WORKING
PAPER

# Estimating Graduation and Dropout Rates with Longitudinal Data

## A Case Study in the Pittsburgh Public Schools

JOHN ENGBERG AND BRIAN GILL

WR-372-PPS

July 2006

Prepared for the Pittsburgh Public Schools




EDUCATION

## Introduction

Conceptual ambiguities and imperfect data conspire to make graduation and dropout rates difficult to calculate. In this paper, we review the challenges of calculating graduation rates, discuss how these challenges have been approached when using cross-sectional data, and describe a method that analyzes longitudinal, student-level data to provide improved estimates of graduation and dropout rates. We then apply that method to estimate graduation and dropout rates for the Pittsburgh Public Schools district-wide and for each high school in the district. We expect that this report will hold interest for readers concerned with the national debate over the calculation of graduation rates and also for readers involved with the Pittsburgh Public Schools.[1]

## Challenges in Defining Graduation and Dropout Rates

Although graduation is itself an unambiguous event, defining and measuring the graduation rate for a particular school, district, or state is not as straightforward as it might seem. The first challenge is identifying the pool of students to be used as the denominator in measuring the graduation rate. A graduation rate based on all students who enter ninth grade four years earlier will be much smaller than a graduation rate based on all students who enter twelfth grade the previous fall. Because many students drop out prior to twelfth grade, using entering twelfth graders as a base for the graduation rate calculation means that the rate will not be useful for measuring the success of retaining and educating students throughout their high school years. On the other hand, using all entering ninth graders as the base ignores the fact that many students move from the

[1] This study was sponsored by the Pittsburgh Public Schools, and conducted under the auspices of RAND Education.

district; whether they graduate elsewhere is usually not known in the district they left. Meanwhile, other students are moving into the district after ninth grade, producing graduates who were not part of the original entering cohort.

Moreover, some students take more than four years to graduate. Should the district be penalized for requiring a student to repeat a grade in high school? Calculating a four-year graduation rate implicitly does so, whereas a five-year (or six-year) graduation rate will include students who are either retained or who leave school temporarily. This distinction is very important when using longitudinal individual data to calculate graduation rates, whereas graduation rates based on aggregate data implicitly include all graduates—no matter how many years they take to complete high school.

Grade repetition creates ambiguities not only in the count of graduates but also in the denominator. In many high schools, substantial numbers of students repeat ninth grade. If all ninth graders were included in the denominator for determining graduation rates, the repeating students would be double counted. They would be included in the base in both their first *and second* time through ninth grade. It is most appropriate to use the number of students who are entering ninth grade for the first time as the base denominator, *but data sets often fail to distinguish first-time ninth-graders from repeaters.*

The identification of dropouts creates an additional complication. In Pittsburgh, for example, district information systems record whether a student left school to enroll in a school outside the district (public or private), join the military, work, or because of death

or incarceration—in addition to a catch-all category called "quit school." But the patterns in the data suggest that many of those who were coded in the data as moving to another school are actually dropouts. For example, some high schools have reported rates of transfer out of the district that are much higher in grades eleven and twelve than those in grades nine and ten. Although it is possible that more families move when their children are older, a more plausible explanation for this pattern is that some students who drop out are incorrectly coded as having transferred to other schools.

**Attempts to Resolve These Issues With Cross-Sectional Data**

In the absence of comprehensive longitudinal student data systems, analysts have attempted to use cross sections of student data, augmented with U.S. Census Bureau or other administrative data, to calculate graduation and dropout rates that account for the issues described above. Recent examples include Greene and Winters (2006); who calculate graduation rates for all states and for the 100 largest districts; *Education Week* (2006), which likewise calculates rates for states and large districts; and Warren (2005), who reviews methods and provides several alternative calculations for all states.[2]

These cross-sectional methods begin with the Common Core Data available through the National Center for Education Statistics.[3] The Greene and Warren methods aim to estimate the proportion of high-school students who ever graduate (regardless of how

[2] *Leaving Boys Behind: Public High School Graduation Rates.* Jay Greene and Marcus Winters. Center for Civic Innovation. Civic Report No. 48. April 2006. (http://www.manhattan-institute.org/pdf/cr_48.pdf). *Education Week*, "Diplomas Count: An Essential Guide to Graduation Policy and Rates) (Washington, DC: Editorial Projects in Education, 2006).
"State-Level High School Completion Rates: Concepts, Measures and Trend," John Robert Warren. *Education Policy Analysis Archives*, Volume 13. 2005. (http://epaa.asu.edu/epaa/v13n51/).

[3] Available online at http://nces.ed.gov/ccd/.

long it takes them) by calculating the ratio of the number of diplomas granted in one year to an estimate of the number of first-time ninth-grade entrants four years earlier.

Analysts have used a variety of methods to adjust for the number of students who repeat ninth grade. One method is to use the number of eighth-graders in the previous year as an estimate of first-time ninth-graders, because retention rates among eighth graders generally are lower. This approach, however, misses students who enter the public school system following the eighth grade. Therefore, other analysts prefer to use an average of reported eighth-, ninth-, and tenth-grade enrollment from three successive years as an estimate of the number of first-time ninth-graders.

Although these methods commonly use a four-year interval between the number of ninth-graders and the number of graduates employed in the calculation, they should not be interpreted as four-year graduation rates. The data sources do not distinguish graduates who spent four years in high school from those who took more or less time. The count of graduates will omit students from the chosen ninth-grade cohort who took shorter or longer than four years, but it will include graduates from earlier cohorts who took longer than four years and later cohorts who took less than four years. The calculated rate will be approximately equal to the true graduation rate for all entering students—provided that net migration into and out of the district is zero.

In order to adjust for population growth or decline, analysts have used U.S. Census Bureau or state data center estimates of changes in the population of high-school age

children during the span of years covered by the high school years in question. This adjustment is especially important in regions that are experiencing dramatic growth or decline in school enrollment. However, it is possible to make these adjustments only for large administrative units for which these estimates are available, such as states or counties. Moreover, these adjustments are based on net migration—the difference between the number of students transferring in and the number transferring out. The method ignores the fact that some of the graduates are not the same students that entered ninth grade—it implicitly assumes that students transferring in during high-school years are interchangeable with students transferring out.

The recent *Education Week* report, in contrast to the Greene and Warren methods, attempts to estimate the proportion of high-school students who graduate *on time* (in four years). *Education Week* uses data from just two consecutive years, creating an artificial, synthetic cohort made up of students in successive grades.[4] Dropouts are accounted for by examining the ratios of the numbers of tenth, eleventh, and twelfth graders and graduates in a given year to the numbers of ninth, tenth, eleventh, and twelfth graders the previous year. A lower ratio is assumed to correspond to a higher dropout rate. Multiplying these ratios together yields an estimate of the on-time graduation rate for ninth-graders.

The Pennsylvania Department of Education (PDE) takes a different approach to calculating official graduation rates and dropout rates, using a series of aggregate cross-

[4] *Education Week*, "Diplomas Count: An Essential Guide to Graduation Policy and Rates" (Washington, DC: Editorial Projects in Education, 2006).

sectional data elements to create a quasi-longitudinal measure that follows an artificial cohort of students. Specifically, the graduation rate for a high school is defined as the total number of graduates in a year divided by this number of graduates, plus the sum of twelfth -grade dropouts the same year, eleventh grade dropouts the preceding year, and so on back to ninth grade. This method is not truly longitudinal, however, because it does not account for grade repetition, and because it includes among the count of graduates students who transferred into the school after ninth grade.

PDE defines a dropout as any "student who, for any reason other than death, leaves school before graduation without transferring to another school/institution."[5] PDE defines the dropout rate for a school as the number of dropouts during a 12-month period divided by the fall enrollment of the school. For a school district, the dropout rate is defined as the number of dropouts during a 12-month period divided by the district's total enrollment in grades seven through twelve. This rate differs conceptually from those presented in this paper in that it is based on total enrollment during a particular year rather than on a single cohort followed through high school. District-wide graduation rates reported by PDE appear low partly because they include middle-school grades, for which dropout rates (as officially reported) are near zero. As noted above, however, PDE's graduation-rate calculation departs from its dropout rate calculation, attempting to use only the dropouts from a particular cohort.

Many of the challenges associated with trying to measure graduation rates with aggregate cross-sectional data can be accommodated in a more natural way by using longitudinal,

[5] Available online at http://www.pde.state.pa.us/k12statistics/cwp/view.asp?a=3&q=112801.

student-level data. In the next section, we examine how the availability of longitudinal data can lead to new methods of estimating graduation and dropout rates that explicitly address many of the naturally occurring flows in and out of schools.

**Preferred Definition of Graduation and Dropout Rates**

We present not only our preferred definition of *graduation rate*, but also a variety of other definitions, allowing the reader to understand how the rates would change if different time points, definitions of dropping out, or methods for addressing students who move were used to estimate the rate.

Our preferred definition is derived from the final record for each student, up to five years after entering ninth grade. Based on the last observed record in the Pittsburgh Public Schools, we preliminarily categorize each student as a graduate, a dropout, a mover, or as still enrolled. Table 3 (in Appendix A) provides the codes for each of these categories. Approximately 20 percent of students are coded as movers, many of whom are likely to be dropouts, either because they are initially miscoded or because they drop out after moving. We therefore calculate the probability that each of these movers is either a graduate, a dropout, or still enrolled (outside of the Pittsburgh Public Schools) based on the observed patterns for demographically similar students who remain in the district.

This probability is based on a statistical model that estimates the relationship between the observed outcome for students who do not leave the district and their sex, race, age upon entry to ninth grade, and whether or not they changed schools within the district. Each of

these characteristics is significantly associated with the probability of graduating, dropping out, or still being enrolled after five years. Boys are more likely to drop out than girls, and black students are more likely to drop out than other students. Students who are older when they enter ninth grade and students who change schools are much less likely to graduate than younger, less transient students. More details regarding this method are included in Appendix B.

We apply this estimated relationship to each of the students who moves out of the Pittsburgh Public Schools. Based on his or her sex, race, age upon entry into ninth grade, and the implied change of schools when moving out of the district, we calculate the probability that each of these movers will be a graduate, a dropout, or still enrolled after five years. Our preferred graduation and dropout rate calculations are based on the observed status of students who remained in the district as well as the predicted status of students coded as leaving the district.

This method relies on several assumptions that cannot be tested with the information available to us. We implicitly assume that after conditioning on the demographics of the student (gender, race, age at entry to ninth grade), the outcomes for students who leave the district are similar to those who stay. Furthermore, we assume that changing schools within the district has a similar relationship to high school outcomes as that of changing schools when leaving the district. However, the method does not rely on an assumption about causality—it could be that changing schools merely reflects other disruptions in a student's life rather than actually causing a student to be less likely to graduate.

**Graduation and Dropout Rates in the Pittsburgh Public Schools**

Table 1a presents estimated five-year graduation rates, and Table 1b presents dropout rates in the Pittsburgh Public Schools. The rates are presented separately for each school, and also by sex and race. All calculations are based on two cohorts of students—those who first attended ninth grade in 1999–2000 and those who first attended in 2000–2001. Because there is no way to know definitively what happened to students who left district schools, these are, necessarily, *estimates* of true graduation rates, incorporating actual graduation and dropout rates for students recorded as such, as well as predicted graduation and dropout rates for students who leave the district.

District-wide, the estimated five-year graduation rate is 64 percent and the estimated dropout rate is 35 percent. Less than 2 percent of entering ninth graders are still enrolled in high school six years later. We favor a five-year rate over a four-year or six-year rate because many students return for a fifth year, whereas the overwhelming majority of students have graduated or dropped out within five years. Pittsburgh has racial and gender differences in graduation rates, with African-Americans and males experiencing lower rates of graduation. (The student population in the Pittsburgh Public Schools is almost entirely African-American or white. For demographic purposes, PPS often categorizes students into two groups: "African-American" and "White and Other".) The estimated graduation rate for African-American students is 58 percent, versus 70 percent for other (primarily white) racial groups. Females graduate at an estimated five-year rate of 69 percent, versus 59 percent for males. The lowest graduation rates are for African-

American males, nearly half of whom drop out. These racial and gender differences are consistent with estimated graduation patterns nationwide.[6]

**Table 1**
**Estimated Five-Year Graduation and Dropout Rates in the Pittsburgh Public Schools, Using Predicted Outcomes for Students Who Leave the District**

1a: Graduation Rates

| | White and Other | | | African American | | | Grand Total |
|---|---|---|---|---|---|---|---|
| School | Females | Males | Both | Females | Males | Both | |
| Allderdice High School | | | 83% | | | 60% | 75% |
| Brashear High School | | | 69% | | | 50% | 63% |
| Carrick High School | | | 71% | | | 43% | 63% |
| Langley High School | | | 59% | | | 44% | 53% |
| Oliver High School | | | 50% | | | 54% | 53% |
| Peabody High School | | | 47% | | | 53% | 52% |
| Perry Traditional Academy | | | 81% | | | 79% | 80% |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 85% | | | 85% | 85% |
| Schenley High School | | | 70% | | | 68% | 69% |
| South Vo Tech High School | | | 53% | | | 59% | 56% |
| Westinghouse High School | | | * | | | 53% | 53% |
| Pittsburgh District-wide | 74% | 66% | 70% | 64% | 51% | 58% | 64% |

1b: Dropout Rates

| | White and Other | | | African American | | | Grand Total |
|---|---|---|---|---|---|---|---|
| School | Females | Males | Both | Females | Males | Both | |
| Allderdice High School | | | 17% | | | 38% | 23% |
| Brashear High School | | | 30% | | | 46% | 35% |
| Carrick High School | | | 28% | | | 55% | 36% |
| Langley High School | | | 39% | | | 46% | 42% |
| Oliver High School | | | 49% | | | 44% | 46% |
| Peabody High School | | | 52% | | | 45% | 46% |
| Perry Traditional Academy | | | 19% | | | 20% | 20% |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 14% | | | 15% | 15% |
| Schenley High School | | | 30% | | | 31% | 30% |
| South Vo Tech High School | | | 46% | | | 39% | 42% |
| Westinghouse High School | | | * | | | 45% | 45% |
| Pittsburgh District-wide | 25% | 33% | 29% | 34% | 46% | 40% | 35% |

Note: Rates are estimated from two cohorts of first-time ninth graders (1999-2000, 2000-2001). Predicted graduation and drop-out probabilities are used for students who leave the district. See Appendix B for details, including confidence intervals for graduation rates in Table 10.

[6] *Leaving Boys Behind: Public High School Graduation Rates.* Jay Greene and Marcus Winters. Center for Civic Innovation. Civic Report No. 48. April 2006. Available online at (http://www.manhattan-institute.org/pdf/cr_48.pdf).

Graduation rates vary considerably in the different high schools throughout Pittsburgh. Because the schools differ in the racial composition of their student populations, a portion of the graduation rate variation among the schools can be attributed to district-wide differences in graduation rates by race. However, Table 1 shows substantial variation in graduation rates across schools among students of the same race. For example, although the graduation rate for all students is 75 percent at Allderdice and 69 percent at Schenley, the graduation rate for African American students is lower at Allderdice (60 percent) than at Schenley (69 percent). Understanding the graduation racial gap in the district requires looking within the individual schools themselves.

**Alternative Definitions of Graduation and Dropout Rates**

The first alternative simply replicates the analysis at the end of four years rather than at the end of five years. At the end of five years, very few students return for another year, so the estimated graduation and dropout rates add to approximately 100 percent. A substantial number of students return for a fifth year, so the four-year graduation and dropout rates add to less than 100 percent. The four-year graduation and dropout rates are presented in Tables 4a and 4b in Appendix A. District-wide, the four-year graduation rate is estimated to be 60 percent. The racial gap in the four-year graduation rates is greater than the racial gap in the five-year rate because African American students are more likely than White and Other students to graduate in their fifth year.

The graduation rates presented in Table 1 above and in Table 4 in Appendix A can be compared to rates that are calculated from cross-sectional data. Using Common Core of

Date from the National Center for Education Statistics (CCD-NCES) data for the 2003 Pittsburgh Public Schools graduating class, there were 1,875 diplomas granted. Based on a ninth-grade class of 3,326 four years earlier, this implies a graduation rate of 56 percent, which is substantially lower than the rate of 64 percent presented in Table 1a. However, if the previous year's eighth-grade enrollment of 2,412 is used as an estimate of entering ninth-graders, the graduation rate is calculated to be 78 percent.

As mentioned above, the previous eighth-grade class is often used as a base because of the high retention of ninth graders, which leads to an overestimate of the number of *first-time* ninth-graders and an underestimate of the graduation rate. But another explanation for a portion of the difference between the size of the eighth- and ninth-grade classes is the influx of students from parochial and other private schools into PPS high schools. The CCD-NCES data do not provide any insight into which source of additional ninth graders is more important—retentions or transfers from private schools. The cross-sectional methods therefore create an uncertainty about the true graduation rate, which by these methods could vary anywhere between 56 percent and 78 percent. Our longitudinal methods avoid this ambiguity.

Choosing a denominator that averages eighth, ninth, and tenth-grade enrollments, as suggested by Greene, produces an estimated graduation rate of 65 percent for the class of 2003—quite close to our preferred, longitudinal estimate of 64 percent.

A hybrid method might use the *actual* number of first-time ninth-graders as the denominator and the actual number of graduates four years later as the numerator. This method, accompanied by a small adjustment for district-wide population change (which uses net migration from a contemporaneous cohort of elementary students, who are too young to drop out) likewise produces a graduation rate of 65 percent for the class of 2003—again, very close to the 64 percent five-year estimate from our longitudinal method (which uses first-time ninth-grade entrants from 1999 and 2000).

The *Education Week* method, which calculates a four-year graduation rate using only two years of data—creating an artificial, synthetic cohort—can also be applied to district-wide data for PPS. Applying this method to NCES data for PPS for the fall of 2002 and 2003 yields an estimated graduation rate of 56 percent. This is only slightly lower than the 60 percent four-year rate that we estimate using the longitudinal data. The difference may be attributable to the cross-sectional method's failure to account for population change or to the difference between using an actual cohort of students (in our longitudinal method) and an artificial, synthetic cohort (in the cross-sectional method).

Another alternative method involves using the longitudinal data but putting students who are coded as "leaving the district" into a separate category, counted neither as graduates nor as dropouts. Here we use all entering ninth-graders as the base for the calculations and calculate the percent of this base that are reported to graduate for the PPS or to drop out. Because the movers make up an alternative category, the graduation and dropout rates are both lower than those in the preferred calculation, and the two rates add to much

less than 100 percent. These rates could be calculated after five years or after four years. We present the five-year version of these graduation and dropout rates in Tables 5a and 5b in Appendix A. We believe, however, that the addition of the third category—movers—is likely to create more confusion than it solves.

Yet another alternative is to eliminate the movers from all calculations. In this case, the base is all entering ninth graders who *are not reported to have moved.* Because we think that many of these movers are actually dropouts whose status was erroneously recorded, we think that this alternative overstates the graduation rate and understates the dropout rate. We present the five-year version of these graduation and dropout rates in Tables 6a and 6b in Appendix A.

The rates that ignore movers can be compared to those calculated using Pennsylvania Department of Education guidelines (described earlier in this report), which implicitly exclude movers from both the numerator and the denominator. The Pittsburgh Public Schools district-wide graduation rate for 2003–04 was reported to be 74 percent, according to state guidelines.[7] This rate is slightly higher than that of the five-year rate of 70 percent that we report in Table 6a. This difference comes about from different treatment of the 9.5 percent of students who remain enrolled in their fifth year. These students are excluded from the Pennsylvania Department of Education denominator but not from the denominator based on the longitudinal data used in our calculations.

[7] Available online at http://www.paayp.com/report_cards/PA/RC05D102027451.PDF.

As a final alternative definition of graduation and dropout rates that were used in this study, The National Governors Association recently endorsed a four-year graduation rate definition that divides the number of on-time graduates in a given year by the number of first-time ninth-graders four years earlier, with adjustments for students transferring into and out of the cohort during the period.[8] This should approximate the results shown in Table 4a, which estimates four-year graduation rates using our method. The accuracy of such a method, however, requires two items currently lacking in PPS data: information on the year of first-time ninth-grade entry for students transferring into PPS, and confidence in the transfer codes for students leaving PPS. The first element is necessary to determine which cohort students transferring in should be assigned to. The second element is necessary to avoid underestimating dropout rates. And finally, as discussed above, we prefer the five-year graduation rate to the four-year rate, because it provides a better estimate of the probability that students will ever graduate.

PPS district-wide averages for the other alternative measures described above (without racial and gender breakouts) also are summarized in Table 2.

[8] National Governors Association Task Force on State High School Graduation Data, *Graduation Counts* (Washington, DC: National Governors Association, 2005).

**Table 2**
**Graduation Rates in the Pittsburgh Public Schools**
**Using Various Methods of Calculation**

| | |
|---|---|
| **Longitudinal methods, using student-level data** | |
| *Five-year Rates Using Predicted Outcomes for Students Who Leave the District (Preferred Method) (Table 1)* | *64%* |
| Four-year Rates Using Predicted Outcomes for Students Who Leave the District (Table 4) | 60% |
| Five-year Rates Eliminating Students Who Leave the District (Table 6) | 70% |
| **Quasi-longitudinal method, using aggregate cross-sectional data** | |
| PDE graduation rate | 74% |
| **Cross-sectional methods** | |
| Ever-graduated rate, using eighth-grade denominator (cf. Warren method) | 78% |
| Ever-graduated rate, using eighth/ninth/tenth grade average denominator (cf. Greene method) | 65% |
| Ever-graduated rate, using actual number of first-time ninth-graders | 65% |
| *Education Week* four-year graduation rate | 56% |

It should be noted that the differences in results between the various alternate methods and the preferred method are likely to vary in different schools and for different populations of students. Methods that produce higher graduation rates in Pittsburgh overall may produce lower estimates for some populations.

## Pittsburgh's Graduation Rate in a National Context

Comparing Pittsburgh's estimated five-year graduation rate of 64 percent to the rates of other districts is not straightforward, because similar longitudinal estimates are unavailable for most districts. Because the Greene method produces an estimate (65 percent) that closely approximates our preferred, longitudinal estimate, however, it may be useful to compare Pittsburgh's graduation rate to Greene-method estimates for other populations. Greene and Winters estimate the national graduation rate for public high schools to be 70 percent. Their estimates for the 100 largest urban districts in the nation

(which do not include Pittsburgh) range from 89 percent at the high end to 42 percent at the low end. A graduation rate of 64 percent would place the Pittsburgh Public Schools squarely in the middle of the distribution among urban districts.

**Recommendations for School District Information Systems**

A national student information system would be the best solution to correct the data problems described above. Such a system would enable tracking of students who move across district and state lines, permitting easy differentiation of movers from dropouts. A statewide information system, such as that now being developed in Pennsylvania, would solve many of the problems because most moves are intra-state. Barring these improvements at the state or national level, there are several steps that the Pittsburgh Public Schools and other districts facing similar challenges can take to improve their understanding of the outcomes of their high school students.

First, districts should review the withdrawal codes that are used for departing students. The district should determine whether the existing codes are adequate to capture the necessary details of exits that students make from the district. In Pittsburgh, for example, the "physically/mentally incapacitated" code is never used in the existing database. If it is not necessary, it should be removed from the system. On the other hand, it is not apparent whether the system has an appropriate code for "maternity leave." Girls who give birth during the school year might be coded with unique withdrawal codes, even if they are temporary withdrawals. Although this example would not directly affect the ability to accurately estimate graduation and dropout rates, a better-designed system of codes

would provide better information about the reasons for withdrawal, permitting the district to design better interventions to reduce dropout rates.

Second, withdrawal codes should have clear documentation regarding when they should be used. Documentation should include examples of complex cases, thereby showing users of the system how to make important distinctions between the various ways in which students can exit the school district. Users should be provided with this documentation and be trained in the appropriate use of the codes. Users from throughout the district should be trained together to ensure consistent use of the codes. Indeed, some central control of code assignments may be necessary to ensure accuracy of data, given that school-based staff may have incentives to under-report dropouts.

Finally, the default withdrawal code for all students leaving district schools should identify them as dropouts (as recommended by the 2005 National Governors Association report). Substituting a "move/transfer" code should require additional documentation that provides explicit evidence that the student has actually enrolled in a public or private school outside of the district. For example, any move code should be accompanied by an indication of a transfer of student records to another district, including contact information for the receiving district and contact information for the individual requesting the records transfer. An occasional audit should be done to examine whether the enrolled are truly present in the receiving district to ensure that move codes are only being used for actual moves.

Similarly, efforts (presumably by school social workers) to track down students who disappear should be documented in the electronic record, when possible. At a minimum, there should be information in the electronic record that makes it easy to locate any associated paper records such as social workers' notes.

**Concluding Remarks**

Graduation and dropout rates require careful definition of terms and can be calculated most precisely by using longitudinal data that follow students even if they leave the administrative entity in question. Consistent definitions and data recording among administrative entities are necessary to compare performance. Although our suggestions regarding improvements to student information systems are aimed directly at tracking students as they change schools and districts, these improvements will also improve the data quality by increasing the ease with which data can be audited.

In the absence of perfect information about graduation and dropping out for students who cease to appear in administrative records, it is possible to provide plausible estimates of graduation and dropout rates by predicting outcomes for students whose outcomes are unknown. For most administrative units of interest—i.e., individual high schools and all but the largest school districts—these methods are likely to produce more-precise estimates of graduation rates than are cross-sectional methods, which require assumptions about grade retention and population change.

The rates presented in Table 1 use longitudinal data in the Pittsburgh Public Schools to estimate graduation rates for the district as a whole and for each of its high schools, stratified by race and gender. Almost 10 percent of students remain enrolled for a fifth year, and many of these graduate; but few students who have not graduated after five years ever complete their degrees. The pattern in Pittsburgh resembles that of national patterns in that graduation rates are lower for males and for African-American students than for females or white students. But the variation in graduation rates among high schools in the district is greater than the variation by race and gender, leaving open the possibility that schools have the ability to influence students' dropout decisions (although some of the differences surely result from unobserved differences in student populations; Pittsburgh's magnet programs, for example, may enroll students with above-average levels of motivation and family support). Why some schools have higher graduation rates than others, and what schools can do to increase their graduation rates are questions that merit immediate and urgent attention—in Pittsburgh and across the country.

# Appendix A: Additional Tables

**Table 3**
**Withdrawal codes**

*Graduates have the following code:*
W09 CERTIFICATE OF GRADUATION

*Dropouts have any of the following codes:*
W00 PLACEHOLDER WITHDRAW CODE
W06 QUIT SCHOOL (17 YEARS OF AGE)
W07 GENERAL EMPLOYMENT CERTIFICATE
W08 FARM/DOMESTIC SERVICE EXEMPTION PERMIT
W12 COMMITTED TO CORRECTIONAL INSTITUTION
W13 DRAFTED/ENLISTED IN ARMED SERVICES
W14 ATTENDED KINDERGARTEN AND WITHDREW
W15 IN CARE OF CHILD CARE AGENCY
W16 RUNAWAY
W17 EXPELLED FROM SCHOOL
W19 IMMUNIZATION INCOMPLETE

*Move codes include*:
W02 PROMOTED/TRANSFERRED TO PUBLIC SCHOOL
W03 PROMOTED/TRANSFERRED TO NONPUBLIC SCHOOL
W04 MOVED FROM THE DISTRICT/ADMINISTRATION UNIT
W05 ATTENDS SCHOOL IN ANOTHER DISTRICT
W10 DECEASED

*Enrolled students have the following code to end the year:*
W18 END OF SCHOOL YEAR

*The following code is in the code book, but not used in the data:*
W11 PHYSICALLY/MENTALLY INCAPACITATED

**Table 4**
**Four-Year Graduation and Dropout Rates in the Pittsburgh Public Schools, Using Predicted Outcomes for Moving Students**

4A: Graduation Rates

| | White and Other | | | African American | | | Grand Total |
|---|---|---|---|---|---|---|---|
| School | Females | Males | Both | Females | Males | Both | |
| Allderdice High School | | | 81% | | | 54% | 73% |
| Brashear High School | | | 65% | | | 46% | 59% |
| Carrick High School | | | 67% | | | 37% | 58% |
| Langley High School | | | 55% | | | 38% | 48% |
| Oliver High School | | | 49% | | | 48% | 48% |
| Peabody High School | | | 41% | | | 45% | 44% |
| Perry Traditional Academy | | | 79% | | | 74% | 77% |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 84% | | | 83% | 84% |
| Schenley High School | | | 68% | | | 63% | 65% |
| South Vo Tech High School | | | 51% | | | 54% | 53% |
| Westinghouse High School | | | * | | | 49% | 49% |
| Grand Total | 72% | 63% | 68% | 59% | 45% | 52% | 60% |

4B: Dropout Rates

| | White and Other | | | African American | | | Grand Total |
|---|---|---|---|---|---|---|---|
| School | Females | Males | Both | Females | Males | Both | |
| Allderdice High School | | | 17% | | | 35% | 23% |
| Brashear High School | | | 29% | | | 42% | 33% |
| Carrick High School | | | 26% | | | 48% | 32% |
| Langley High School | | | 38% | | | 44% | 40% |
| Oliver High School | | | 46% | | | 37% | 40% |
| Peabody High School | | | 49% | | | 41% | 42% |
| Perry Traditional Academy | | | 18% | | | 17% | 18% |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 12% | | | 13% | 13% |
| Schenley High School | | | 29% | | | 28% | 28% |
| South Vo Tech High School | | | 44% | | | 35% | 38% |
| Westinghouse High School | | | * | | | 41% | 41% |
| Grand Total | 24% | 31% | 28% | 30% | 42% | 36% | 32% |

**Table 5**
**Five-Year Graduation and Dropout Rates in the Pittsburgh Public Schools, Using Reported Outcomes for Moving Students**

5A: Graduation Rates

| | White and Other | | | African American | | | Grand Total |
|---|---|---|---|---|---|---|---|
| School | Females | Males | Both | Females | Males | Both | |
| Allderdice High School | | | 77% | | | 51% | 68% |
| Brashear High School | | | 61% | | | 43% | 55% |
| Carrick High School | | | 66% | | | 35% | 58% |
| Langley High School | | | 52% | | | 38% | 47% |
| Oliver High School | | | 37% | | | 45% | 42% |
| Peabody High School | | | 26% | | | 44% | 41% |
| Perry Traditional Academy | | | 74% | | | 76% | 75% |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 81% | | | 79% | 80% |
| Schenley High School | | | 63% | | | 65% | 64% |
| South Vo Tech High School | | | 45% | | | 52% | 49% |
| Westinghouse High School | | | * | | | 45% | 45% |
| Grand Total | 67% | 59% | 63% | 57% | 45% | 51% | 57% |

5B: Dropout Rates

| | White and Other | | | African American | | | Grand Total |
|---|---|---|---|---|---|---|---|
| School | Females | Males | Both | Females | Males | Both | |
| Allderdice High School | | | 8% | | | 22% | 13% |
| Brashear High School | | | 20% | | | 31% | 24% |
| Carrick High School | | | 22% | | | 41% | 27% |
| Langley High School | | | 30% | | | 36% | 32% |
| Oliver High School | | | 26% | | | 25% | 25% |
| Peabody High School | | | 20% | | | 30% | 28% |
| Perry Traditional Academy | | | 12% | | | 14% | 13% |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 10% | | | 8% | 9% |
| Schenley High School | | | 20% | | | 26% | 24% |
| South Vo Tech High School | | | 31% | | | 28% | 29% |
| Westinghouse High School | | | * | | | 30% | 30% |
| Grand Total | 16% | 22% | 19% | 24% | 31% | 27% | 23% |

Note: Because the movers make up an alternative category, the graduation and dropout rates are both lower than those used in the preferred calculation (Table 1). Using this method, the graduation and dropout rates add to much less than 100 percent.

**Table 6**
**Five-Year Graduation and Dropout Rates in the Pittsburgh Public Schools, Eliminating Moving Students**

6A: Graduation Rates

| | White and Other | | | African American | | | Grand Total |
|---|---|---|---|---|---|---|---|
| School | Females | Males | Both | Females | Males | Both | |
| Allderdice High School | | | 90% | | | 69% | 84% |
| Brashear High School | | | 75% | | | 56% | 69% |
| Carrick High School | | | 74% | | | 46% | 67% |
| Langley High School | | | 62% | | | 46% | 56% |
| Oliver High School | | | 59% | | | 64% | 62% |
| Peabody High School | | | 57% | | | 59% | 58% |
| Perry Traditional Academy | | | 86% | | | 84% | 85% |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 89% | | | 91% | 90% |
| Schenley High School | | | 76% | | | 71% | 73% |
| South Vo Tech High School | | | 58% | | | 63% | 61% |
| Westinghouse High School | | | * | | | 59% | 59% |
| Grand Total | 80% | 73% | 76% | 70% | 58% | 64% | 70% |

6B: Dropout Rates

| | White and Other | | | African American | | | Grand Total |
|---|---|---|---|---|---|---|---|
| School | Females | Males | Both | Females | Males | Both | |
| Allderdice High School | | | 10% | | | 30% | 15% |
| Brashear High School | | | 25% | | | 41% | 30% |
| Carrick High School | | | 26% | | | 53% | 32% |
| Langley High School | | | 35% | | | 43% | 38% |
| Oliver High School | | | 41% | | | 36% | 37% |
| Peabody High School | | | 43% | | | 40% | 41% |
| Perry Traditional Academy | | | 14% | | | 16% | 15% |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 11% | | | 9% | 10% |
| Schenley High School | | | 24% | | | 28% | 27% |
| South Vo Tech High School | | | 41% | | | 34% | 37% |
| Westinghouse High School | | | * | | | 40% | 40% |
| Grand Total | 19% | 26% | 23% | 29% | 40% | 34% | 29% |

## Appendix B:
## Predicting Outcomes for Students Who Leave the District

This appendix describes the method used to predict graduation and dropout probabilities for students who move out of the PPS district. We used a multinomial logit regression to estimate the relationship between student characteristics and observed high school outcomes for those students who do not move out of the district. We then used the estimated relationship to predict the outcome for students whose outcome we cannot observe because they move out of the district.

The estimation is intended to capture the separate effects of race, sex, age upon entry into ninth grade, and changing schools during high school. Table 7 provides the results of the estimation. Because many of the explanatory variables are dichotomous, we need to define a set of comparison characteristics. The comparison we use is to a white female who entered ninth grade at the average age (14.5 years old) and did not change high schools. At the end of five years, such a person has 12.5 percent chance of having dropped out, a 0.2 percent chance of still being enrolled, and an 87.3 percent chance of having graduated.

The relative risk ratios reported in Table 7 provide information on the relationship between student characteristics and the likelihood that they have either dropped out, are still enrolled, or have graduated at the end of five years. The relative risk ratios indicate the difference in relative risk between a person with the indicated characteristic and that of the comparison group.

For example, the coefficient in the "Graduated" column for African-American males of 0.589 means that the odds of graduating relative to dropping out are .589 as great for African-American males as for White and Other females. The odds of graduating relative to dropping out are defined as the ratio of the probabilities. For White and Other females, these odds can be calculated from the probabilities provided above. The odds are 87.3 / 12.5 = 4.12. For African-American males, the odds are 58.9 percent of this value, or 2.42. These odds, in combination with the new odds for continued enrollment, are equivalent to 80.0 percent of African-American males graduating, 0.6 percent continuing enrollment, and 19.4 percent dropping out.

The relative risk ratios are multiplicative. An African-American male who enters ninth grade one year older than average (age=15.5) and changes schools during high school has a graduation relative risk of 0.589 * 0.377 * 0.705 * 0.145 = 0.023. This relative risk, in combination with the relative risk for continued enrollment, implies a graduation rate of 13.1 percent for such students, a continued enrollment rate of 4.5 percent, and a dropout rate of 82.5 percent.

The race and gender coefficients suggest that males are significantly less likely to graduate than females, and that black males are significantly less likely to graduate than any other race/gender group. The difference in rates between black males and non-black males is statistically significant. Race and gender do not have a significant relationship with the probability of still being enrolled after five years. We allowed for a nonlinear

relationship between outcomes and age by including both the difference between age and the average age as well as the square of this variable. The risk ratio less than one on the linear term implies that for kids at approximately the average age, older kids have lower probabilities. The risk ratio less than one on the squared term implies an inverted-U shape to this relationship. The coefficients on the age and age-squared variables suggest that students who enter ninth grade at 13.8 years of age have the highest odds of graduating.

**Table 7**
**Relative Risk Ratios from Multinomial Logit Estimation**

Risk is reported for continued enrollment and graduation relative to dropping out.

| | Continued Enrollment | Graduated |
|---|---|---|
| Race and Gender (White and Other Female is comparison) | | |
| African-American Female | 1.048 | 0.917 |
| White and Other Male | 1.103 | 0.727* |
| African-American Male | 1.867 | 0.589* |
| Difference from average age upon entry into ninth grade | | |
| Years older | 0.551* | 0.377* |
| Years older squared | 1.059 | 0.705* |
| School mobility (Did not change schools is comparison) | | |
| Changed schools | 2.622* | 0.145* |

* significant at 1 percent (i.e. $p < 0.01$)

The coefficients from this model are used to predict the probabilities of graduating, remaining enrolled, and dropping out for each individual in the data set for whom these outcomes cannot be observed because they leave the district. The graduation and dropout rates reported in Tables 1 and 4 are calculated by averaging the appropriate predicted probability for students who left the district and the actual outcomes for students who did not leave the district.

This model produces relatively precise estimates of graduation rates for the district, and for the subgroups defined by school, race, and sex. The precision of these estimated rates

depends on the sample sizes for the district and for each subgroup and on how many of the students left the district. Table 8 provides the sample sizes for each of the subgroups. Data on over 6,000 students were used in the analysis, but some of the demographic subgroups in particular schools included fewer than 100 students. The reliability of estimates for small subgroups will be less than the reliability for larger subgroups.

The other factor that affects the precision of the estimates is the number of students who moved out of the district. Table 9 presents the proportion of district leavers for the district as a whole and for each subgroup. Although the overall rate of leaving the district was less than 20 percent, for some subgroups it was substantially higher. For example, over 50 percent of White and Other students at Peabody High School moved out of the district. In combination with the small number of such students, we expect a very imprecise estimate of the graduation rate for these students.

The precision of the estimates of graduation rates based on the model is shown by the confidence intervals presented in Table 10. This table reprints the graduation rate estimates presented in the top panel of Table 1 and adds 95 percent confidence intervals. A wider interval indicates more error (i.e., less precision) in the estimate. There are two sources of error in the estimates. The first is the usual sampling error that arises whenever statistics from a particular sample or cohort are used to make inferences about a larger population. The second source of error arises from the estimation error in the model used to predict graduation rates. The confidence intervals presented in Table 10 are from a bootstrap process that re-samples the data 1,000 times and repeats the estimation each

time, thereby accounting for both sampling and estimation error. The reported intervals are based on a normal approximation.

The confidence intervals range from a width of +/- 1.3 percent for the district as a whole to almost +/ 20% for some sub-groups such as White and Other students at Peabody High School. These confidence intervals can be used to determine which rates are significantly different from each other. If two intervals do not overlap, then the rates differ significantly. For example, the district-wide graduation rate for White and Other Females is significantly higher than each of the other three demographic groups and the district-wide rate for African-American Males is significantly lower than that for the other three groups. The rate for White and Other Males is not significantly different from the rate for African-American Females. All of the school graduation rates have a confidence interval less than +/- 6 percent, indicating that two schools with estimated rates that differ by more than 12 percentage points are significantly different.

The reported precision of these estimates takes as given the model that we used to predict graduation probabilities. Different models would produce different estimates, with more or less bias and precision. Unfortunately, it is not possible to test all the assumptions on which our model is based, so we do not know if other models would be better. However, we performed several analyses to check the sensitivity of the outcomes to our modeling assumptions.

We considered including standardized test-scores among the variables used to predict graduation, but scores were missing for approximately one-third of the students (largely because the students entered the district in high school, after the eighth-grade state test). Among the students for whom eighth-grade scores on the Pennsylvania System of Student Assessment are available, scores are related to the probability of graduation. Although it might have been useful to have had scores for all of the students, we were able to confirm that the scores of students leaving the district—after controlling for age, sex, and race—were similar to those of students transferring within the district. This suggested that little predictive power is lost by the absence of the scores.

We also considered using only students who moved within the district as the estimation sample for the outcome predictions. The current model uses all students and includes covariates for demographic characteristics and movement within the district. Such a model assumes that the association of demographic characteristics with graduation and dropping out is the same for movers and non-movers. Using only movers would loosen this assumption, but at the cost of using a smaller sample for estimation. We find that the predicted outcomes arrived at using only movers differ minimally (by no more than one percentage point for district-wide results by race and sex and for schools' overall results) from those using both movers and stayers with only a covariate to indicate moving. Therefore, we prefer to use all students in the estimation in order to gain the benefit of the additional sample size.

**Table 8**
**Sample Size of Analytic Cohorts,**
**Pittsburgh Public Schools**

| | White and Other | | African American | | Grand Total |
|---|---|---|---|---|---|
| School | Females | Males | Females | Males | |
| Allderdice High School | 276 | 327 | 128 | 149 | 880 |
| Brashear High School | 280 | 267 | 123 | 132 | 802 |
| Carrick High School | 250 | 260 | 102 | 96 | 708 |
| Langley High School | 101 | 128 | 76 | 78 | 383 |
| Oliver High School | 100 | 122 | 245 | 217 | 684 |
| Peabody High School | 35 | 46 | 196 | 205 | 482 |
| Perry Traditional Academy | 146 | 143 | 140 | 143 | 572 |
| Pgh HS Crt/Prfm Arts (CAPA) | 71 | 37 | 54 | 23 | 186 |
| Schenley High School | 137 | 162 | 232 | 217 | 748 |
| South Vo Tech High School | 61 | 76 | 126 | 71 | 334 |
| Westinghouse High School | * | * | 176 | 153 | 329 |
| Pittsburgh District-wide | 1458 | 1568 | 1598 | 1484 | 6108 |

* Fewer than 10 White and Other students attended Westinghouse High School, so this subgroup was dropped from the analysis.

**Table 9**
**Percentage of Students Who Move Out of the District,**
**Pittsburgh Public Schools**

| | White and Other | | African American | | Grand Total |
|---|---|---|---|---|---|
| School | Females | Males | Females | Males | |
| Allderdice High School | 12% | 17% | 24% | 28% | 18% |
| Brashear High School | 20% | 17% | 24% | 22% | 20% |
| Carrick High School | 10% | 12% | 21% | 24% | 14% |
| Langley High School | 13% | 18% | 20% | 14% | 16% |
| Oliver High School | 34% | 39% | 24% | 37% | 32% |
| Peabody High School | 49% | 59% | 26% | 25% | 30% |
| Perry Traditional Academy | 13% | 15% | 6% | 14% | 12% |
| Pgh HS Crt/Prfm Arts (CAPA) | 10% | 8% | 11% | 17% | 11% |
| Schenley High School | 17% | 16% | 6% | 11% | 12% |
| South Vo Tech High School | 31% | 17% | 20% | 15% | 20% |
| Westinghouse High School | * | * | 19% | 29% | 24% |
| Pittsburgh District-wide | 17% | 19% | 18% | 23% | 19% |

**Table 10**
**Five-Year Graduation (95% Confidence Intervals)**
**Pittsburgh Public Schools**
**Using Predicted Outcomes for Students Who Leave the District**

| | Graduation Rates | | | | | | |
|---|---|---|---|---|---|---|---|
| | White and Other | | | African American | | | Grand Total |
| | Females | Males | Both | Females | Males | Both | |
| Allderdice High | | | 82.5% (79.9, 85.1) | | | 59.6% (53.4, 65.9) | 75.3% (72.7, 77.9) |
| Brashear High | | | 68.8% (64.7, 72.8) | | | 50.4% (43.6, 62.6) | 62.9% (59.3, 66.5) |
| Carrick High | | | 70.5% (66.4, 74.6) | | | 43.1% (34.9, 51.4) | 62.9% (59.0, 66.8) |
| Langley High | | | 58.6% (52.0, 65.2) | | | 44.2% (35.4, 53.0) | 52.8% (47.5, 58.1) |
| Oliver High | | | 50.4% (42.0, 58.7) | | | 54.4% (49.3, 59.6) | 53.1% (48.8, 57.4) |
| Peabody High | | | 47.0% (30.1, 63.8) | | | 52.8% (47.2, 58.4) | 51.8% (46.4, 57.3) |
| Perry Trad. Ac. High | | | 80.7% (76.7, 84.8) | | | 79.3% (74.7, 83.9) | 80.0% (77.0, 83.1) |
| Pgh HS Crt/Prfm Arts (CAPA) | | | 85.4% (79.5, 91.3) | | | 84.5% (77.4, 91.5) | 85.0% (80.6, 89.5) |
| Schenley High | | | 70.0% (64.7, 75.2) | | | 67.8% (63.4, 72.1) | 68.7% (65.3, 72.1) |
| South Vo Tech High | | | 53.1% (43.6, 62.6) | | | 58.7% (51.5, 65.9) | 56.4% (50.8, 62.0) |
| Westinghouse High | | | * | | | 53.2% (47.2, 59.2) | 53.2% (47.2, 59.2) |
| Pittsburgh District | 74.6% (72.4, 76.9) | 66.2% (63.8, 68.6) | 70.3% (68.6, 71.9) | 64.6% (62.1, 67.0) | 51.4% (48.6, 54.3) | 58.2% (56.3, 60.2) | 64.1% (62.9, 65.4) |