4. Require the organization to destroy all PII from education records when the information is no longer needed for the purposes for which the study was conducted, and specify the time period in which the information must be destroyed. Your agreement must require the organization to destroy the PII from education records when it is no longer needed for the identified study. You should determine the specific time period for destruction based on the facts and circumstances surrounding the disclosure and study. The parties to the written agreement may agree to amend the agreement to extend the time period if needed, but the agreement must include a time limit.

Are there mandatory provisions for written agreements under the audit or evaluation exception?

Yes. The mandatory provisions for written agreements under the audit or evaluation exception are similar to, but slightly different from, the provisions required for written agreements under the studies exception. Section 99.35(a)(3) specifically requires that the following provisions be included in written agreements under the audit or evaluation exception:

1. Designate the individual or entity as an authorized representative. Your agreement must formally designate the individual or entity as an authorized representative.

2. Specify the PII from education records to be disclosed. Your agreement must identify the information being disclosed.

3. Specify that the purpose for which the PII from education records is being disclosed to the authorized representative is to carry out an audit or evaluation of Federal- or State-supported education programs, or to enforce or to comply with Federal legal requirements that relate to those programs. Your agreement must state specifically that the disclosure of the PII from education records is in furtherance of an audit, evaluation, or enforcement or compliance activity.

4. Describe the activity with sufficient specificity to make clear that it falls within the audit or evaluation exception. This must include a description of how the PII from education records will be used. Don't be vague—the agreement must describe the methodology and why disclosure of PII from education records is necessary to accomplish the audit, evaluation, or enforcement or compliance activity.

5. Require the authorized representative to destroy the PII from education records when the information is no longer needed for the purpose specified. Your agreement should be clear about how the PII from education records will be destroyed.

6. Specify the time period in which the PII must be destroyed. Your agreement must provide a time period for destruction. You should determine the specific time period for destruction based on the facts and circumstances surrounding the disclosure and activity. The parties to the written agreement may agree to amend the agreement to extend the time period if needed, but the agreement must include a time limit.

7. Establish policies and procedures, consistent with FERPA and other Federal and State confidentiality and privacy provisions, to protect PII from education records from further disclosure (except back to the disclosing entity) and unauthorized use, including limiting use of PII from education records to only authorized representatives with legitimate interests in an audit, evaluation, or enforcement or compliance activity. The agreement must establish the policies and procedures, consistent with FEPRA and other Federal and State laws, to protect PII from education records from further disclosure or unauthorized use.

Can an entity receiving PII from education records disclose it in a way that allows individual students to be identified?

No. Absent consent from the parent or eligible student, FERPA provides that the PII from education records cannot be published in a way that would allow individual students and their parents to be identified. The organization conducting the study, audit, or evaluation can use PII from education records to conduct the study, audit, or evaluation, but results must be published in a way that protects the privacy and confidentiality of the individuals involved. For example, when publishing tables, cell suppression and other methods of disclosure avoidance can be used so that students cannot be identified through small numbers displayed in table cells.

Under the audit or evaluation exception, what is your responsibility to use "reasonable methods" to ensure that your authorized representative is FERPA-compliant to the greatest extent practicable? (§99.35(a)(2))

When you disclose PII from education records under the audit or evaluation exception, you are required to use "reasonable methods" to ensure to the greatest extent practicable that your authorized representative is FERPA-compliant. This specifically means ensuring that your authorized representative does the following:

1. Uses PII from education records only to carry out an audit or evaluation of Federal- or State-supported education programs, or for the enforcement of or compliance with, Federal legal requirements related to these programs. You should make sure that the proposed audit or evaluation is legitimate, and require in your written agreement that your authorized representative use the PII from education records only for that audit, evaluation, or enforcement or compliance activity. You should not disclose all of your PII from education records; rather, you should determine which specific elements your authorized representative needs and disclose only those.

2. Protects the PII from education records from further disclosures or other uses, except as authorized by you in accordance with FERPA. Your agreement must specify that your authorized representative may not further disclose the PII from education records, unless authorized. Approval to use the PII from education records for one audit or evaluation does not confer approval to use it for another.

3. Destroys the PII from education records when no longer needed for the audit, evaluation, or enforcement or compliance activity. Your agreement must specify that your authorized representative is required to destroy the PII from education records when it is no longer needed and specify the time period in which the PII must be destroyed.

Are there best practices that support reasonable methods?

Yes. While it is vital for organizations to comply with FERPA and its regulations, FERPA represents the floor for protecting privacy, not the ceiling. Accordingly, the Department is also specifying best practices, in which we describe actions we recommend you take to ensure that your authorized representative is protecting privacy to the greatest extent possible. Best practices are broader than FERPA compliance and describe recommended actions you should take to ensure that your authorized representative is FERPA-compliant to the greatest extent practicable.

These best practices may apply to data sharing under both the audit and evaluation exception and the studies exception. Please keep in mind that not all of the following best practices are appropriate in every instance, and this list does not include every possible protection. Before disclosing PII from education records under one of these exceptions, you should examine the following list and tailor your practices as necessary and appropriate.

- *Convey the limitations on the data.* You should take steps to ensure your authorized representative knows the limitations on the use of the data (i.e., that the data is only to carry out the audit or evaluation of Federal- or State-supported education programs, or to enforce or to comply with Federal legal requirements that relate to those programs).
- *Obtain assurances against redisclosure.* You should obtain assurances from your authorized representative that the data will not be redisclosed without permission, including such assurances that your authorized representative will provide you (the disclosing entity) the right to review any data prior to publication and to verify proper disclosure avoidance techniques have been used.
- *Be clear about destruction.* You should set clear expectations so your authorized representative knows what process needs to be followed for the proper destruction of PII from education records.
- *Maintain a right to audit.* You should maintain the right to conduct audits or other monitoring activities of your authorized representative's policies, procedures, and systems.
- *Verify the existence of disciplinary policies to protect data.* You may want to verify that your authorized representative has appropriate disciplinary policies for employees that violate FERPA. This can include termination in appropriate instances.
- *Verify the existence of a sound data security plan.* You may wish to verify before disclosing PII from education records that your authorized representative has a sound data security program, one that protects both data at rest and data in transmission. You have a responsibility to determine if your authorized representative's data security plan is adequate to prevent FERPA violations. The steps that you may need to take in order to verify a sound data security program are likely to vary with each situation. In some cases, it may suffice to add language to the written agreement that states what data security provisions are required. In other cases, it may be more prudent for you to take a hands-on approach and complete a physical inspection. Additionally, your written agreements could specify

required data security elements, including requirements related to encryption, where the data can be hosted, transmission methodologies, and provisions to prevent unauthorized access.

- *Verify the existence of a data stewardship program.* You may want to examine your authorized representative's data stewardship program. Data stewardship should involve internal control procedures that protect PII from education records and include all aspects of data collection—from planning to maintenance to use and dissemination. The Department believes that a good data stewardship plan would have support and participation from across the organization, including the head of the organization, management, legal counsel, and data administrators, providers, and users. The plan should detail the organization's policies and procedures to protect privacy and data security, including the ongoing management of data collection, processing, storage, maintenance, use, and destruction. The plan could also include designating an individual to oversee the privacy and security of the PII from the education records it maintains. For more information, we have posted for comment a technical brief: "Data Stewardship: Managing Personally Identifiable Information in Electronic Student Education Records" that can be found at http://www2.ed.gov/policy/gen/guid/ptac/technical-briefs.html

- *Disclose only PII from education records that is needed.* When you consider disclosing PII from education records to an authorized representative for an audit, evaluation, or enforcement or compliance activity, you may want to explore which specific data elements are necessary for that activity and provide only those elements. You should take care to ensure that you are not disclosing more PII from education records than needed for the stated activity and purpose. You should also explore whether PII from education records is actually required, or whether de-identified data would suffice.

- *Know to whom you are disclosing data.* You may want to require your authorized representative to conduct background investigations of employees who will have access to PII from education records, or you may want to conduct these investigations yourself. Additionally, you may want to require your authorized representative to disclose past FERPA or data management violations. If you discover past violations, you would want to explore the circumstances behind the violation, and discover all information that would allow you to make an informed judgment on whether the individual or entity is likely to be a responsible data steward. This may include discovering whether the violation was covered up, including if it was voluntarily reported to affected students or FPCO, and whether appropriate breach response procedures were followed.

- *Verify training.* You may want to verify that your authorized representative has a training program to teach its employees about FERPA and how to protect PII from education records, or you may want to train your authorized representatives yourself.

<u>Are there best practices for written agreements?</u>

You should consider the following items for inclusion in your written agreements for work under both the audit or evaluation exception and the studies exception. We note that this list may not cover everything you want in your agreement—you should look to the facts and circumstances surrounding the disclosure agreement and include all terms necessary to be clear about roles, responsibilities, and expectations for safeguarding PII from education records.

- *Bind individuals to the agreement.* It can be important to bind not just the entity to whom you are disclosing PII from education records, but also the individuals who will be accessing that data. There are several ways to accomplish this result. One way is to identify the individuals in the agreement itself, and have them execute the agreement in their individual capacity as well as having a representative execute the agreement for the entity. Alternatively, your agreement can require individuals accessing the PII from education records to execute affidavits of nondisclosure or other documentation indicating their individual agreement to handle the PII from education records properly.

- *Agree on limitations on use of the PII from education records.* Your agreement should be clear about limitations on the use of the PII from education records, meaning that it can only be used for the activities described in the agreement. Your agreement may also address methodological limitations, for example, identifying which data sets, if any, the PII from education records may be linked.

- *Agree to not redisclose.* The most basic provision of the agreement is to make clear that the PII from education records is confidential and must not be redisclosed through direct data disclosures or publishing results that allow individuals to be directly or indirectly identified. FERPA-permitted entities may wish to require that specified disclosure avoidance methodologies be applied, or may wish to review all results prior to publication, or both.

- *Specify points of contact/data custodians.* Your written agreements should specify points of contact and data custodians (the individuals directly responsible for managing the data in question).

- *Mention Institutional Review Board (IRB) review and approval.* While FERPA does not mention IRBs, research proposals involving human subjects may have to be reviewed and approved by IRBs, if required under protection of human subject regulations of the Department and other Federal agencies. If IRB review and approval is required or expected, this may be noted in the written agreement.

- *State ownership of PII from education records.* You may wish for your agreement to be clear that, in disclosing PII from education records to an entity, you are in no way assigning ownership of the PII or records to that entity, and that it may only be redisclosed with your permission or otherwise in compliance with FERPA and its regulations.

- *Identify penalties.* Your agreement could include penalties under State contract law such as liquidated damages, data bans of varying length, and any other penalties the parties to the agreement deem appropriate. You may want your agreement to create third-party beneficiary rights, e.g., allowing parties injured by a data breach to sue for damages. While FERPA itself has little flexibility for sanctions, you can include a wide range of appropriate sanctions in your written agreements.

- *Set terms for data destruction.* As discussed previously, written agreements for both studies and audits and evaluations are required to contain provisions dealing with the destruction of PII from education records when those records are no longer needed. The agreement could include a method for documenting the destruction, such as the use of notarized statements.

- *Include funding terms.* If the agreement involves cost reimbursement, these details could be specified.

- *Maintain right to audit.* You may want to include the right to conduct audits or otherwise monitor the entity to which you are disclosing PII from education records to periodically affirm that the entity has appropriate policies and procedures in place to protect the PII from education records.

- *Identify and comply with all legal requirements.* It is important to remember that FERPA may not be the only law that governs your agreement. The agreement could broadly require compliance with all applicable Federal, State, and local laws and regulations, and identify the legal authority (whether express or implied) that permits the audit, evaluation, or enforcement or compliance activity.
- *Have plans to handle a data breach.* While no one anticipates a data breach, data loss may occur. You may wish to include specific procedures in your written agreements detailing the parties' expectations in the event that PII from education records is lost, including specifying the parties' responsibilities with regard to breach response and notification and financial responsibility.
- *Review and approve reported results.* If applicable, the written agreement could specify the parties' agreements with respect to publication of results. For example, you may wish to review and approve reports prior to publication to make sure that they reflect the original intent of the agreement.
- *Define terms for conflict resolution.* The agreement could specify procedures for how disputes between the parties would be resolved.
- *Specify modification and termination procedures.* The agreement could specify how it can be modified or terminated. You may wish to provide specific provisions for termination based on improper handling of PII from education records.

What do I do if the terms of the written agreement are violated?

If the entity to which you have disclosed PII from education records without consent (whether under the studies exception or the audit an evaluation exception) violates the terms of the written agreement, you should evaluate your options under the penalty and termination provisions of the agreement. You may want to stop disclosing PII from education records to that organization, or pursue legal redress. If you have reason to believe that the entity has violated FERPA, you should contact FPCO.

How should the public be informed?

It is a best practice to keep the public informed when you disclose PII from education records.

- *Inform the public about written agreements.* Transparency is a best practice. You might want to post your data sharing agreements on your Web site, or provide some equivalent method to let interested parties know what data you are disclosing, the reasons it is being disclosed, and how it is being protected. While the Department generally recommends public posting of written agreements, parties are encouraged to review their contractual data security provisions carefully and redact, prior to publication, any provisions that may aid those seeking unauthorized access to systems. In certain instances a separate confidential IT Security Plan may be appropriate. For more information on data security best practices, see the Privacy Technical Assistance Center (PTAC) Web site: http://www2.ed.gov/policy/gen/guid/ptac/index.html

Who should I call if I have questions?

If you would like more information about best practices to protect PII from education records, contact the PTAC Help Desk at PrivacyTA@ed.gov or 855-249-3072.

If you are a parent, eligible student, school, LEA, or SEA and would like more information on FERPA, please call FPCO at 1-800-872-5327.

Appendix D. Additional questions for confirming specific data availability

Source: Authors.

The following list of specific K–12 data considerations can be used to help with planning a study that uses SLDS data.

Socioeonomic status: Are any data elements available that are typically used to create socioeconomic status variables? Is parental education available? How does the state define "economically disadvantaged" for federal accountability? Are both "economically disadvantaged" and FRPL eligibility available? Are "free" and "reduced-price" lunch status available separately?

School and district identifiers: Is the NCES ID provided? Are school and district names available along with any IDs?

Enrollment: At what time(s) of year is enrollment determined? When are students assigned to schools in the database? Do some students have multiple enrollment records?

State assessments: What tests are given in what subjects and grades? Does the state have a high school exit exam? End-of-course tests? When and how often are the different tests given? How is performance reported (scale score, raw score, proficiency level, and so forth)? Is student growth calculated for any of the tests? Have any of the tests changed over time so that trend data are not comparable? Is documentation available that addresses these questions?

Course-taking: Are course records available? For what grades and subjects? Is the course section or teacher identified for each course? How far back in time are the course records available? What is the coverage of course records at different grade levels? In particular, what proportion of high school students has complete course records? Does coverage vary significantly by district or year? Does the state have a standard course classification or coding system? Does the state have a standard or comprehensive course catalog? How can this catalog be obtained?

Grades: Are course grades available? For what grade levels? Is overall GPA available? For what grade levels? Is the GPA cumulative? Does the state have a standard grading scale? If so, what is it? If different grading scales are used across the state, are data included in the SLDS that indicate what scale is being applied?

Credits: Are course credits available? For what grade levels? Are both the credits attempted and credits earned provided? Are summary credit measures available? If so, how are they defined? Does the state have a standard credit system? If so, what is it (such as the Carnegie unit)? If different credit systems are used across the state, are data included in the SLDS that indicate what credit system is being applied?

College Board and ACT data: Are these data available in the SLDS at the student level? Which of the following are available: SAT, AP, ACT? Are both composite scores and subject scores provided?

Teachers: Can teachers be linked to specific students, classrooms, or course sections? To specific courses or schools? What information is available on teacher qualifications, experience, professional development, and compensation?