MAY 2014

COLLEGE READINESS INDICATOR SYSTEMS

A Technical Guide to College Readiness Indicators

The University of Chicago Consortium on Chicago School Research

CRIS

RESOURCE SERIES

Districts now have access to a wealth of new information that can help target students with appropriate supports and bring focus and coherence to college readiness efforts.[1] However, the abundance of data has brought its own challenges. Schools and school systems are often overwhelmed with the amount of data available. The capacity of districts to determine which data to include in their indicator systems to evaluate past efforts, monitor progress, and make strategic plans for the future lags behind the push for data use.[2] Indicator systems that incorporate too many data elements can be cumbersome and confusing. Data elements that are only weakly or spuriously related to college success can dilute the data system's potential to improve student outcomes and divert scarce resources to approaches that are less effective. By focusing on the best indicators of college success, schools and districts can target the right students for the right kinds of support, effectively evaluate their efforts, efficiently allocate resources, and bring coherence to their push to improve college readiness.

This technical guide is designed to help districts select the indicators that have the highest leverage for improving students' college outcomes. It is written with two audiences in mind. First, it assists district leadership in understanding how practical considerations intersect with the technical issues that must be considered when evaluating the extent to which a set of indicators matters in improving students' later outcomes. However, decisions about what indicators and outcomes to include in an indicator system should not be made by district leaders alone. Therefore this technical guide is also meant for district data analysts and educational researchers. It outlines the analyses that will help districts choose among the multitude of potential indicators they could incorporate into their data systems and college readiness efforts. It addresses one of the characteristics of effective college readiness indicators: being valid for the intended purpose (see *Selecting Effective Indicators*). For indicators to be considered valid, they must measure what they are intended to measure and predict the outcome of interest. This guide is not intended to provide an exhaustive list of issues for either audience to consider; rather it is meant to help each ask the right questions when selecting indicators to incorporate into an indicator system.

The technical guide is a part of a series of resources produced by the partners in the College Readiness Indicator Systems (CRIS) initiative: the Annenberg Institute for School Reform at Brown University (AISR), the John W. Gardner Center for Youth and Their Communities (Gardner Center), and the University of Chicago Consortium on Chicago School Research (UChicago CCSR). Although it is based on work conducted at UChicago CCSR on the Chicago Public Schools (CPS) and at the Gardner Center on the San Jose Unified School District (SJUSD), the technical guide is intended to be flexible enough to be used in other districts that are collecting different types of data and serving different populations of students. It is written to guide analysis in districts with multiple high schools, but the principles will also generally apply to analysis done for individual schools, for smaller districts, or at the state level.

After reviewing the benefits of a college readiness indicator system and describing how indicators should fit into the larger picture of district priorities and data use, this technical guide describes seven parts of the validation process. These parts of the process fall broadly into two stages: data preparation and validation.

How to cite this document:

University of Chicago Consortium on Chicago School Research. (2014). *A technical guide to college readiness indicators.* College Readiness Indicator Systems Resource Series. Seattle, WA: Bill & Melinda Gates Foundation.

DATA PREPARATION

Specifying variables for outcomes and indicators	Defining the sample	Examining descriptive statistics

VALIDATION

- Examining and comparing bivariate relationships between indicators and outcomes
- Developing a parsimonious but predictive set of indicators
- Setting cut points
- Making sense of it all

The data preparation stage includes the preliminary activities and decisions that must be completed before the work of validating indicators can begin. While the parts are described in different sections, they are not sequential steps; in practice, their order will often differ and activities from one part may occur in conjunction with those from another.

The validation stage includes four steps for comparing, selecting, and making sense of a set of indicators that are most predictive of the chosen high school and college outcomes. Each step is first described in general terms of the considerations and decisions that need to be addressed, and then examples detail how UChicago CCSR and the Gardner Center have handled the decisions. These examples are meant to serve as illustrations rather than rigid rules for other districts to follow.

What Is College Readiness?

Much of the early research on student performance indicators focused on high school graduation as an outcome; however, more recently, attention has turned to postsecondary outcomes, particularly college enrollment and college degree attainment. Efforts to improve postsecondary outcomes have largely coalesced under the banner "college readiness." David Conley (2007) defines college readiness as "being sufficiently prepared to enroll and succeed in a non-remedial, credit-bearing general education course during the first year in a post-secondary institution offering a baccalaureate degree or transfer to a baccalaureate granting institution."[3] Both Conley's work and the CRIS project have recognized that academic preparation alone is not sufficient for students to be successful in college. We emphasize a definition of college readiness that extends beyond academic preparation to include the concepts of academic tenacity and college knowledge.

This more expansive definition of college readiness brings new challenges for districts. It will require that they look beyond the traditional set of academic indicators that currently inhabit accountability systems and seek additional measures of academic tenacity and college knowledge (see *Menu of College Readiness Indicators and Supports*). Moreover, research can often send mixed messages about what is important for student success. For example, while numerous studies have documented the importance of high school grades for predicting high school graduation and college readiness, there is no real agreement about what other factors matter. Some studies emphasize the importance of taking Advanced Placement (AP) classes, others emphasize the role of noncognitive factors, and still others point to test scores as highly predictive of college readiness. This can be confusing to schools and districts and can make it difficult to prioritize among different indicators.

Furthermore, there are potential differences across districts and geographic regions as to how well indicators may predict outcomes for their students. Schools and districts have different data systems, with different ways of capturing information. And it is possible that what matters for one population of students may be different from what matters for a very different population of students, particularly given the types of postsecondary options that may be available in a given geographic area. Districts should consider what indicators matter the most for their population and their context and then reassess their needs as the student population or district priorities change.

Using a broader definition of college readiness may require districts to seek new data or delve more deeply into data not previously used for reporting. For indicators that have not been widely validated across student subgroups and contexts, districts may need to pay particular attention to whether these newer indicators are valid for their student population and context. This technical guide is designed to help districts conducting their own validation of indicators determine which are most predictive of their students' outcomes and should be included in their indicator systems.

Clarifying the Purpose of the Indicator System

Before beginning the data preparation and validation stages, it is important to have a clear sense of the purpose of the indicator system, how it will be used, and by whom (see *Selecting Effective Indicators*). There should also be clearly defined student outcomes that the indicators are intended to predict.

The process of validation should be guided by the larger process of how the indicators will be used and what the reporting system will look like. Decisions about the purposes of the indicator system, and which outcomes the system is intended to improve, should occur before any validation or selection of indicators.

Indicators can provide information at different levels for use by various actors within an indicator system. This has implications for how they are designed and validated. In this technical guide, we focus on student-level indicators but also discuss setting- and system-level indicators. Indicators can be used to identify individual students for intervention or leverage points for schools to focus their efforts on. They also can be used to assess a student's progress or to evaluate the effectiveness of policies and practices for improving college readiness. Indicators can further be used to identify areas for improvement in schools (setting-level indicators) and across the district (system-level indicators), or to inform the process of planning supports and interventions. Often, setting- and system-level indicators are aggregations of individual-level indicators—averages or percentages. However, indicators that work well at the individual level do not always work in the same way at the school level. The selection of indicators and their validation depend on how they are intended to drive improvements in student outcomes and school processes.

While the ultimate purpose of an indicator system is to improve students' educational attainment, indicators are also sometimes used as mechanisms for districts to communicate their priorities and evaluate progress toward intermediate goals around school culture and educational practices. If the purpose of an indicator is to drive particular changes in school practice, it may not matter whether it has a strong, direct relationship with college outcomes. For example, if the purpose of the indicator system is to build a college-going culture, and a district is using AP participation to help drive that effort, a district may choose to include AP course-taking as a measure of academic preparedness even if other indicators are more predictive of college outcomes. Likewise, a district may want to track the extent to which schools are implementing particular policies or track progress toward providing resources to students (e.g., counselor to student ratios or graduation requirements).

Key Terms and Definitions

This guide uses the following key terms and definitions:

OUTCOME: The measure of the goal that the indicator system is meant to support, such as college enrollment. In statistical terms, the outcome is the dependent variable.

INDICATOR: A measure that identifies whether students are likely to achieve the outcome or whether schools are supporting students in achieving the outcome. In statistical terms, the indicator is an independent variable.

VARIABLE or POTENTIAL INDICATOR: An indicator that has not yet been validated.

INDICATOR SYSTEM: The structures and supports that link indicators to action and include the tools for reporting indicators and tracking progress; the supports for building the capacity of adults to access, understand, and act upon the indicators; and the strategies, supports, and interventions for students identified by the indicators.

Data Preparation

To determine whether certain variables forecast outcomes like college enrollment, GPA, or graduation, the data used must meet certain requirements. The goal of the data preparation stage is to arrive at a comprehensive dataset of clearly defined variables that include the potential indicators, outcomes, subgroups, and background characteristics that will be used in the validation stage.

First and foremost, districts must collect data on students over time and assign each student a unique identifier (e.g., a student ID number) that stays with that student, even if the student changes schools. These identifiers should, in turn, connect students to relevant indicator, subgroup, outcome, and background variables.[4] (Indicator, subgroup, and outcome variables are considered in greater detail in subsequent sections of this guide.) To examine the relationship between indicators and postsecondary outcome variables, such as those measured by the National Student Clearinghouse (NSC) data, or postsecondary records in state data systems, it is critical that a unique identifier link all data sources.

Before beginning analysis, the data need to be cleaned so that errors are detected and corrected and so that each variable is precisely and consistently defined. One of the primary reasons for engaging in this process is to, as much as possible, have consistent definitions of variables so that the relationships between indicator and outcome variables can be compared over time and across students and subgroups of students. Over time, inconsistencies in data elements may come about for various reasons. A district may choose a different data system, warehouse provider, or data interface. A new policy may be enacted that requires or encourages data to be collected or defined in a different way.[5] New systems might use different names for the same variable, while variables with the same labels may in fact capture different information.[6] Such changes could also mean that queries used to produce complete datasets change, either in obvious or subtle ways. Whatever tools are used to access data, districts should try to maintain consistent definitions governing the collection and coding of underlying data. When changes are made, it is important to document these changes and ensure that everyone has a consistent and accurate notion of what the information being retrieved captures.

For example, if the variable denoting whether a student was enrolled in algebra includes pre-algebra in some years but not in others, then the relationship of the variable with later outcomes will not be the same in all years. For student-level indicators, a change in the relationship between indicators and outcomes could result in flagging the wrong students for intervention. For school- or system-level indicators, a change in the relationship between indicators and outcomes could lead to a misunderstanding of school or district progress, or an incorrect assessment of the efficacy of programs or strategies. To highlight changes in variable definitions, graphs or tables that show trends over time can use different symbols or colors to alert people that the variables have changed. When possible, analysts can try to equate old and new variables so that they have similar meanings over time. When variables change meaning over time, special studies may be required to understand the effect of the changes.

Keeping in mind the broad considerations described above, we now describe three key parts of the data preparation stage that should be attended to before moving on to analysis:

1. Specifying variables for outcomes and indicators

2. Defining the sample

3. Examining descriptive statistics

We describe the parts separately but they are not sequential steps and often overlap with one another.

Specifying Variables for Outcomes and Indicators

This step has two goals:

1. To identify variables that measure the potential indicators and outcomes of interest
2. To state clearly the decision rules defining each variable

Identifying Outcomes

Potential indicators cannot be validated without clearly defined outcome measures. Myriad outcomes could be considered important in a college readiness indicator system—high school graduation, strong college GPAs, enrollment in any college, enrollment in a four-year college, enrollment in a selective college, persistence past the first year, bypassing remedial college coursework, college graduation, etc. It is critical that the outcomes used reflect the purposes of the indicator system because the best predictors of one outcome might not be the best predictors of another. It even is possible that an indicator that is highly predictive of one outcome is of little or no use in predicting another. For example, a student-level indicator suitable for flagging students who are likely to drop out of high school may not be the best one to use for predicting four-year college graduation. Stakeholders may have different thoughts as to what the most important outcomes for students are. If it is not clear what the indicator system is intended to monitor or who it should identify, these stakeholders may be disappointed if the system is ineffective for improving outcomes that they care about the most.

While many outcomes are useful for understanding students' postsecondary success, the selection of outcomes used in an indicator system will be limited by the data that are available to districts. A commonly used source for data on college enrollment, persistence, and graduation is the NSC. This data source provides information on colleges across the country, allowing districts to examine postsecondary outcomes among their students who attend out-of-state institutions. However, about 7% of college enrollment is not captured by the NSC, either because the institutions do not participate in the NSC or because students have blocked their NSC records from reporting.[7] Another disadvantage of the NSC data is that it does not provide information on students' coursework, remedial course-taking, or course grades. Thus, while districts can determine whether their graduates are enrolling in college and persisting from year to year, these are blunt measures of their performance and may be more difficult to accurately predict than outcomes such as students' college grades or credit completion. Although the dynamics of college persistence make prediction more difficult, persistence may nonetheless capture some factors that are ultimately critical for degree attainment but that are not captured by more narrowly academic outcomes like college freshman GPA.

Example of Outcomes: Persistence in Four-Year Colleges

Two recent research reports from UChicago CCSR have used two-year persistence rates in four-year colleges, among CPS students who enrolled in four-year colleges, as an outcome.[8] Examining college persistence among students who enrolled in college separates the issue of persistence from that of enrollment when evaluating the relationship between indicators and outcomes. To create this outcome, several decisions had to be made.

One decision was who to consider as an enrollee in a four-year college. College enrollment and persistence are based on NSC data, which include the name of the postsecondary institution, the start and end date for the term in which the student was enrolled, and the student's status (e.g., full- or part-time). NSC also provides a graduation date if a student graduates from college. Students were considered enrolled in a four-year college if they made the immediate transition to college, had a term start date before November 1 and a term end date after September 15, and enrolled full-time. Using these dates ensured that a student's enrollment was captured regardless of the college's academic calendar.

It was also important to define persistence; UChicago CCSR coded students as persisting for two years if they were enrolled for four consecutive terms including the initial fall term following their high school graduation.[9] For spring semesters, students are considered enrolled if the term start date is before May 1 and the term end date is after March 1. The term dates are intended to capture at least one quarter of enrollment for students whose schools are on the quarter system. Students do not have to be enrolled in the same institution for all four terms to be considered as persisting in college.

In states that have longitudinal data systems that include postsecondary information, districts might be able to access information on college course-taking, course performance, and remedial course placement for students who enter public postsecondary institutions. Such records would be particularly useful for evaluating the quality of academic indicators of college readiness, as they could be used to determine whether the preparation students receive in high school is satisfactory for strong academic performance in college. The disadvantage of these data systems is that they do not include records for students who attend college out of state, and these students may systematically differ from those who decide to remain in state.

In general, postsecondary outcomes depend not only on whether students are academically prepared, but also on whether they can navigate issues such as financial aid, social integration into college, and personal situations that may pull them away from persisting or succeeding in their college classes. As a result, academic indicators such as GPA, high school coursework, and test scores might be less strongly associated with college persistence and college graduation than they are with academic outcomes such as remedial coursework, students' college grades, or credit completion. Indicators of academic tenacity, college knowledge, or other noncognitive factors might be more relevant to college persistence and graduation—which are the outcomes that matter the most for students. Designers of the indicator system should consider the mechanisms through which these distal outcomes are likely to be affected, as this should guide the selection of appropriate indicators.

Another issue to consider when selecting outcomes is whether the outcomes are suitable for different populations of students. While a district may be focused on increasing four-year college enrollment and graduation rates, that bar may be too low for high-achieving students who should be aiming for selective colleges and too high for low-achieving students who have little chance of succeeding in a four-year college. It is important that districts and schools make sure the indicator system can capture appropriate goals and benchmarks for different populations of students. While an indicator system might track aspirational outcomes so that schools with higher-achieving students are able to evaluate their students' progress, the district also should make sure to use outcomes that are targeted toward a typical student in the district.

Specifying Potential Indicators

When choosing which variables to include in the validation process, several factors are important to keep in mind (see *Selecting Effective Indicators*). Indicators should be meaningful and accessible to school staff. They should describe student experiences or school practices that are easily understood by administrators and practitioners. Even statistically valid indicators will be of limited utility if users have difficulty interpreting the information. A key prerequisite to data use is building the capacity of practitioners and administrators to understand what information is and is not contained in an indicator and how to translate that information into action.

Statistical validity is also critical. College readiness indicators should be valid predictors of college outcomes or of school processes that are intended to improve college outcomes. Scarce school resources make it particularly important to focus efforts in areas that are actionable and likely to lead to improvements. In Chicago, for example, UChicago CCSR's research showed that 9th grade failure rates were strongly predictive of high school graduation. Course failure is highly predictive of high school dropout, not just because it identifies students who will likely drop out, but also because failure to accrue credits prevents students from graduating. A number of schools began efforts to reach out to students who were at risk of failure in 9th grade, providing support and strategies to help them pass their courses. In many schools, this resulted in a large reduction in failure rates and an increase in students who were on track for graduation after their 9th grade year.[10] Subsequent research has now shown considerable improvements in graduation rates in those schools, coinciding with the efforts to reduce 9th grade failures.[11] Thus, the indicator system not only provided a means of identifying students' risk of graduating or dropping out, but also focused attention on what matters for high school graduation.

Decisions about how to construct indicators will need to consider how well those indicators work at different levels and whether they meet the different needs for which they are intended. It is easier for schools to set goals and strategies if the setting-level indicators are derivations of the same student-level indicators; however, the same indicators may not work as well at both levels. In general, it is more effective to begin with the validation of student-level indicators, and then build and validate the setting- and system-level indicators around those student-level indicators that are strongly predictive of college success. This ensures that the system is focused on those indicators that matter the most for college outcomes.

Once the student-level indicators are chosen, then setting-level indicators can be designed to provide insight into school and district progress around those indicators and whether efforts are being effectively organized around the areas that matter the most for improving student outcomes. As described below, sometimes the setting-level indicators will be aggregations of the student-level indicators. Other times, districts might want to include process-based indicators that are known to affect student outcomes or that are aligned with district policies. These might include measures of school climate or school practices.

Student-level indicators are useful for intervention, monitoring, and goal setting. They are more useful for intervention if they can be updated regularly enough to track progress and intervene if needed (e.g., student attendance rates, mid-term course grades, and completion of the Free Application for Federal Student Aid [FAFSA]). It is also important that they differentiate among students. For example, whether a student is on track for graduation provides only a general sense of his or her progress. Students' 9th grade GPA can be used to further pinpoint their probability of graduating high school or to target students with the appropriate level of support.

Setting-level indicators can provide useful information about the performance of schools. They show whether schools are making progress on the indicators that drive postsecondary success. Setting-level indicators include school-level aggregations of student-level indicators. However, they may not be exactly the same as the student-level indicators. Sometimes averages and percentages of success on student indictors are not very informative at the school level. Averages at the school level may mask differences among students within the school that matter for improvement strategies. For example, the indicator system may use students' attendance rates to flag students who need intervention, but a school's average attendance rate cannot explain the scope of attendance problems in the school. For example, consider two schools with the same 90% attendance rate: At one school almost all students may have a 90% attendance rate, while at the other school most students could have an attendance rate of 95% or higher but a significant fraction of students could have very low attendance, thus skewing the school's overall attendance rate. In this case, developing categories of attendance rates may be helpful in understanding the patterns of attendance and in identifying students who need intervention.

Certain indicators, such as FAFSA completion, taking algebra in 8th grade, or participating in a college access program, are most readily described in terms of occurring or not occurring. Dichotomous indicators like this are easily understood and easily tracked over time. An indicator that contains multiple categories or is a continuous measure of performance can be more difficult to interpret but can provide more information for differentiating among students, and often provides a better prediction of later outcomes. For example, 9th grade GPA (a continuous indicator) is slightly more predictive of high school graduation than whether or not students are on track for graduation (a dichotomous indicator) because the on-track indicator does not differentiate between students who were on track at the end of 9th grade and earned high grades from those who were on track and earned lower grades.[12]

In general, the more information that is used to create an indicator, the more precise the indicator will be. For example, including both core courses and non-core courses when calculating students' GPA provides a more robust picture of students' overall performance in their classes than core courses alone. Likewise, combining test scores from multiple years and subjects provides a more precise estimate of academic skills than a score from one year or subject. However, the robustness and precision of the indicator need to be balanced with the ease with which it can be understood and translated into reporting tools and linked to strategies and interventions. Indicators based on information accumulated over a longer period of time will necessarily be less sensitive to short-term interventions.

Indicators may also be confounded with each other or with other factors. For example, using a weighted GPA (in which students receive additional GPA points for taking honors, AP, or other advanced classes) confounds grades earned with types of classes taken. If the purpose of the indicator is simply to identify which students are more likely to attain an outcome, it may not matter that the indicator mixes different types of preparation. However, if the purpose is to diagnose what could be done to help students be more likely to obtain a degree, then it is useful to have separate indicators that allow practitioners to distinguish between whether changing course-taking patterns or improving course performance is a better approach to improving a given student's college readiness.

Finally, we should recognize that the division between indicators and outcomes can be somewhat fluid. For example, UChicago CCSR research has identified having a cumulative GPA of 3.0 as being a useful benchmark for whether students who enroll in four-year colleges graduate within six years.[13] In this case, GPA is being used as a college readiness indicator, but it may also be useful for a district to use cumulative GPA of 3.0 as an outcome in an analysis of 9th grade indicators so that the results can be used to identify 9th graders who are in need of interventions and supports for college readiness.

Defining the Sample

Another key part of the data preparation stage is defining the sample of students whose data will be used in the validation. This section lays out three areas where decisions must be made to define the sample.

The process of defining the sample is:

- guided by clear and consistent decision rules around which students will be included in the analytic sample;
- informed by how the indicators will be used and for which population of students and schools;
- limited by the availability of data in particular years (e.g., a change from one standardized test to another); and
- limited by the availability of data on certain variables for certain students (e.g., certain tests scores may not be available for English language learners [ELLs]).

The decision rules for determining which students to include in the analytic sample may differ from how a district decides which students are included in publicly reported data. The analytic sample is used to understand the relationship between potential indicators and outcomes, rather than to provide information on all students.

Step 1: Determining Cohorts

The first step in defining the sample is determining the broad parameters of the analysis and the data required. This includes deciding what type of cohort is being analyzed (e.g., grade cohorts, age cohorts, graduating high school students, or college enrollees) and how many years of data are available for the potential indicators and outcomes of interest. A cohort approach is very useful for longitudinal analyses in which student outcomes are compared over multiple years. Since it requires eight to 10 years to follow students from 9th grade to college graduation, it may make sense to use cohorts of graduating seniors when examining college outcomes. For example, in 2013, the most recent cohort we would be able to track to a six-year college graduation rate is the students who were in 9th grade during the 2003–04 school year. On the other hand, if the purpose of the indicator is to flag 9th grade students who are falling off the path to college graduation, then looking at 9th grade cohorts in addition to graduating senior cohorts is essential. Determining what kind of cohort to use is a balance between the purpose of the indicator and the availability of data. In every case, it is essential that students be assigned to *only one cohort* for a given analysis. For a variety of reasons, students can appear in the same administrative dataset multiple times, even when the data are drawn from the same school term.[14]

Grade-based cohorts assign students to a cohort based on whether they were enrolled in a given grade during a given year. Grade-based cohorts, particularly 9th grade cohorts, are useful in developing indicators for high schools. Grade repetition is a particularly noteworthy issue when creating an analytic sample based on students' grade level. At any given grade level, there are students who are registered in that grade for the first time, as well as students who have been retained and are enrolled in the grade for the second or even third time. Even when only one year of data is used, it is important to assign students to one cohort because the performance and outcomes of retained students can differ in the retained year versus their first year at a grade level. Including retained students in multiple cohorts biases the sample toward disproportionate representation of low-achieving students.

For example, SJUSD used several grade-based cohorts in their indicator analysis. First, they examined several college outcomes for students who began 8th grade in 2003–04 through 2005–06. This meant that three cohorts of students could be tracked from middle school through college to see what indicators predicted outcomes of interest. Similarly, SJUSD built grade-based cohorts of ELLs beginning in kindergarten to see how quickly ELLs were reclassified and whether this reclassification rate influenced postsecondary outcomes.

Graduating cohorts are based on students who graduate during a given school year, and may include those who graduated during the following summer term.

Depending on the purpose of the indicator, graduating cohorts may exclude students who graduated from alternative high schools or who received a GED because these students did not meet the same graduation requirements as graduates of regular high schools; these decisions will depend on the district's philosophy and the purpose of the indicator system. Graduating cohorts are often used in analyses of postsecondary outcomes, particularly college enrollment. When examining college performance, persistence, or graduation, the indicator system might continue to follow all high school graduates, or it might limit the analysis to college enrollees or students who enroll in a four-year college.

Occasionally, UChicago CCSR has used age-based cohorts rather than, or sometimes in addition to, grade-level or graduating cohorts. Age cohorts are calculated based on students' ages at a given date in the year. The date will depend on the cutoff for school enrollment. For example, in Chicago, students are eligible for kindergarten if they are five years old as of September 1. A defining feature of age cohorts is that students remain in the same age cohort regardless of their progression or retention from grade to grade. This is an advantage when studying policies that may affect the number of students who progress in school, such as test-based promotion policies, or for comparing academic progress for different student subgroups. Age cohorts can also be more inclusive than grade-based cohorts, which may not include students who do not make it to the older grades (e.g., 9th grade cohorts would not include students who drop out in 8th grade).

Step 2: Determining Which Students and Schools to Include

The second step in defining the sample is determining which students and schools should be included and excluded from the analysis. If, in clarifying the purpose of the indicator, it is determined that the results should be generalizable to the entire district (e.g., school accountability for graduation rates) then all students and schools should be included. However, if the purpose is to analyze typical public school students, then exclusion of some students and schools may be more appropriate. Students enrolled in alternative, magnet, selective enrollment, or other special types of high schools may have substantially different school experiences and patterns of mobility than students enrolled in neighborhood schools. As a result, it may not make sense to include these students in the validation sample, as they may bias the results.[15] For example, some students transfer from outside a district into alternative schools because viable options (e.g., dropout recovery programs) do not exist in other districts or because they are enrolled in schools that are not meant to be long-term options (e.g., "juvenile hall" schools). Including these students and schools in the analysis may distort the relationship between the indicators and outcomes that exists for more typical students. Schools that serve students with severe disabilities are another group that may have significantly different outcomes based on the indicators than observed for other students. Subsequent analysis can examine whether the indicators function in the same way at different types of schools.

After determining which students to include in the analytic sample, it is important also to identify any subgroups of students for whom the relationship between the indicator and outcome may be different, or for whom different interventions or strategies might be appropriate. Such subgroups might include students with disabilities or those enrolled in alternative schools, and again the selection of subgroups should be guided by the purpose and goals of the indicator system. Membership in a subgroup of interest must be captured in the data set. When identifying subgroups of students, it is important to understand how changes in policies or district practices or changing demographics in the district may influence the composition of subgroups.[16] A significant change in the composition of a subgroup can affect the relationship between indicators and outcomes; when subgroup performance is compared over time, it is helpful to understand whether changes in composition may be influencing the relationship between indicators and outcomes.

For example, changes in classification policies around who is considered an ELL can result in more or fewer students classified, resulting in different measured outcomes for that subgroup. In California, districts must follow certain state and federal guidelines for classifying ELLs, but they can also use additional local criteria, which may differ across districts. Policy changes can also have an effect. In 2006, the State of Illinois changed the test used to measure English proficiency, which may have affected who was identified as an ELL. ELL membership is also constantly changing over time; as students gain proficiency in English, they exit ELL subgroup status and are replaced by newly enrolled students who have not yet achieved proficiency. Taking into account former and current ELLs is essential for evaluating a district's success with its ELL population. For some purposes, it may be more useful to classify students according to whether they are now *or ever have been* classified as ELL.

Step 3: Establishing Decision Rules for School Enrollments

The third step is establishing decision rules for what constitutes a school enrollment. These decisions must be made for both high school (or middle school) and college enrollments. Determining the decision rules around enrollment for when a student should be included in the analytic sample is complicated. The decision might be based on the length of time a student has been enrolled in the district (e.g., at least 10 days, at least one semester, etc.) or on whether a student was enrolled at a particular point in time (e.g., by September 30 or the end of the school year). Key considerations include the purpose of the indicator and the type of cohort being used (grade cohorts, age cohorts, or high school graduating cohorts).

If the cohort is defined as college enrollees, decision rules should include whether only full-time students are counted, since students who take only one or two courses are less likely to graduate. It should be determined whether students must make an immediate transition to college, because when students delay enrollment the role of high schools in helping them gain access is diminished. It also should be decided how long a student must be enrolled (one or two terms) to be considered enrolled in college.

For example, SJUSD used a variety of definitions of college enrollment. In some models, any student who showed up in the college's administrative data system was counted as enrolled. In others, only students who were still enrolled at the beginning of the next semester were included. There are many ways to define enrollment, and it is important to understand how different definitions may change the relationship between indicators and outcomes.

Care should be taken with where to attribute enrollment for highly mobile students. Students may have multiple enrollments in schools because they transferred schools within the same district mid-semester or mid-year. This is particularly common in school transition years such as 9th grade, as students may register for one school but then show up in a different school in the district, or may only attend a school for a short time before deciding to attend a different school. Students may also leave and re-enter the district, even within the same school year. Other students may be attending night school or a special program in one school in addition to their regular school and thus have more than one enrollment.

The decision rules about what constitutes an enrollment and establishing which enrollment is primary should guide the determination of which observations to include in the analytic sample. Excluding mobile students from the analysis entirely can bias results as these students often have weak academic performance. The question of how to manage mobility becomes particularly sensitive when school-level measures are being reported. Schools are differentially affected by mobility; high-performing schools tend to lose their weakest students and accept only the strongest students as transfers. However, low-performing schools tend to have high rates of mobility in and out of district, which may lead to greatly reduced sample sizes if mobile students are not included. As a result of student mobility, the sample size is rarely consistent when following a single cohort across years or comparing an initial grade cohort to its graduating cohort. Mobility also introduces problems in determining to which school a particular student should be assigned for the purpose of producing setting-level indicators that are aggregations of student-level indicators. If students are being followed longitudinally, it often makes sense to assign them to the school in which they were enrolled at the beginning of the period of study, when the cohort was defined. For graduating cohorts, it makes sense to assign them to the school from which they graduated. Each of these decisions will be shaped by the local context, as well as state accountability policies.

Another key consideration is the limitations on data availability for populations such as highly mobile students, ELL students, or students with disabilities. For example, a district may want to use 8th grade test scores and attendance to flag incoming 9th grade students for additional supports to help them pass their coursework. If the district excludes students with disabilities from testing, the samples for the analysis of the relationship between attendance and 9th grade grades would differ from that for test scores. In this case, it will also be important to make sure that using different samples does not bias the results. Because it is preferable to use the same analytic sample across indicators for comparability purposes, it may be worth exploring ways to address missing data.[17]

Example of Decision Rules: Creating a Dataset of First-Time 9th Graders

UChicago CCSR has a longstanding body of research that looks at early indicators of high school dropout. The early research focused on first-time 9th graders and defined them as students who enrolled in the 9th grade and were not registered in grades 9–12 in any previous year. Because the findings were meant to apply to a broad range of students, as many schools and students as possible were included in the analysis. Public charter schools were excluded from the analysis because UChicago CCSR lacked data on public charter school students' grades. Subsequent work tracked 9th grade cohorts into later grades and to high school graduation.

To create a dataset for first-time 9th grade cohorts, UChicago CCSR researchers started with the group of students who enrolled in 9th grade in a given academic year and who received course grades for at least one semester (thereby excluding students who were only enrolled for a very short period of time). The researchers then looked at enrollment records for up to five years prior to exclude students who were previously enrolled in 9th grade. First-time 9th grade students with verified transfers out of CPS at any point during high school were also excluded from the sample.

For analysis of indicators after 9th grade, new students who enter the district after 9th grade were added after looking at enrollment records for up to three years after the initial 9th grade school year. Although these students would not be included in any analysis of 9th grade course performance (because they would not have any CPS records from 9th grade), they were included in analyses of later high school outcomes, such as 10th or 11th grade on-track status, and graduation rates. These students were added to 9th grade cohorts based on their age; that is, they were assigned to the cohort in which they would have belonged had they entered 9th grade at age 14. This approach was used because the available records did not indicate when these students actually entered 9th grade for the first time, and we could not infer that information from their initial CPS grade level placement, since many students repeat a grade once they enter CPS. We used the same approach for students with disabilities who enrolled in a CPS high school for the first time and for whom no grade level information was available.

Examining Descriptive Statistics

Once the decision rules have been set and the data set has been assembled, the third part of data preparation is to produce thorough descriptive statistics on both indicator and outcome variables. The goal is to gain a better understanding of the data and the basic relationships among variables, check for any discrepancies in the data or issues with missing data, and identify the baseline sample sizes for the different cohorts. Descriptive statistics (means, standard deviations, ranges, Ns, and frequency distributions for variables coded into categories) should be produced for each variable being considered on the cohorts of students included in the analysis, at the school level, and for any relevant subgroups. Each variable should be checked for anomalous values and missing data. This is also the time to check for duplicate records. When discrepancies occur, analysts often will need to examine individual student records, and comparisons across schools, to determine why there are inconsistencies. The descriptive statistics should be calculated separately for each year as well as for each cohort. This can reveal whether there have been significant changes in coding, recordkeeping, or variable definitions over time. Any unexpected patterns should be examined further.

The descriptive statistics allow us to get an indication of how much variation there is between students on both indicator and outcome variables. Variation on indicators is important: If all students look exactly the same on a particular potential indicator, that variable cannot predict variation in student outcomes. Variation on outcomes is similarly important: If few students achieve a given outcome, it may be necessary to consider additional outcomes. Similarly, at the setting level, it also is important to look for variation across schools to make sure that the indicators and outcomes provide information that can be used to evaluate progress and guide strategies.

Examining descriptive statistics also can reveal issues with the availability of data on different indicators. Of particular concern is that nonrandom patterns of missing data can lead to biased results that are not applicable to certain groups of students.[18]

For example, of the students who graduated from a CPS high school in 2005, only 84% had an ACT score, even though the state requires that all juniors take the ACT. This suggests that there may be a reason for students not taking the ACT that is worth investigating, such as subgroups of students who may not be accurately represented in the data. A subgroup analysis shows that approximately 25% of the students who did not have an ACT score were special education students; some may have been exempted from taking the exam. Another 10% attended alternative schools. Another large portion of students, 25% of all graduates without ACT scores, were never classified as juniors during the spring test-taking period and so did not take the ACT. Finally, students who were missing ACT scores had significantly higher absence rates and so may have missed the test because they were not in school and did not take a make-up exam.

Validation: Describing the Relationship Between Indicators and Outcomes

In this section, we describe a four-step process for examining the predictive validity of potential indicators:

1. Examining and comparing bivariate relationships between indicators and outcomes
2. Developing a parsimonious but predictive set of indicators
3. Setting cut points
4. Making sense of it all

This section is primarily written for data analysts who will be conducting the validation, although district administrators may also find it useful for understanding the key considerations that go into this aspect of validating an indicator.

Examining and Comparing Bivariate Relationships between Indicators and Outcomes

A first step in evaluating the predictiveness of a potential indicator is to look at the bivariate relationship between the indicator and the outcome. The goal of this step is to determine which potential indicators are most predictive of the outcomes of interest. These relationships can be preliminarily examined using descriptive statistics and graphical displays. For example, it is useful to calculate the mean value of the outcome (e.g., the percentage of students who graduate from a four-year college) at each level of the indicator (e.g., GPA by tenth of a grade point) and then display these findings graphically. Graphical displays can provide an understanding of the strength and shape of the relationship between indicator and outcome variables and can show whether the relationship between indicator and outcome is similar at all levels of the indicator or whether there are thresholds where the relationship becomes stronger.

When analyzing the bivariate relationship with continuous variables, it is helpful to maintain as much detail as possible about the relationship between indicator and outcome. However, each point that is shown should represent a large number of students to reduce sampling error. Thus, instead of graphing points for variables with many values (e.g., attendance rate), it is generally necessary to divide the variable into ranges ("bins") (e.g., 98–100% attendance, 96–98%, etc.). When examining relationships among variables, it is good practice to graph only data points that are based on 100 or more students. Results based on fewer than 100 students can give a distorted picture of the actual relationship between indicator and outcome. Additionally, when reducing the number of values a variable can take, it is important to re-examine the relationships it has with the outcome to ensure that the relationship is not an artifact of bin size.

For categorical variables, contingency tables ("cross-tabs") can be used to examine the relationship between variables. For example, SJUSD generated two-by-two tables on the outcome of college enrollment (enrolling versus not enrolling) by each of a series of indicators like completion of an algebra course in 8th grade (completion versus noncompletion). This approach shows what percentage of students who did or did not complete algebra in 8th grade went on to enroll in college.

Once the nature of the relationship is understood between each potential indicator and the outcome, the next step is to determine which variables actually have the strongest relationships with the outcome. A quick way to do this is to compare correlations of each indicator with the outcome. If the descriptive statistics showed that some indicators have nonlinear relationships with the outcome (where the strength of the relationship changes at different values of the indicator), then it might be more appropriate to run regression equations to better model the relationship. This might mean transforming the indicator variable and/or including a quadratic term, or chunking values of the indicator into categories and representing category membership by including a series of dummy variables in the equation. Comparisons between potential indicators should then be done using statistics such as an R^2 or pseudo-R^2 that describe how much variability in the outcome is explained by the potential indicator. Analysts can then use the R^2 statistic from the models to compare the size of the relationship of each indicator with the outcome.

In addition to pseudo-R^2s from logistic regression equations, measures of sensitivity and specificity can further illuminate the predictiveness of potential indicators of dichotomous outcomes. The sensitivity of an indicator measures the proportion of students achieving a given outcome who are correctly identified as doing so; it also is known as the true positive rate. In a study seeking to identify which students are likely to drop out of college, for example, a highly sensitive indicator would be able to correctly identify a large proportion of those students who leave college without a degree. The specificity of an indicator is the proportion of students not achieving a given outcome who were correctly identified as not doing so; it also is known as the true negative rate (e.g., the proportion of those students not dropping out who are correctly identified). It is useful to examine the inverse of the specificity of an indicator (1 – specificity), also known as the false positive rate. It is a measure of the proportion of students who are incorrectly identified as achieving an outcome. For example, in the case of dropping out of college, the false positive rate would be the proportion of students who graduated but were incorrectly predicted to drop out. Similarly, the inverse of the sensitivity of an indicator, the false negative rate or Type II error, is the proportion of students who dropped out of college but were identified as not being at risk of doing so. As schools and districts consider using indicators to guide intervention strategies, they must balance these different pieces of information about indicators. If the purpose of the indicator is to identify students, schools will want to correctly identify as many at-risk students as possible, so that targeted interventions can be put in place, but at the same time minimize the number of false positives so that valuable resources are not wasted. It can be helpful to graph the sensitivity and specificity for each indicator.[19]

It is a good idea to look at the bivariate relationship between an indicator and an outcome with as many years of data as possible. While graphing multiple years of data may be somewhat cumbersome, focusing on the R^2 or pseudo-R^2 is a relatively quick way to ensure that the relationship between indicator and outcome is stable over time. The bivariate relationship between indicator and outcome should also be examined for subgroups of students, as well as for the general population and for different samples. Differences by subgroups may mean that interventions designed to move an indicator may have differential effects on the outcome. One potential issue is that an indicator may be much less predictive, or more predictive, for one group than another group. In this case, the formula and indicators to use for flagging risk may be different. Alternatively, the general relationship of indicators with outcomes may be the same, but the threshold that indicates risk may be different. This was the case in Chicago for specific subgroups of students. Analyses of 9th grade early warning indicators of high school graduation for students with disabilities and ELLs showed similarly sized relationships between the same early warning indicators and outcomes, indicating the same general system of indicators could be used for these subgroups. However, the thresholds for risk were lower for particular subgroups of ELL students than among the students in the general population.[20]

Developing a Parsimonious but Predictive Set of Indicators

The set of indicators with the strongest relationship to the outcome should be identified using the bivariate statistics. The next step is to use multivariate analyses to determine whether the indicators can be used in combination with one another or whether they are redundant with one another. Indicators that have a strong relationship with an outcome simply because they are related to other factors may not be good choices for an indicator system. Conversely, indicators that have an indirect relationship with the outcome may be useful if efforts to change them also change the mediating factors.

To determine whether an additional indicator provides independent (i.e., nonredundant) information, we use a series of simple regression models (using either ordinary least squares for continuous outcomes or logistic regression for dichotomous outcomes) in which the outcome is regressed on one or more indicators. The first model used includes only the indicator having the strongest bivariate relationship with the outcome. The indicator with the second strongest bivariate relationship is then added to the model (including interaction terms between the first and second indicators). If the adjusted R^2 from the second model is not significantly greater than that for the first model, then the second indicator does not add much predictive power over and above the first indicator and can be removed from the model. However, if the adjusted R^2 increases somewhat, the two indicators together provide more information about who is likely to achieve the outcome of interest. The process of adding additional indicators to the model, checking the adjusted R^2, and removing redundant indicators should be continued until the entire set of indicators with the strongest bivariate relationships with the outcome has been evaluated. Analysts should use the model-based R^2 or pseudo-R^2 rather than the significance test for a given coefficient to determine whether a variable remains in the model. There is often a high degree of multicollinearity in these models, and coefficients for individual indicators may be insignificant simply because they are related to other predictors in the model. Also, when using large samples, indicators may be significant predictors of an outcome but may not actually add much to the predictive power of the model.

Setting Cut Points

Establishing the predictive relationship between indicators and outcomes does not, in itself, provide decision rules to guide the practical use of the indicator. While it may seem obvious that for an indicator like GPA, "higher is better," school and district personnel require guidance as to how indicators should be used to trigger interventions or to allocate resources. Decision rules are most often formulated in terms of cut points, along with rules stating that if an indicator falls below (or sometimes above) the specified level, then some action is to be taken.

It is generally best to base cut points on data about both the distribution of the indicator (its mean and variability, overall and for subgroups) and the relationship of the indicator to valued outcomes. Distributional data will indicate what proportion of students (or schools) would be flagged for action if the cut point were set at a particular level. Too high a threshold for action might fail to trigger intervention in some cases where it was needed, but too low a threshold might flag so many cases for action that support structures for remediation are overwhelmed and resources are diluted. Sensitivity analyses will help determine at what point the relationship between indicators and outcomes is strong and which represent a feasible level for a large proportion of students. A similar logic applies to choosing cut points as to choosing outcomes: Indicator cut points should not be set so high that few students will achieve them, nor so low that the relationship between indicator and outcome weakens substantially.

The relationship between indicator and outcome also is critical. Attendance below some threshold may be a serious risk factor, but it could turn out that above that threshold, further improvements in attendance do not matter much for the outcome. Cut points should be set at a level where a student's standing above versus below the cut point has a clear, empirical relationship to the probability of attaining the outcome of interest.

Cut points for desired outcomes (e.g., a specified probability of a college freshman GPA of B– or better) may be mapped back to corresponding cut points for indicators. Throughout the system, outcomes at one level may be indicators for the next level. Thus, college readiness may be mapped backward to successively lower grade levels, creating a roadmap for on-track progress through the system.

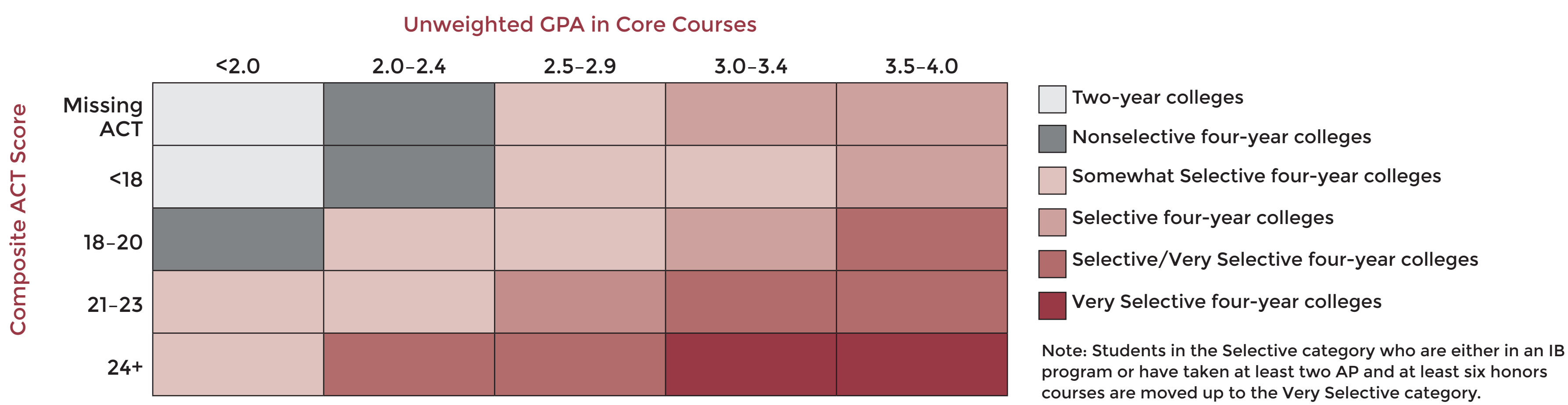
While cut points are a useful component of a parsimonious indicator system, dividing the entire distribution of measures into too few categories may result in an imprecise metric of accomplishment that is less than optimal for gauging improvements over time. The size of year-to-year improvements can depend entirely on whether there are many students with scores that are near the cutoff. Small improvements in the measure can result in many more students meeting the benchmark if there are many students close to the cutoff score, while large overall improvements can go unnoticed if there are few students with scores close to the cutoff. It is therefore important that the proportion of students meeting a benchmark is not used to measure school or district improvement; rather, meeting benchmarks serves as a minimum goal for students and for targeting interventions. The distribution of the underlying measure should be monitored over time.

Making Sense of It All

Validating student indicators can provide districts and schools with important information about the areas of student performance that are most strongly linked to student outcomes. If an indicator is highly predictive of a given outcome and it captures malleable behaviors, attitudes, or skills, then implementing interventions for students who have poor performance in one or more of those areas may ultimately lead to improved outcomes for those students.

Creating categories based on cut points can provide guidance to practitioners about how to target interventions and supports for students based on the needs highlighted by the indicators. One example of using categories and cut points to provide guidance to practitioners and evaluate how schools and students are performing is the college match measure developed by UChicago CCSR. College match compares the type of colleges students enroll in (e.g., two-year college or Barron's selectivity category) to the type of college the student would likely have access to given their course performance (unweighted GPA in core classes), their ACT scores, and their involvement in college preparatory AP and International Baccalaureate (IB) coursework (see Figure 1). The cut points for ACT scores and for GPA were established based on both ease of understanding and empirically testing the robustness of different cut point decision rules. The categories indicate the minimum GPA, ACT scores, and advanced course-taking that CPS graduates would need for a high likelihood of acceptance to a college of a given selectivity. A student is considered to be enrolled in a "match" college if they enroll in an institution that aligns with the type of school to which their qualifications would give them access. While college match was originally intended to provide a quantitative measure of the extent to which students and schools were participating in the college search and application process, it has also become a tool used by counselors and other practitioners to guide students in recognizing the range of their college options.

FIGURE 1 Categories for access to college types based on CPS graduates' GPAs and ACT scores and patterns of enrollment



Source: Roderick, M., Nagaoka, J., Coca, V., & Moeller, E.; with Roddie, K., Gilliam, J., & Patton, D. (2008). *From high school to the future: Potholes on the road to college.* Chicago, IL: University of Chicago Consortium on Chicago School Research.

Conclusion

College readiness indicator systems have the potential to make the complicated task of preparing students for college more manageable. The information provided by indicators can help practitioners and administrators focus their attention on key leverage points for improving the postsecondary outcomes for students. Determining what indicators to use in identifying leverage points and what supports and interventions are needed is the foundation of this work. The indicators selected to be a part of a system should be based on evidence of a strong link between the indicator and the intermediate or postsecondary outcome of interest, and they should serve the goals of the district. This technical guide is meant to walk analysts through the process of validation and allow them to ask critical questions as they proceed with this essential work.

Endnotes

1. The rise in data use over the past decade has largely been driven by federal legislation, such as the No Child Left Behind Act and the American Recovery and Reinvestment Act, which include provisions for data use, and by policy and foundations, such as the Bill & Melinda Gates Foundation and Michael & Susan Dell Foundation, which have made substantial investments in data as a strategy for education reform.
2. Dantow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing school systems use data to improve instruction for elementary students*. Los Angeles, CA: University of Southern California, Center on Educational Governance.
3. Conley, D. T. (2007). *Redefining college readiness*. Eugene, OR: Education Policy Improvement Center.
4. Background variables, including income level, parental education, and race, may not necessarily be indicators of performance but may be related to outcomes or needed for subgroup comparisons. They can also be used to help identify students' levels of risk for a particular outcome.
5. For example, the size of the special education population in CPS increased substantially after an initiative intended to end social promotion was implemented. See: Miller, S. R., & Gladden, R. M. (2002). *Changing special education enrollments: Causes and distributions among schools.* Chicago, IL: University of Chicago Consortium on Chicago School Research. The increase could be due to better identification of actual disabilities or it could be due to reclassifying students so that they are not subject to the policy. As the proportion of students who were classified as having a disability increased, it is likely that the composition of this group (background characteristics and achievement level) also changed.
6. For example, in compliance with state and federal requirements, CPS created three new race categories. Students that were previously categorized as "Asian / Pacific Islanders" are now categorized as either "Asian" or "Native Hawaiian and Pacific Islanders." Students who have multiple races but were previously categorized as white, Hispanic, or African American are now categorized as "Multiple Race."
7. Other issues with the NSC data are discussed in Dynarski, S. M., Hemelt, S. W., & Hyman, J. M. (2013). *The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes.* Cambridge, MA: National Bureau of Economics.
8. Coca, V., Johnson, D. W., Kelley-Kemple, T., Roderick, M., Moeller, E., Williams, N., & Moragne, K. (2012). *Working to my potential: The postsecondary experiences of CPS students in the International Baccalaureate Diploma Programme*. Chicago, IL: University of Chicago Consortium on Chicago School Research; and Roderick, M., Coca, V., Moeller, E., & Kelley-Kemple, T. (2013). *From high school to the future: The challenge of senior year in Chicago Public Schools*. Chicago, IL: University of Chicago Consortium on Chicago School Research.
9. This is meant as an example; there are many other reasonable ways to define persistence.
10. Allensworth, E. (2013). The use of ninth-grade early warning indicators to improve Chicago schools. *Journal of Education for Students Placed at Risk, 18*(1), 68–83.
11. Roderick, M., Kelley-Kemple, T., & Johnson, D. W. (forthcoming). *Validating the on-track indicator.* Chicago, IL: University of Chicago Consortium on Chicago School Research.
12. Allensworth, E., & Easton, J. Q. (2007). *What matters for staying on-track and graduating from Chicago Public High Schools: A close look at course grades, failures and attendance in the freshman year.* Chicago, IL: University of Chicago Consortium on Chicago School Research.
13. Roderick, M., Nagaoka, J., & Allensworth, E.; with, Coca, V., Correa, M., & Stoker, G. (2006). *From high school to the future: A first look at Chicago Public School graduates' college enrollment, college preparation, and graduation from four-year colleges.* Chicago, IL: University of Chicago Consortium on Chicago School Research.
14. Other issues can come from duplicates created when merging data sources or the flaws and quirks of administrative data and human data entry errors. In any case, it is good practice to check for duplicates and then troubleshoot around causes and find ways to address them.
15. If alternative, magnet, selective enrollment, or other special types of high schools are not included in the analytic sample, UChicacgo CCSR treats transfers in and out of these schools as if they were from out-of-district schools.
16. See earlier discussion about data cleaning starting on page 5.
17. A discussion of options for imputing missing data is beyond the scope of this resource. See, for example, Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage; and Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: John Wiley & Sons, Inc.
18. In the case of data not missing at random, it is particularly important to conduct sensitivity analyses to assess how sensitive results are to different ways of addressing missing data.
19. Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. *The High School Journal, 96*(2), 77–100.
20. Gwynne, J., Pareja, A. S., Ehrlich, S. B., & Allensworth, E. (2012). *What matters for staying on-track and graduating in Chicago Public Schools: A focus on English language learners*. Chicago, IL: University of Chicago Consortium on Chicago School Research.