

High School Completion Rates¹

Rob Warren

University of Minnesota

2/5/2010

There are very few social phenomena as important in the United States as completing high school. Business leaders, academics, policy makers, and others know that the American economy is fueled in part by a steady supply of well-educated workers; indeed states and local communities frequently seek to improve their business climates by producing a more educated work force. The trend toward greater accountability in public education has meant that state and local high school completion rates are now key metrics (along with test scores) used to evaluate the performance of teachers, schools, districts, and state education agencies. These rates are front-page news, and they have real political and practical consequences for schools and school personnel. For individual young people, of course, completing high school is rightly seen as a prerequisite for social and economic success. Researchers in any number of disciplines have shown that failing to complete high school means a lifetime of markedly increased risk for poverty, violence, incarceration, family turmoil, and ill health, among other maladies.

Given its economic, social, political, personal, and academic importance, it is manifestly imperative to develop accurate measures of the high school completion rate (or else of the high school dropout rate, which most people take to equal one minus the completion rate). Such measures are useful to business leaders seeking to attract or relocate businesses; to families seeking to settle in new communities; to taxpayers and policymakers looking to evaluate the performance of public school systems; to academics interested in explaining the social, economic, and other correlates of educational attainment; and to a diverse array of other parties.

Most observers begin with a sense that it should be quite straightforward to ascertain the high school completion or dropout rate. However, confusion sets in when those observers are confronted with widely publicized high school completion and dropout rates.

What was the high school completion rate in 2005? According to the U.S. Department of Education's *Digest of Education Statistics 2007*, "74.7 percent of public high school students graduated on time [in 2004-05]" (Snyder et al. 2008: 3). Similarly, "about 71 percent of 9th graders make it to graduation four years later [in 2005]" according to *Diploma Counts 2008* (Editorial Projects in Education 2008). However, the *Digest* also reported a dropout rate for 16 to 24 year olds of only 9 percent in 2005 (Table 105) and the Annie E. Casey Foundation's *Kid's Count 2008 Data Book* reported a dropout rate for 16 to 19 year olds of only 7 percent in that year (2008). How can we make sense of a 70 to 75 percent completion rate in light of a dropout rate below 10 percent?

The situation gets even more confusing at the state level. In Texas in 2005, for example, the percentage of 9th graders who completed high school was either 74 percent or 69 percent, according to the *Digest of Education Statistics* and *Diploma Counts 2008*, respectively. However, the Texas Education Agency (TEA) reported that the "overall graduation rate for the [Texas] class of 2005 Grade 9 cohort was 84.0 percent" (Ramirez et al. 2006: viii). In the same report, the TEA noted that the "longitudinal dropout

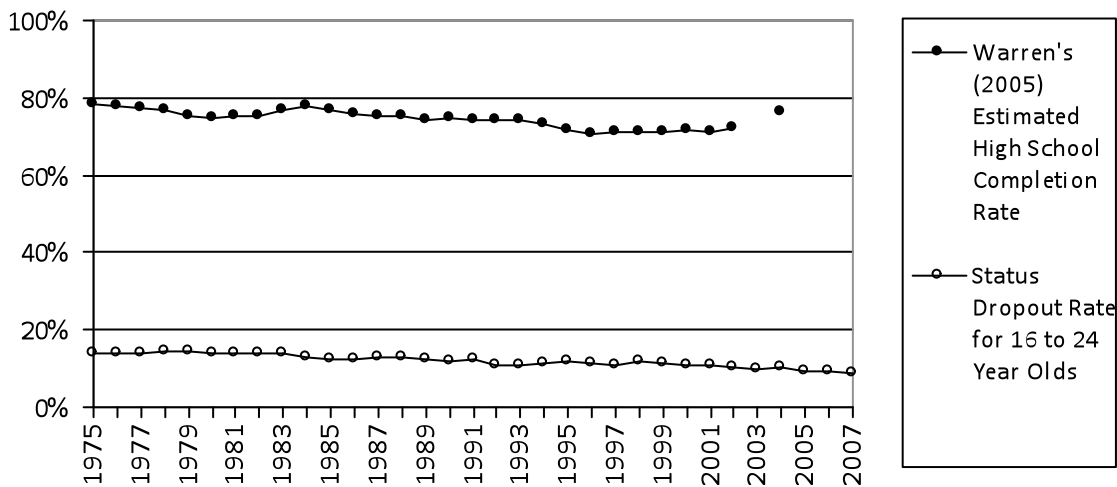
¹ Most of the material in this chapter was originally prepared by me for the National Research Council's workshop on "Improved Measurement of High School Dropout and Completion Rates," which was convened in Washington DC in October 2008. The published report from that workshop borrows material from my initial contribution.

rate for the class of 2005 Grade 9 cohort” was only about 4 percent. The *Kid’s Count 2008 Data Book* reported a dropout rate for Texas 16 to 19 year olds of about 8 percent in 2005. How should Texans make sense of high school completion rates that range widely from 69 to 84 percent in 2005, especially in light of dropout rates below 10 for that year?

This confusion contributes to misunderstandings about the absolute rates at which young people complete and drop out of high school, but it also leads to seriously flawed conclusions about differences in these rates *over time* and *across localities*.

How has the high school completion rate changed *over time* in the United States? Figure 1 shows the status dropout rate for 16 to 24 year olds and Warren’s Estimated Completion Rate for 1975 through 2004 (U.S. Department of Education 2006; Warren 2005; Warren and Halpern-Manners Forthcoming). The former measure is a high school dropout rate. The latter is a high school completion rate. Perhaps surprisingly, *both* figures declined modestly over these three decades. That is, somewhat fewer young people completed high school. At the same time, somewhat fewer young people dropped out.

Figure 1. Dropout and Completion Rates, 1975 to 2004

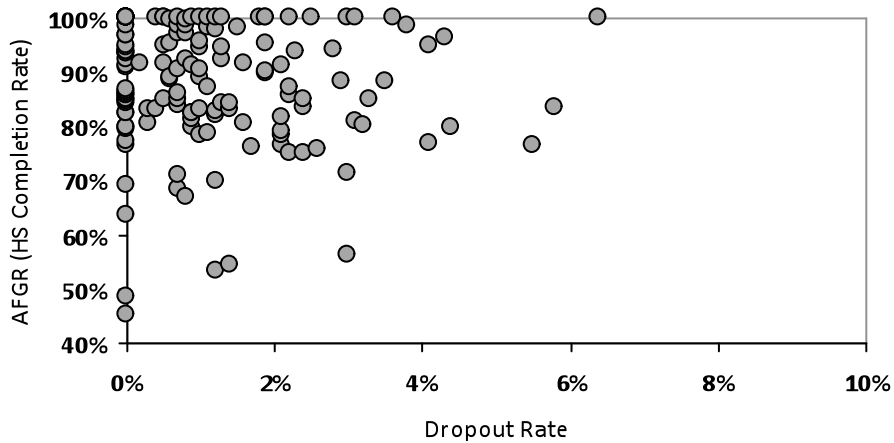


How do completion and dropout rates vary *across locales*? For example, business leaders or potential homebuyers might wonder which school districts have the best record of accomplishment with respect to rates of high school completion. Figure 2 plots the U.S. Department of Education’s Adjusted Freshman Graduation Rate² for 2006 against state-reported dropout rates³ in 2006 for 149 school districts in South Dakota. We might expect a very strong negative relationship between these two measures; in fact, their correlation is only about -0.07.

² Obtained from the U.S. Department of Education at <http://nces.ed.gov/ccd/drpagency.asp> on 12/28/2009.

³ Obtained from the South Dakota Department of Education’s website at <http://www.doe.sd.gov/ofm/applications/statdigest/default.asp> on 12/28/2009.

Figure 2. AFGR vs Dropout Rate in South Dakota School Districts



Why do widely used measures of high school completion and dropout rates differ so much from one another? As I explain below, the answer has to do with (a) differences in what various estimates are designed to accomplish; (b) differences across estimates in the conceptual and technical definition of both the numerator and denominator of the rates; and (c) differences in the accuracy of the data used to produce these estimates. Although it is intuitively appealing to try to decide on a *single* common metric for quantifying such a centrally important social phenomenon, below I argue that there is no single “best measure” of high school completion or dropout. *Although some measures are much better (i.e., more accurate) than others, the “right” measure depends on the purpose for which it is intended.* In the end, I argue for the development and use of a set of statistically sound common metrics that would accurately portray rates of high school completion and dropout and that would serve the several academic and policy purposes for which such measures are used.

Three Reasons that Different Measures Paint Different Pictures

The first reason that dropout and completion rate estimates differ so much from one another is that they differ with respect to what they are trying to measure. Analysts operationalize high school completion and dropout rates in different ways because they have different conceptual or practical reasons for making those measurements. There are at least three main reasons for measuring high school completion and dropout rates.

One reason to measure high school completion and dropout rates is to describe the amount (or lack) of human capital in a population. Motivated by this purpose, the objective is to quantify the share of people who have not completed high school and who are no longer enrolled. A second reason to measure high school completion and dropout rates is to characterize schools’ “holding power” (Hartzell et al. 1992). How well do schools move young people from the first day of high school to successful high school completion? A third reason to measure these rates is to characterize students’ success at successfully navigating high school from beginning to end. In contrast to measures of “holding power” -- which are intended to describe an attribute of *schools* --- these measures are designed to describe an attribute of *students*. How well do students manage to move themselves from the first day of high school to successful high school completion?

Note that different actors use high school dropout and completion rates for different purposes. For example, economists or business leaders may be interested in characterizing levels of human capital in a population, while education policy makers may instead focus on quantifying schools' holding power. On the other hand, education researchers may be more interested in characterizing students' experiences navigating educational institutions. It seems unlikely that any one common metric — one methodology for describing completion or dropout rates — can satisfy these diverse demands and interests, however technically sound or valid it may be.

For the purposes of describing amounts of human capital in a population, the timing of high school completion --- how long ago or at what age people completed high school --- is not important. Nor does it matter exactly how young people complete high school --- that is, by obtaining a diploma, a GED, obtaining a certificate of completion, completing an adult education program, or some other way. Anyone who has completed high school is considered to have surpassed a necessary threshold, regardless of his or her age at the time of surpassing it. For the latter two purposes, however, both the timing of high school completion and the manner in which young people complete high school can be important. For example, schools may only be deemed successful at moving a young person through to completion if those young people obtained regular diplomas "on time," typically within four years.

Given these differences in intended purpose, it becomes less puzzling to read in the *Digest of Education Statistics* that "74.7 percent of public high school students graduated on time" despite the fact that only 9 percent of 16 to 24 year olds were dropouts in 2005 (Snyder et al. 2008). The former estimate is explicitly intended to describe the share of young people who complete high school on time and by obtaining a diploma---an attribute of schools. The latter estimate is clearly intended to describe the share of young people who lack the human capital associated with high school completion. Presumably many of the $100 - 74.7 = 25.1$ percent of the 9th graders in the fall of 2001 who did not go on to graduate from high school with a diploma by spring 2005 were still enrolled or will complete high school later, via a GED or another alternative credential. Given the 9 percent dropout rate, we might presume that eventually about $100 - 9 = 91$ percent of these young people will eventually complete high school one way or another.

The second reason that high school dropout and completion rates differ so much from one another is that there are technical differences in how various measures are constructed. This is true even among measures whose purpose is to quantify the same concept (e.g., "holding power" or "human capital"). All high school completion and dropout rates boil down to a ratio: In the numerator is the number of high school completers or dropouts and in the denominator is the number of people at risk of completing or dropping out of high school. However, even when analysts are trying to measure the same concept, they frequently differ with respect to who counts as a "completer" or "dropout" in the numerator and who is considered "at risk" in the denominator.

If we are measuring high school completion rates, who should be counted as a "success" in the numerator? Those who obtain regular high school diplomas surely count. But what about GED recipients? Or those who obtain certificates of completion? And should we count a student as a "success" if they do not complete high school in a timely manner? If we are measuring high school dropout rates, who should be counted as a "failure?" Everyone agrees that students who leave school, never return, and never obtain any high school credential should count as dropouts. But what about students who drop out and then obtain a GED or complete a state-certified adult education program? Or students who drop out for a while and then re-enroll in high school? Even when analysts agree on

the numerator and can quantify “success” or “failure,” they frequently differ in their definition of the denominator --- or those at risk of “success” or “failure.” How should the denominator account for geographic migration into or out of a population of interest? How should students who are expelled or otherwise “pushed” out of high school be counted in the denominator? When students transfer from one school to another during high school, in which school’s denominator should they be included?

The third reason that high school dropout and completion rate estimates differ so much from one another is that they differ with respect to the accuracy of the data used in their construction. As a result, even when measures are intended to quantify the same concept and agree on the technical definition of the numerator and denominator, their estimates may differ quite a bit. There has been considerable debate in recent years, for example, about the accuracy of measures that rely on individuals’ self-reports of their high school completion status (Heckman and LaFontaine 2007; Mishel and Roy 2006; Warren and Halpern-Manners 2007). Another prominent example is the recent debate about how to accurately measure the number of 9th graders who begin high school for the first time in a particular year (e.g., Warren 2005). These and other accuracy issues are described in more detail below.

A Catalogue of High School Completion and Dropout Rates

This section includes definitions of status rates, event rates, individual cohort rates, and aggregate cohort rates and reviews of their intended purposes, technical definitions, and data sources. Subsequent sections include discussions of issues of accuracy and bias, both in the technical definition of particular measures and in the underlying data typically used to produce them. In the end, I discuss which measures most accurately and reliably represent high school completion and dropout rates for each of the three basic purposes described above.

Status Rates

A status rate reports the fraction of a population that falls into a population sub-category at a given point in time. The most common and visible example is the status dropout rate --- the share of people in a population who are high school dropouts --- but status enrollment rates and status completion rates are occasionally reported as well. For example, in *Dropout Rates in the United States, 2006* the U.S. Department of Education reported that 9.3 percent of 16 to 24 year olds were not enrolled in school and did not have any high school credential in October of that year (Laird et al. 2008). In that same month, 87.8 percent of 16 to 24 year olds were status completers --- that is, they were not enrolled in high school and held some sort of high school credential. These measures’ focus on 16 to 24 year olds --- as opposed to the population of all adults --- is designed to offer cohort-specific information. Some analysts use slightly different age ranges. For example, the Annie E. Casey Foundation reports status dropout rates for 16 to 19 year olds in its annual reports (e.g., 2008).

The numerator of the status “dropout” rate reflects the number of people who have not obtained any high school credential and are not working toward one. The fact that many dropouts subsequently re-enroll in high school, obtain GEDs, or earn high school credentials in other ways is immaterial; so is the age at which young people complete high school.

Status dropout and completion rates are usually calculated using cross-sectional data on individuals in the target population. All that is required is information about individuals’ age, enrollment status, and high school completion status. Status dropout (and occasionally completion) rates are routinely calculated using data from the decennial U.S. Census, the Current Population Survey (CPS), and the

American Community Survey (ACS). The Census and (in recent years) ACS data contain a sufficient number of observations to report status dropout and completion rates at the national, state, and sub-state levels.

All status dropout and completion rates are measures of the amount (or lack) of human capital in a population. They are very poor measures of schools' holding power or of young people's success at navigating the secondary school system and persisting in school. This is because such measures do not differentiate between those who obtain high school credentials by graduating with a diploma and those who obtain GEDs or other alternative credentials, and do not consider whether credentials were obtained after the typical ages of high school completion. A very low status dropout rate may reflect very high holding power of schools, or it may belie a situation in which schools have very low holding power and many young people obtain alternate credentials in their late teens or early twenties. What is more, status rates do not consider the location of high school dropout or completion. A geographic area with a very low status dropout rate may have schools with very high holding power, or it may attract many in-migrants who either are enrolled or have high school credentials. For example, counties with higher-technology industries or large post-secondary institutions tend to have relatively lower status dropout rates. This likely says more about the human capital of people who move to those counties and less about those counties' schools' holding power.

Event Rates

An event rate reports the fraction of a population that experiences a particular event over a given time interval; by definition, everyone in the population is at risk of experiencing that event during the period. The most frequently reported example is the event dropout rate --- the proportion of students who exit school during an academic year without completing high school (Komiski 1990). Between October 2005 and October 2006, 3.8 percent of 15 to 24 year olds who began that year in grades 10, 11, or 12 left school without obtaining a diploma or an alternate credential (Laird et al. 2008). Some states report variants of this measure among their students. For example, North Carolina (2004) eliminates the age restriction and reports these rates for 9th through 12th graders and separately for 7th through 12th graders.

It is intuitively tempting to use a series of grade-level-specific event *dropout* rates to infer an event *completion* rate that describes completion across the four years of high school. Imagine four separate event dropout rates, focusing on 9th, 10th, 11th, and 12th graders in academic years X, X+1, X+2, and X+3, respectively --- call them $E_{9th, X}$, $E_{10th, X+1}$, $E_{11th, X+2}$, and $E_{12th, X+3}$. One might take $(1 - E_{9th, X}) \times (1 - E_{10th, X+1}) \times (1 - E_{11th, X+2}) \times (1 - E_{12th, X+3}) = G$ to represent the fraction of incoming 9th graders who persist across the four years of high school, but this would be inaccurate. First, the tendency for students to repeat grades can downwardly bias these grade-specific event dropout rates; $E_{9th, X}$, in particular, will be downwardly biased in most populations and thus G will be upwardly biased. Second, students who drop out of two different grades upwardly bias these grade-specific event dropout rates and downwardly bias G.

Event dropout and completion rates can be calculated using either cross-sectional data or longitudinal data on individuals in the target population. All that is required is information about individuals' enrollment status in two consecutive academic years, their completion status in the second of those years, and (under some formulations) age. Enrollment status is typically measured at the beginning of each academic year so that event dropout measures can more clearly represent the dropout during well-defined academic years. Using longitudinal data, current enrollment status is measured in each of two consecutive academic years. This is, for example, how North Carolina estimates its event dropout

rate for 7th to 12th graders (2004) using longitudinal administrative data. Using cross-sectional data, enrollment status in the previous year is measured retrospectively and current enrollment status is measured contemporaneously. For example, the October supplement to the CPS asks about enrollment in both the current and the previous October. The event dropout rate is routinely calculated using data from the October supplement to the CPS. Because of sample size restrictions, it is sufficiently reliable only when calculated for the nation as a whole and for a few larger states (Winglee et al. 2000).

Because they are measures of the share of a population that experiences a particular event --- typically high school dropout --- over the course of a specific time interval, event dropout and completion rates can be used either to describe schools' holding power or else to describe young people's ability to navigate successfully the school system. Whether an event dropout rate fairly characterizes the former or the latter is a matter of (a) how "success" and "failure" are defined in the numerator and (b) how the population is defined in the denominator. If the goal is to measure something about schools --- their "holding power" --- then the numerator is determined by how schools define "success" (e.g., they are explicit about whether GEDs and other alternative credentials are treated as equivalent to regular diplomas) and the denominator is restricted to those continuously residing in a well-defined geographic area (typically school districts or states). The resulting event dropout rate should thus speak to the experiences of only those students for whom the school district or state is formally responsible. If the goal is to measure something about students --- the rate at which they succeed in navigating the secondary school system --- then the denominator need not be geographically restrictive.

Individual-Level Cohort Rates

Individual-level cohort rates are derived from longitudinal data on individuals, all of whom share a common characteristic at one point in time (e.g., they are all students entering high school). These rates report the fraction of individuals who transition into a particular status by a subsequent point in time. In the case of dropout and completion rates, students are usually followed from the beginning of high school until at least the normal time of high school completion. Individual-level cohort rates typically report the fraction of students who ended the period as high school dropouts or completers. For example, Mishel and Roy (2006) use data on students who were 8th graders in the spring of 1988, and report that 12 years later 83.0 percent had obtained regular diplomas, 7.7 percent had earned GEDs, and 9.3 percent had obtained neither credential. Likewise, the Texas Education Administration used a cohort rate to determine that "[f]or the class of 2007 Grade 9 cohort, 78.0 percent of students graduated, and 2.0 percent received GEDs" (McMillion et al. 2008).

The longitudinal data used to compute individual-level cohort dropout and completion rates has typically come from one of two places. The first is state- or district-level administrative data on students who are under the supervision of schools. Here, analysts typically select students who were first-time 9th graders in the fall of one academic year (constituting the denominator), and then count the number of students who obtain diplomas (for cohort graduation rates), obtain any secondary credential (for cohort completion rates), or leave school without obtaining any credential (for cohort dropout rates). As with event dropout rates, states and districts differ with respect to what counts as "success" and "failure" in the numerator of individual-level cohort dropout rates computed using administrative data. Some agencies count only regular diplomas as "successes," while others also count GEDs and alternative credentials.

The second source of longitudinal data used for computing individual-level cohort dropout and completion rates is the series of major longitudinal studies undertaken by the National Center for

Education Statistics (NCES).⁴ These include the 1972 sample of seniors in the National Longitudinal Study of the High School Class of 1972 (NLS); the 1980 and 1982 samples of sophomores and seniors in High School & Beyond (HS&B); the sample of 8th graders in the 1988 National Educational Longitudinal Study (NELS); and the sample of sophomores in the 2002 Educational Longitudinal Study (ELS). In each case, NCES began by selecting stratified, nationally representative samples of students in the focal grade(s) in the base year. Those students were then followed periodically, allowing for the computation of cohort dropout and completion rates. Here, the denominator of individual-level cohort rates consists of all sampled students who were in the same grade at the same point in time. The numerator can be defined by the analysts, as each data set includes information about whether, when, and how students completed school.

Although similar in many ways, individual-level cohort dropout and completion rates computing using longitudinal administrative data differ in a few ways from those computed using longitudinal data (from NCES or elsewhere) on samples of students. Most importantly, because the former are constructed in order to characterize dropout and completion rates in a particular state or school district, the denominator of these rates must be adjusted to account for entry and exit into the population of interest. Most prominently, the denominators of such rates are frequently adjusted to account for migration into and out of the jurisdiction in question and for transfers to other educational settings (like GED or adult education programs). Likewise, some states or localities adjust the denominator to account for students who are incarcerated or expelled or who die. Conceptually, these students are no longer considered to be at risk of dropping out from or completing high school. Individual-level cohort dropout and completion rates computing using NCES or other longitudinal survey samples are typically computed for the nation as a whole, and so issues of migration into and out of the population of interest are less of a concern.

Individual-level cohort dropout and completion rates computed using longitudinal administrative data from schools or districts are typically intended to characterize *schools* --- that is, to describe their holding power or their success in graduating students. Spurred by the language in the Elementary and Secondary Education (“No Child Left Behind,” or NCLB) Act of 2001, all 50 states have entered into the National Governors Association’s (NGA) Compact on State High School Graduation Data. Among other things, the compact stipulates that all states will construct a “standard, four-year adjusted cohort graduation rate” which is calculated by “dividing the number of on-time graduates in a given year by the number of first-time entering ninth graders four years earlier.” In the numerator, graduates are those receiving a regular high school diploma in four academic years. In the denominator, adjustments are made to account for (among other things) in- and out-migration. In October 2008, former U.S. Secretary of Education Spellings announced regulations as part of No Child Left Behind that move toward codifying this definition of individual-level cohort graduation rates.

In contrast, individual-level cohort dropout and completion rates computing using longitudinal data on samples of students are typically intended to characterize *students* --- that is, to describe their success in navigating through post-secondary educational institutions. The difference in purpose is partly attributable to the nature of the longitudinal cohort samples themselves. The NCES cohorts, for example, are infrequently conducted and are not sufficiently large to characterize school districts or even states; they are thus not useful for NCLB or other school accountability purposes. Because student-level microdata from NCES and other sample surveys are publicly available, analysts are

⁴ It is also possible to compute rates using other longitudinal sample surveys like the National Longitudinal Surveys.

generally more focused on understanding the individual student-level correlates of high school completion or dropout. Researchers are also at liberty to construct individual-level cohort dropout and completion rates that suit their own purposes and that differ with respect to the technical definitions of both the numerator and denominator. Finally, because these sample surveys include students in both public and private schools, researchers are in a position to generalize their estimates to all students.

Aggregate Cohort Rates

Aggregate cohort rates are designed to approximate individual-level cohort rates in situations in which individual-level longitudinal data are not available. Beginning with a count of the number of individuals who share a common characteristic at one point in time (e.g., students entering high school), aggregate cohort rates estimate the fraction of individuals who transition into a new status (e.g., high school completion) by a subsequent point in time. In the case of dropout and completion rates, this frequently takes the form of a rate with the number of dropouts or completers in a cohort in the numerator and an estimate of the number of people at risk of dropping out or completing in the denominator. For example, the NCES Averaged Freshman Graduation Rate “estimates the proportion of public high school freshmen who graduate with a regular diploma 4 years after starting 9th grade” (Laird et al. 2008: 2).

Aggregate cohort dropout and completion rates are primarily based on administrative data collected from schools. All that is required is information about the number of students completing and/or dropping out at a point in time and the estimated number of students at risk of doing so; this information is usually tied to specific cohorts of incoming students. The denominator usually takes the form of an estimate of the number of first-time 9th graders in a particular cohort, which can be calculated in various ways (described below). Many states, for example, have computed aggregate cohort completion rates by dividing the number of completers in the spring of academic year X by the sum of the number of completers in academic year X and the number of dropouts in academic years X, X-1, X-2, and X-3.

In recent years, the most widely publicized aggregate cohort graduation rates have been computed by dividing the numbers of regular diploma recipients in academic year X by an estimate of the number of first-time 9th graders in academic year X-3. Prominent examples include Swanson’s Cumulative Promotion Index (Editorial Projects in Education 2008; Swanson and Chaplin 2003); the NCES Averaged Freshman Graduation Rate (Seastrom et al. 2006a; Seastrom et al. 2006b); Greene’s Adjusted Completion Rate (Greene and Forster 2003; Greene and Winters 2006); and Warren’s Estimated Completion Rate (Warren 2005; Warren and Halpern-Manners 2009). These measures are all based on data from the State Nonfiscal Survey of the NCES’ Common Core of Data (CCD). The CCD data --- which are described in more detail below --- include annual counts of numbers of graduates and numbers of students enrolled in each grade for all states and school districts in the United States. All of these measures also make some effort to adjust the denominator --- which is based on the *total* number of 9th graders --- to account for migration into and out of the “at risk” population and to account for bias introduced by the fact that some 9th graders are not first-time 9th graders.

As described above, individual-level cohort dropout and completion are employed either to describe *schools’* holding power (when the rates are based on longitudinal administrative data) or to describe *students’* success at persisting in school (when the rates are based on longitudinal samples of students). As approximations of individual-level cohort rates, aggregate cohort rates have serious limitations for both purposes. Aggregate cohort rates are frequently used to characterize the holding power of states, districts, and schools’ --- as exemplified, for example, by the publicity surrounding the Editorial Projects

in Educations' annual *Diploma Counts* (e.g., 2008) reports. However, they are conceptually imperfect for this purpose because of their inability to distinguish "on time" graduates from other graduates in the numerator --- all that is known is the total number of graduates. As a result, they are less useful for accountability purposes because they do not meet the graduation rate definition spelled out by NCLB or by the NGA Compact. When used to characterize students' success at persisting in school, this issue about numbers of "on time" graduates is less consequential. However, aggregate cohort rates based on data from public schools say nothing about the experiences of those students who attend private schools, and so they are less useful as descriptors of students' experiences. For all purposes, the inability to adjust aggregate cohort rates' denominators properly to account for migration and grade retention leads to systematic bias under most conditions (as described below).

Although there are inherent weaknesses in aggregate cohort rates for either of their main purposes, they do have one major advantage: They can be computed for every state and every local education agency in the country in a technically consistent manner, and they are available annually going back for many years. This allows for comparisons over time and across locales. The same cannot be said of individual-level cohort rates: When they are based on longitudinal administrative data, they are generally not computed in a consistent manner across locales (although the NGA Compact may change that) or over time. When they are based on longitudinal survey data on students, they cannot be generalized to districts or even (in many cases) states. For any analyses of change in rates over time and/or across locales, aggregate cohort rates are all that are available. This fact makes understanding their conceptual and technical problems (described below) all the more important.

Accuracy and Bias in Dropout and Completion Rates

The most widely used of the status, event, individual-level cohort, and aggregate cohort rates described above come from one of three sources of data: Cross-sectional sample surveys (e.g., the CPS and the Census), longitudinal sample surveys (e.g., NELS), and administrative data (e.g., CCD and state agency data). Observers have frequently claimed that discrepancies between various high school dropout and completion rates are due to biases or other limitations inherent in the data that underlie them. How valid are claims of bias or flaws in the major sources of data used for constructing the most widely used dropout and completion rates? Each of the status, event, individual-level cohort, and aggregate cohort rates described above can be decomposed into two component parts: Their counts of the number of people at risk of completing or dropping out (in their denominators) and their counts of actual completers or dropouts (in their numerators). After an overview of the major data resources used for constructing dropout and completion rates, the sections that follow describe their strengths and weaknesses with respect to deriving accurate numerators and denominators.

Overview of Main Data Resources

This section describes the design and weaknesses of the major data resources used to compute high school dropout and completion rates. These include cross-sectional sample surveys (using the CPS as the main example); longitudinal sample surveys (using NELS as the main example); cross-sectional administrative data (using the CCD as the main example); and longitudinal administrative data.

1. Cross-Sectional Sample Surveys

The most widely used cross-sectional sample surveys for the purposes of measuring high school dropout and completion are the CPS⁵, the decennial US Census, and (in more recent years) the ACS. Because the decennial Census is not conducted more frequently, and because the ACS is a relatively new resource, the CPS has served as the central cross-sectional data resource for decades.

The CPS is a monthly survey of more than 50,000 households and is conducted by the Bureau of the Census for the Bureau of Labor Statistics. Households are selected in such a way that it is possible to generalize to the nation as a whole and, in recent years, to individual states and other specific geographic areas. Individuals in the CPS are broadly representative of the civilian, non-institutionalized population of the United States. In addition to the basic demographic and labor force questions that are included in each monthly CPS survey, questions on selected topics are included in most months. Since 1968 the October CPS has obtained basic monthly data as well as information about school enrollment—including current enrollment status, public versus private school enrollment, grade attending if enrolled, most recent year of enrollment, enrollment status in the preceding October, grade of enrollment in the preceding October, and high school completion status. In recent years, the October CPS has also ascertained whether high school completers earned diplomas or GED certificates.

There are a number of conceptual and technical problems with CPS-derived measures of high school dropout and completion, particularly when computed at the state level. First and foremost, the sample sizes are not large enough to produce reliable estimates of rates of high school completion or dropout at the state or sub-state levels (Kaufman 2001; Winglee et al. 2000). Even when data are aggregated across years—for example, in the Annie E. Casey Foundation’s *Kids Count* (2008) measure—the standard errors of estimates for some states are frequently so large that it is difficult to make meaningful comparisons across states or over time. What is more, by aggregating across years the resulting measure no longer pertains to specific cohorts of incoming students; as a result, CPS-based measures are not useful for assessing schools’ holding power or for describing the dropout or completion rates of specific cohorts of young people.

Second, until 1987 it was not possible to distinguish high school completers from GED recipients in the CPS; since 1988 October CPS respondents who recently completed high school have been asked whether they obtained a diploma or GED, but there are serious concerns about the quality of the resulting data (Chaplin 2002; Kaufman 2001). Third, as noted by Greene and Winters (2002: 7), “[status] dropout statistics derived from the Current Population Survey are based on young people who live in an area but who may not have gone to high school in that area.” To the extent that young people move from state to state after age 18, CPS-based state-level high school dropout rates—particularly status dropout rates based on 16 to 24 year olds—may be of questionable validity (see also Kaufman et al. 1992). Fourth, some observers have expressed concern about coverage bias in the CPS, particularly for race/ethnic minorities. The CPS is representative of the civilian, non-institutionalized population of household residents in the United States, and so young people who are incarcerated, in the military, or homeless are not represented. To the extent that these populations differ from the rest of the population with respect to frequency and method of high school completion, there is the potential for bias in estimates. Finally, substantial changes over time in CPS questionnaire design, administration, and survey items have made year-to-year comparisons difficult (Hauser 1997; Kaufman 2001).

⁵ The CPS is not, of course, really a cross-sectional survey. Households are included in eight monthly CPS surveys over the course of 16 calendar months. However, because of difficulties associated with linking surveys, deriving appropriate longitudinal weights, and other technical considerations, virtually all researchers use the CPS as though they were cross-sectional.

It is possible to overcome some, but not all, of these limitations of the CPS by using data from the ACS or the decennial Census. Sample sizes are larger, enhancing the reliability of state- and urban-level estimates. Both the ACS and the decennial Census include individuals who are institutionalized or in the military, and so generalizability is enhanced. However, it is still not clear how accurately ACS respondents report whether they obtained GEDs or regular high school diplomas. In addition, the ACS and Census share with the CPS the limitation that sampled young people may not be living in the state in which they attended high school. As a result, the ACS and the Census are useful for constructing status (but not event) dropout and completion rates that describe the human capital of populations. Measures derived from the CPS, the ACS, and the Census are not well suited to describing schools' holding power or to describing young people's success at navigating the secondary school system.

2. Longitudinal Sample Surveys

As noted above, the most widely used longitudinal sample surveys are those produced periodically by the NCES. This discussion focuses on NELS because it has been the center of so much research and debate on the measurement of high school dropout and completion rates in recent years (Greene et al. 2006; Kaufman 2004; Mishel 2006; Mishel and Roy 2006).

NELS is a longitudinal survey of the 8th grade student cohort of 1988. In the base year, the sample included approximately 25,000 randomly selected students in 1,000 public and private schools across the United States. In addition to the data collected from student interviews, NELS contains information from parents, school administrators, teachers, and student transcripts. The initial student cohort has been followed-up on four occasions, in 1990, 1992, 1994, and 2000. Students who dropped out of school between survey waves were also interviewed, and in early follow-ups, the sample was "freshened" with new sample members in order to make the first and second follow-ups cross-sectionally representative of 1990 sophomores and 1992 seniors, respectively. The content of the surveys include students' school, work, and home experiences; educational resources and support; parental and peer influences; educational and occupational plans and aspirations; delinquency; and many others (Curtin et al. 2002).

For the purposes of measuring high school dropout and completion rates, the key feature of NELS (and other longitudinal sample surveys) is that it includes information about whether and when cohort members dropped out of school and whether and how they obtained secondary school credentials. A key design feature of NELS is the availability of transcript data on high school enrollment, dropout, and completion. In the absence of coverage bias and non-participation, NELS data would provide very accurate estimates of high school dropout and completion rates --- albeit for a single cohort.

A number of technical issues raise questions about the accuracy of dropout and completion rates based on NELS (Kaufman 2004); these issues also arise in the context of other longitudinal sample surveys. First, the base year NELS sample excluded many students with limited English proficiency or mental or physical disabilities; NCES gathered supplementary information from these students later, but it is not clear how often this supplemental information is used in calculating NELS-based dropout and completion rates. Second, as noted by Kaufman (2004: 119), "[s]ince NELS is a sample survey, it is subject to the same potential for bias due to non-response and undercoverage that CPS has." Third, transcripts data are frequently unavailable for dropouts or alternative completers; this is due in part to the logistical difficulties inherent in collecting such data and in part to non-response by schools (Ingels et al. 1995). Some of these problems are overcome by the use of sample weights in the NELS, but in the

end NELS --- like all longitudinal sample surveys --- has a difficult time retaining difficult to follow populations like high school dropouts.

3. Administrative Data

Each state maintains its own system for counting the numbers of students who are enrolled in each grade (usually at the beginning of each academic year) and the numbers of students who obtain regular diplomas and other high school completion credentials. These counts are usually aggregated up from the schoolhouse level, and are increasingly linked to longitudinal data systems. At the national level, cross-sectional administrative data on enrollments and numbers of completers are compiled as part of the Common Core of Data (CCD).

Compiled by NCEES, the CCD is the federal government's primary database on public elementary and secondary education. Each year the CCD survey collects information about all public elementary and secondary schools from local and state education agencies. One component of the CCD—the State Nonfiscal Survey—provides basic, annual information on public elementary and secondary school students and staff for each state and the District of Columbia. CCD data from the State Nonfiscal Survey includes counts of the number of students enrolled in each grade in the fall of each academic year and the number of students who earned regular diplomas, who earned other diplomas, and who completed high school in some other manner in the spring of each academic year.

Few observers have raised concerns with CCD enrollment counts. These counts closely correspond to states' own reports of enrollment numbers (for good reason) and to CPS estimates of the number of public school students enrolled in grades 8 and 9; as discussed elsewhere (Warren and Halpern-Manners 2009), discrepancies in enrollment counts between the CCD and the CPS are likely attributable to biases in the CPS. Some observers have raised concerns about CCD counts of high school completers; at issue is whether diplomas issued by community colleges or adult education programs are included among the CCD counts of regular diplomas or alternative credentials (Mishel and Roy 2006). It is not clear how substantial these biases might be.

One obvious limitation of CCD data --- and indeed all state administrative data --- is that it pertains exclusively to public school students. When high school dropout and completion rates are used for the purposes of describing levels of human capital in a population or for describing young people's success at navigating the secondary education system, this limitation is important. In the fall of 2003, 8.3 percent of secondary school students were enrolled in private schools (Snyder et al. 2008: Table 52).

Beyond these cross-sectional, state-produced enrollment counts, states also frequently make use of longitudinal administrative data to produce high school dropout and completion rates. Each state utilizes somewhat different data collecting, reporting, and aggregation procedures, but in general there have been few concerns about states' reports of the numbers of students in each grade or of the numbers of students obtaining regular diplomas. The most prominent controversies surrounding the use of longitudinal administrative data concern the *use* (or *misuse*) of those data. While there is little concern about states' abilities to accurately count the number of students who begin high school as 9th graders (for example), there is frequent concern about how states account for factors like migration, incarceration, expulsion, and enrollment in alternative educational programs as they manipulate their 9th grade enrollment counts to cause them to reflect the number of students at risk of obtaining high school diplomas some time later. For example, in 2004 and 2005 the Houston public schools were

sharply criticized for reclassifying nearly all dropouts as either enrollees in alternative educational programs or out-migrants --- thus removing them from the denominators of dropout rate.

The Denominators: Differences in Estimates of the Size of the At-Risk Population

The denominators of high school dropout and completion rates represent the number of students at risk of having dropped out or completed high school at a particular point in time. This section describes differences between cross-sectional sample surveys, longitudinal surveys, and administrative data with respect to their estimates of the size of the at risk population.

According to NCES reports, the NELS cohort of 8th graders in spring of 1988 generalizes to a population of approximately 2.46 million public school 8th graders during the 1987-88 academic year (Hafner et al. 1990; Kaufman et al. 1992). The 1987 CCD State Nonfiscal Survey indicates that there were about 2.84 million public school 8th graders in the fall of 1987; there were about 2.85 million public school 8th graders in that year according to the 1987 October CPS. The relatively low numbers of NELS students in 1988 may be due to the base year exclusions of sampled students with English language problems or physical or mental disabilities. Without these exclusions, the NELS base year sample would have generalized to a population of about 2.60 million public school 8th graders (Spencer et al. 1990). Given the vast differences between these sources of data with respect to how this information was gathered, it may be surprising that their level of agreement is so high.

A similar story holds among 10th graders using data that are more recent. In fall 2002, there were 3.65 million public school 10th graders according to the October CPS and 3.58 million public school 10th graders according to the CCD. In the ELS cohort, there were roughly 3.15 million public school 10th graders --- again, this NCES survey excluded young people with language problems English language problems or physical or mental disabilities.

Although the information above suggests that the CPS and CCD data generally agree about the number of enrolled public school students, there are also important discrepancies between these sources of data. As reported by Warren and Halpern-Manners (2009), the estimated numbers of 10th, 11, and 12th graders is higher in the CPS than in the CCD ... and this discrepancy has grown over time. As those authors argue, the best explanation for this pattern is that some young people who are claimed by CPS respondents to be enrolled in 10th, 11th, and particularly 12th grade are in fact not officially enrolled in high school at all. This may come about through CPS respondents' unwillingness to report that their high school-aged children have left high school (a social desirability issue), or it may arise because of honest confusion about young people's official enrollment status.

In larger states, it is possible to compare enrollment counts from the CPS, the CCD, and states' own internal administrative data. In general, states' own enrollment data and the CCD are virtually identical in most years. This should come as no surprise --- both are derived from states' own accounting of enrollments, aggregated up from the schoolhouse level. In contrast, because of sample size issues the CPS data are much more variable, sometimes yielding higher counts than states' administrative data or the CCD and sometimes yielding lower counts. These CPS sample size issues are one reason that status dropout and completion rates generally include a wide range of ages (i.e., 16 to 24 year olds) as opposed to focusing more narrowly on specific cohorts of students.

The Denominators: Differences in the Estimated Number of First Time 9th Graders

Cohort measures typically characterize rate dropout or completion among a group of young people who began high school at the same time. Whereas individual-level cohort rates typically use *actual* counts of the number of students at risk of dropping or completing in the denominator, aggregate cohort rates --- based centrally on cross-sectional data on enrollments and completions --- must *approximate* the number of students at risk. CCD-based measures, for example, typically begin with the number of 9th graders in a cohort in the denominator, and then adjust that number to account for various factors that lead the number of enrolled 9th graders at the beginning of one academic year to differ from the true number of students at risk of graduating or dropping out three academic years later.

Many observers have noted that aggregate cohort rates based on CCD (or similar) data are potentially biased by migration, grade retention, mortality, enrollment in ungraded programs, and other factors that lead to biases in the estimated number of students at risk of completing or dropping out (e.g., Pallas 1990). The most widely used CCD-based synthetic cohort rates have thus all sought to adjust the denominator (which begins with the total number of 9th graders enrolled in an academic year) to account for migration, grade retention, and (indirectly) mortality.

Unfortunately, efforts to adjust the denominators of CCD-based aggregate cohort rates to better approximate true cohort rates have largely been unsuccessful (Warren 2005; Warren and Halpern-Manners 2009). Because of problems in their denominators, widely publicized CCD-based aggregate cohort rates like Swanson's Cumulative Promotion Index (Editorial Projects in Education 2008; Swanson and Chaplin 2003); the NCES Averaged Freshman Graduation Rate (Seastrom et al. 2006a; Seastrom et al. 2006b); Greene's Adjusted Completion Rate (Greene and Forster 2003; Greene and Winters 2006); and Warren's Estimated Completion Rate (Warren 2005) are biased. These are systematic errors, not random errors. These measures are accurate only under very specific (and generally unlikely) demographic circumstances. The direction and magnitude of their biases vary over time and across states in ways that further weaken their utility for most purposes.

The Numerators: Estimating the Numbers of Completers and Dropouts

The preceding sections described issues in quantifying the number of students "at risk" of graduating from or dropping out of high school in the denominators of status, event, cohort, and aggregate cohort dropout and completion rates. This section describes differences between estimates based on cross-sectional sample surveys, longitudinal surveys, and administrative data with respect to their numerators, or their counts of the number of people who complete or drop out of high school.

Cross-sectional sample surveys --- most notably the CPS, the decennial Census, and the ACS --- allow respondents to classify themselves as either high school completers or non-completers. The status "high school dropout" is an inferred residual category, assigned to respondents who have not obtained a secondary school credential and who are not enrolled in school. In recent years, these sample surveys have included additional questions that seek to differentiate high school graduates from those who obtain GEDs. A number of observers have raised questions about the validity of these self-reports of *how* respondents complete high school (Chaplin 2002; Warren and Halpern-Manners 2007). What is more, no effort is made in these cross-sectional sample surveys to differentiate individuals who obtained regular high school diplomas from those who earned certificates of completion or who completed high school by completing community college or adult education programs. In the end, cross-sectional sample surveys do a relatively poor job of representing the number of regular diploma

recipients. Finally, these sample surveys include no information about when respondents completed high school, and so it is not possible to measure on-time high school completion.

Longitudinal sample surveys --- like NELS, ELS, and their predecessors --- typically supplement self-reports of high school completion and dropout with transcript data on those events. It is thus possible to construct status, event, and cohort dropout or completion rates. In combination with information about the timing of dropout and completion events, it is also possible to differentiate on-time completion from later completion. Although the transcript data frequently available in these longitudinal sample surveys includes detail about when respondents completed high school and what sort of public school credential they obtained, there is less (and less well-validated) information about credentials obtained from GED, community college, and adult education programs. An exception is the ELS data, which includes detailed administrative data from the GED Testing Service.

States' longitudinal administrative data include information about enrollment in and exits from public schools. Most students exit the school system by obtaining a regular diploma or alternative credential, and states' longitudinal data systems generally count these things in a consistent manner. Where they differ, though, is in their accounting of other exits from the system. Districts and states vary with respect to how and how well they track and account for students who transfer to schools in other districts or states; who transfer to private schools, correctional facilities, or other institutional settings; or who are expelled, who die, or who exit the system in other ways. This means that states' longitudinal data generally preclude enumerating students who obtain credentials from GED programs or private schools. What is more, exiting the system by dropping out is rarely directly observed. Students are presumed to have dropped out if they do not exit the system by completing high school, transferring, or through some other quantifiable transition. It is difficult to validate such indirect techniques for counting dropouts.

Cross-sectional state administrative data --- for example, as compiled in the CCD --- generally include only annual counts of the numbers of regular high school diplomas, GEDs, and other credentials issued by schools, districts, and states. As such, it is impossible to distinguish on-time completers from other completers. Some observers have raised concerns about whether alternative credentials are consistently distinguished from regular diplomas in the CCD (Warren 2005), and GED data are frequently missing in the CCD. As with longitudinal administrative data, dropout is a residual outcome, and is not directly observed in cross-sectional state administrative data. Because the CCD and similar data are usually compiled from data produced by public school agencies, very little information is available about private high school completions. Consequently, analysts sometimes supplement CCD data on public school completions with similar data on private schools from the bi-annual NCES Private School Survey (Tourkin et al. 2008).

To summarize, the type of data used to construct dropout or completion rates determines what can be counted in the numerator. Status dropout and completion rates are typically based on cross-sectional sample survey data, and thus they do less well at differentiating regular diploma recipients from other high school completers and they do not distinguish on-time completers from later completers. Event dropout and completion rates can be constructed from cross-sectional survey data, longitudinal survey data, or longitudinal administrative data. Depending on which type of data is used, event rates may or may not distinguish between types of completions; may or may not include private school completers; may or may not directly enumerate dropouts; and may or may not differentiate on-time completers from other completers. Cohort dropout and completion rates can be constructed from longitudinal

sample survey data or longitudinal administrative data, and this choice determines analysts' ability to observe dropout events and to include private school completers.

Discussion

There are three important reasons to measure high school dropout and completion rates: To assess the level of human capital in a population, to characterize schools' "holding power," and to describe students' success at navigating the secondary school system. For each purpose, how can we move toward a set of "common metrics?" What are the best available techniques for measuring rates of high school dropout and completion, given each purpose? What are the weaknesses in those "best techniques," and how might they be improved?

Describing Levels of Human Capital

Status dropout or completion rates computed from cross-sectional sample surveys are best suited to describing levels of human capital in a population. Because the goal is to describe the share of all individuals who have obtained a high school credential, it is important to utilize data that include people who may have obtained those credentials from public schools, private schools, GED programs, community colleges, adult education programs, or elsewhere. This is not possible with state administrative data, because it only pertains to people who went to school in that state (as opposed to the target population, which includes everyone in an age range who lives in that state) and because those data generally do not include those who complete high school outside of the public school system.

Status completion and dropout rates have traditionally been based on CPS or decennial Census data. The CPS has known flaws for this purpose --- it excludes the non-civilian and institutionalized populations and includes too few respondents to generalize reliably to the state or substance levels --- and the long-form of the Census has been phased out. Since the ACS sample is about 16 times larger than the CPS, and since it generalizes to the entire population of US households, status completion rates should more frequently be constructed using these data.

For the purposes of describing levels of human capital, status completion or dropout rates are imperfect in a couple of respects. One key weakness is their treatment of all high school credentials as equivalent. Economists and others have long questioned the labor market value of the GED (e.g., Cameron and Heckman 1993); little is known about the labor market value of other alternative credentials. The 2008 ACS was the first to distinguish regular diploma recipients from GED recipients. It would be useful for the ACS to add a third category in order to distinguish regular diploma recipients from those who obtain certificates of completion or who complete high school via community college or adult education programs. As described below, there is also potential payoff to adding just a few questions about the year and state/country in which individuals completed high school. A second key weakness of status completion rates for describing levels of human capital in a population concerns the validity and reliability of respondents' reports of whether (and how) they completed high school. This issue could be better understood, and perhaps put to rest, with a validation study in which high school and GED records are matched to a random subset of ACS records. This would require obtaining information from the selected respondents about where, when, and how they obtained their high school credentials.

Describing Schools' Holding Power

For purposes of describing schools' holding power, dropout and completion rates must directly and reliably pertain to specific locales and to specific cohorts of students, and they must be available annually. This implies the use of cohort rates based on longitudinal administrative data. As described above, the biggest drawback to using cohort rates based on these data is that different states and locales have historically defined the numerators and denominators differently. In the numerator, most states have only counted regular diploma recipients as "successes," but some have been less restrictive in their definition of "success." In the denominator, states have typically begun with counts of first time 9th graders, but they have differed in how they adjust that figure to account for migration and other exits and entries in order to have the denominator reflect the number of students at risk of graduating or dropping out. As a result, the public has sometimes been distrustful of these rates, trend analyses have been hampered, and cross-state comparisons have been next to impossible. In this respect, the movement toward universal adoption of the standards laid out in the National Governors Association Compact (2008) is a great step forward. If states faithfully and consistently implement the standards laid out in this agreement --- which call for restricting the numerator of cohort rates to regular diploma recipients and which spell out standards for deriving the denominator --- then we will eventually reach a point at which these cohort rates can be compared over time and across states.

In the meantime, however, many situations call for the use of aggregate cohort completion rates based on cross-sectional administrative data like the CCD. This includes any situation in which the goal is to describe trends over time and/or differences across states in rates of high school dropout or completion. Policy makers or the members of the public may wish to consider changes over several years in their state or district's dropout rate. Or they may wish to compare dropout rates in their state or district to those in neighboring states or districts. Researchers may want to model state- or district-level dropout or completion rates in order to understand the factors that influence schools' holding power or to understand the consequences of schools' holding power for any number of social, economic, or other outcomes. All of these situations require measures of dropout or completion that are consistent over time and/or place. Until consistently define cohort rates based on longitudinal administrative data have been around for a number of years, aggregate cohort rates based on CCD or similar data will have to suffice. Unfortunately, data limitations preclude the construction of aggregate cohort dropout rates.

Since there will likely be important roles for aggregate cohort measures based on CCD or similar data for foreseeable future, it is important to routinely acknowledge weaknesses in those measures. As described above, the most widely publicized variants of these measures are demonstrably biased. Warren's (2005) ECR is currently the least biased of these measures, and should be used whenever synthetic cohort completion rates are computed at the national or state levels. There is currently no scientifically sound CCD-based aggregate cohort measure than can be constructed for use at the level of cities or districts. It is conceivable that one could adjust for migration using different data or different algebra than Warren (2005), and that this would allow for these applications of the ECR.

Describing Students' Progress through the Secondary School System

As described above, measures of *students'* progress through the system of secondary education are conceptually and technically distinct from measures of *schools'* holding power. Measures that describe students should be based on data that include private school students and that give analysts control over how to define "success" and "failure" in the numerator. They should also allow for an understanding of how dropout and completion vary by students' social, demographic, and other

characteristics. For these reasons, longitudinal administrative data from schools are not appropriate. Neither are cross-sectional administrative or sample survey data (because they do not allow analysts to follow students over the course of high school).

For these reasons, individual-level cohort rates based on longitudinal sample survey data (like NELS or ELS) are best suited for describing students' progress through the secondary school system. From a practical standpoint, however, the reliance on such surveys is limiting. They are relatively expensive and infrequently conducted. They suffer from problems of coverage bias and sample attrition (as described above). They cannot be used to describe the experiences of students in particular states, cities, or districts. So while it is certainly worth investing in longitudinal sample surveys (for these and many other purposes), it is worth considering some alternatives.

For example, it would be relatively easy to construct an individual-level cohort measure of students' success at completing their secondary schooling by adding just one item to the ACS: "In what year did you first enroll in the 9th grade?" In the absence of any other information, this would allow for the construction of an event completion rate (which is essentially the same as an individual-level cohort rate when the time lapse is the standard length of high school). This cohort measure could be made to apply to particular states --- and with much greater statistical power --- by adding another item that asks about the state in which individuals first enrolled in the 9th grade, and it could become an on-time individual-level cohort completion rate by adding a question about the year in which individuals obtained their high school credential. To be confident of the validity of these rates, however, it would first be wise to validate reports of whether, when, and how individuals completed high school against high school and GED Testing Service records (as described above).

Conclusion

A large share of the confusion about how often young people drop out of or complete high school stems from mismatches between people's reasons for quantifying these rates and the specific measures and data that they employ. There are many reasons to describe schools' holding power --- especially in this age of school accountability --- but CPS-based status dropout rates and individual-level cohort dropout rates derived from longitudinal NCES surveys are very dull instruments for achieving this end. There are probably as many good reasons to describe students' success at progressing through America's system of secondary schools, but this end is not achievable with CCD-based aggregate cohort completion rates or even cohort rates based on states' longitudinal data systems. Likewise, there are good reasons to describe the human capital in a population, but it is important to recognize that none of the measures developed in the last several decades take us much further toward that goal than the status dropout rates that were in use in the 1970s. There are ways to better achieve all three of these goals, but the biggest step we can take toward clarifying our understanding of the high school dropout/completion situation in the United States would be to be consistently clear and forthcoming about *why* we are measuring high school dropout/completion, what measures and data are best suited for this purpose, and what limitations are inherent in those data and measures.

References

- Annie E. Casey Foundation. 2008. *Kids Count 2008 Data Book*. Baltimore, MD: Annie E. Casey Foundation.
- Cameron, Stephen V. and James J. Heckman. 1993. "The Nonequivalence of High School Equivalents." *Journal of Labor Economics* 11:1-47.
- Chaplin, Duncan. 2002. "Tassels on the Cheap." *Education Next* 2:24-29.
- Curtin, Thomas R., Steven J. Ingels, Shiyong Wu, and Ruth Heuer. 2002. *User's Manual: NELS:88 Base-Year to Fourth Follow-up: Student Component Data File*. NCES 2002323. Washington, DC: National Center for Education Statistics, US Department of Education.
- Editorial Projects in Education. 2008. *Diplomas Count 2008: School to College: Can State P-16 Councils Ease the Transition?*, Vol. 27 (special issue), no. 40, June 5, 2008.
- Greene, Jay P. and Greg Forster. 2003. *Public High School Graduation and College Readiness Rates in the United States*. New York: Center for Civic Innovation, Manhattan Institute.
- Greene, Jay P. and Marcus A. Winters. 2002. *High School Graduation Rates in the United States*. New York: Center for Civic Innovation, Manhattan Institute.
- . 2006. *Leaving Boys Behind: Public High School Graduation Rates*. New York: Center for Civic Innovation, Manhattan Institute.
- Greene, Jay P., Marcus A. Winters, and Christopher B. Swanson. 2006. "Missing the Mark on Graduation Rates: A Response to 'The Exaggerated Dropout Crisis'." Pp. 39, 42 in *Education Week*.
- Hafner, Anne, Steven Ingels, Barbara Schneider, and David Stevenson. 1990. *A Profile of the American Eighth Grader: NELS:88 Student Descriptive Summary*. NCES 90458. Washington, DC: National Center for Education Statistics, US Department of Education.
- Hartzell, Gary, Jack McKay, and Jack Frymier. 1992. "Calculating Dropout Rates Locally and Nationally with the Holding Power Index." ERIC Document Reproduction Service No. ED 343 953.
- Hauser, Robert M. 1997. "Indicators of High School Completion and Dropout." Pp. 152-184 in *Indicators of Children's Well-Being*, edited by R. M. Hauser, B. V. Brown, and W. R. Prosser. New York: Russell Sage Foundation.
- Heckman, James J. and Paul A. LaFontaine. 2007. "The American High School Graduation Rate: Trends and Levels." Retrieved from <http://ftp.iza.org/dp3216.pdf> on April 21, 2008. Discussion Paper No. 3216, December 2007. Bonn, Germany: Institute for the Study of Labor.
- Ingels, Steven J., Kathryn L. Dowd, John R. Taylor, Virginia H. Bartot, Martin R. Frankel, Paul A. Pulliam, and Peggy Quinn. 1995. *User's Manual: NELS:88 Second Follow-Up: Transcript Component Data File*. NCES 95377. Washington, DC: National Center for Education Statistics, US Department of Education.
- Kaufman, Phillip. 2001. "The National Dropout Data Collection System: Assessing Consistency." Paper presented at the Harvard Civil Rights Project, January 13, 2001. Available on the web at: <http://www.law.harvard.edu/civilrights/publications/dropout/kaufman.html>.
- . 2004. "The National Dropout Data Collection System: History and Search for Consistency." Pp. 107-130 in *Dropouts in America: Confronting the Graduation Rate Crisis*, edited by G. Orfield. Cambridge, MA: Harvard Education Press.
- Kaufman, Phillip, Marilyn M. McMillen, and Denise Bradby. 1992. *Dropout Rates in the United States: 1991*. NCES 92129. Washington, DC: National Center for Education Statistics, US Department of Education.
- Komiski, Robert. 1990. "Estimating the National High School Dropout Rate." *Demography* 27:303-311.
- Laird, Jennifer, Emily Forrest Cataldi, Angelina KewalRamani, and Chris Chapman. 2008. *Dropout and Completion Rates in the United States: 2006*. NCES 2008053. Washington, DC: US Department of Education, National Center for Education Statistics.

- McMillion, Robin, Jason Ramirez, and Linda Roska. 2008. *Secondary School Completion and Dropouts in Texas Public Schools, 2006-07*. Accessed at http://www.tea.state.tx.us/research/pdfs/dropcomp_2006-07.pdf on 9/23/2008. Austin, TX: Department of Assessment, Accountability, and Data Quality, Division of Accountability Research, Texas Education Administration.
- Mishel, Lawrence. 2006. "The Exaggerated Dropout Crisis." Pp. 40 in *Education Week*.
- Mishel, Lawrence and Joydeep Roy. 2006. *Rethinking High School Graduation Rates and Trends*. Washington, D.C.: Economic Policy Institute.
- National Governors Association. 2008. *Implementing Graduation Counts: State Progress to Date, 2008*. Washington, D.C.: National Governors Association Center for Best Practices.
- North Carolina State Board of Education. 2004. *North Carolina Public Schools: Statistical Profile 2004*. Raleigh, NC: North Carolina State Board of Education.
- Pallas, Aaron M. 1990. "Conceptual and Measurement Issues in the Study of School Dropouts." in *Research in Sociology of Education and Socialization*, vol. 8, edited by R. G. Corwin and K. Namboodiri. Stamford, CT: JAI Press.
- Ramirez, Jason, Robin McMillion, and Linda Roska. 2006. *Secondary School Completion and Dropouts in Texas Public Schools, 2004-05*. Accessed at http://www.tea.state.tx.us/research/pdfs/dropcomp_2004-05.pdf on 9/18/2008. Austin, TX: Department of Assessment, Accountability, and Data Quality, Division of Accountability Research, Texas Education Administration.
- Seastrom, Marilyn, Chris Chapman, Robert Stillwell, Daniel McGrath, Pia Peltola, Rachel Dinkes, and Zeyu Xu. 2006a. *A Review and Analysis of Alternative High School Graduation Rates, Volume 1. Users Guide to Computing High School Graduation Rates, Volume 2, An Analysis of Alternative High School Graduation rates*. NCES 2006–604 and 605. Washington, D.C.: National Center for Education Statistics, U.S. Department of Education.
- Seastrom, Marilyn, Lee Hoffman, Chris Chapman, and Robert Stillwell. 2006b. *The Averaged Freshman Graduation Rate for Public High Schools From the Common Core of Data: School Years 2002–03 and 2003-04*. NCES 2006-606rev. Washington, D.C.: National Center for Education Statistics, U.S. Department of Education.
- Snyder, Thomas D., Sally A. Dillow, and Charlene M. Hoffman. 2008. *Digest of Education Statistics, 2007*. NCES #2008-022. Washington, D.C.: US Government Printing Office.
- Spencer, Bruce D., Martin R. Frankel, Steven J. Ingels, Kenneth A. Rasinski, and Roger Tourangeau. 1990. *NELS:88 Base Year Sample Design Report*. NCES 90463. Washington, DC: National Center for Education Statistics, US Department of Education.
- Swanson, Christopher B. and Duncan Chaplin. 2003. *Counting High School Graduates When Graduates Count*. Washington, D.C.: Education Policy Center, Urban Institute.
- Tourkin, Steven, Nancy Swaim, Xiaodong Guan, Jennifer Peterson, Dennis Schwanz, Toni Warner, Randall Parmer, Meagan Wilson, Stephen Broughman, and Elizabeth Walter. 2008. *2003-2004 Private School Universe Survey (PSS) Data File User's Manual and Survey Documentation*. NCES 2008314. Washington, DC: National Center for Education Statistics, US Department of Education.
- U.S. Department of Education. 2006. *Digest of Education Statistics, 2005*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.
- Warren, John Robert. 2005. "State-Level High School Completion Rates: Concepts, Measures, and Trends." *Education Policy Analysis Archives* (Retrieved March 13, 2006 from <http://epaa.asu.edu/epaa/v13n51/>). 13.

- Warren, John Robert and Andrew Halpern-Manners. 2007. "Is the Glass Emptying or Filling Up? Reconciling Divergent Trends in High School Completion and Dropout." *Educational Researcher* 36:335-343.
- . 2009. "Measuring High School Graduation Rates at the State Level: What Difference Does Methodology Make?" *Sociological Methods & Research* 38:3-37.
- . Forthcoming. "Measuring High School Graduation Rates at the State Level: What Difference Does Methodology Make?" *Sociological Methods & Research*.
- Winglee, Marianne, David Marker, Allison Henderson, Beth Aronstamm Young, and Lee Hoffman. 2000. *A Recommended Approach to Providing High School Dropout and Completion Rates at the State Level*. NCES2000305. Washington, DC: National Center for Education Statistics, US Department of Education.
- Young, Beth and Lee Hoffman. 2002. *Public High School Dropouts and Completers From the Common Core of Data: School Years 1991-92 through 1997-98*. NCES 2002382. Washington, DC: National Center for Education Statistics, US Department of Education.