IMPROVING GRADUATION RATES:
DATA-DRIVEN DROPOUT PREVENTION

Twenty-one states and the District of Columbia are working to implement graduation requirements designed to prepare all students for college and careers. At the same time, federal and state policies are putting new pressure on high schools to improve graduation rates. Those dual goals—raising graduation rates while raising graduation standards—are the right ones for America's high schools but many educators wonder whether it is possible to achieve both simultaneously and how much it will cost. Fortunately, the latest research shows that the dropout problem is not an inevitable, immutable feature of American education. Demographics matter, but what happens in schools has a greater impact on whether students stay in school and graduate. Most students who do drop out follow identifiable pathways and research suggests we can do a much better job predicting which students are most likely to drop out; the key is better use of data.

By tracking students over time, systems can identify where in the pipeline they need to target the greatest resources. For example, a number of studies have found that "transition years" are major stumbling blocks on the way to graduation for many students. Although a small group of students exhibit early warning signs in elementary school, large numbers develop them either in sixth grade or ninth grade, years when they are transitioning into generally less supportive but more demanding school settings. Second, systems can identify the schools where resources should be targeted—not only the schools with high numbers of at-risk students but also the schools which contribute to the problem above and beyond what those individual risk factors would suggest. Researchers in Chicago found that dropout rates vary widely across the district's public high schools—even after taking into account students' race, gender, poverty, prior academic achievement and whether they are overage for their grade. Research has also found that ninth grade failure rates vary widely across high schools—even after taking into account individual risk factors—as do rates of recovery from ninth grade failure.

Finally, systems can analyze their at-risk and dropout populations to craft a set of solutions that recognize the many kinds of needs within them. For example, New York's Office of Multiple Pathways has conducted analyses that have led to different solutions for different populations— from alternative programs for dropouts just a few credits shy of graduating to "transitional high schools" for overage, under-credited students who find it very difficult to recover and graduate in a large, traditional high school setting.

COMMON EDUCATIONAL RISK FACTORS

Since the early 1990s, several longitudinal studies zeroing in on individual school districts have shed light on *how*, in addition to *why*, students drop out of school. In Philadelphia, researchers found that 50 percent of all eventual high school dropouts could be identified as early as sixth grade on the basis of just four educational indicators—low attendance, receiving a poor classroom behavior mark from one or more teachers, failing mathematics or failing English. Among those eighth graders with low attendance or a failing grade in math or English, fewer than 75 percent graduate on time. And ninth graders who show no risk factors in eighth grade but have low attendance, earn fewer than two credits, or fail to be promoted also have at least a 75 percent chance of dropping out.

A study conducted in Fall River, Massachusetts revealed that the district had two very distinct types of dropouts — those who left school between 7th and 9th grades (early dropouts), and those who left during 10th through 12th grades (later dropouts). Not only did members of these two groups leave school at different points in the pipeline, they also had very different educational careers — exhibiting different risk factors at different points in time. This research also uncovered powerful evidence that student performance during the transition years — from elementary to middle school and from middle to high school — is a decisive predictor of which students would later drop out.

By tracking students over time, researchers in Chicago created an "on-track indicator" that accurately identifies 85 percent of the ninth grade students who will eventually drop out of Chicago's public high schools. A student is considered "on track" at the end of ninth grade if he or she has accumulated enough course credits to earn promotion to tenth grade while receiving no more than one F (based on semester marks) in the core academic subjects.

Table 1. *Examples of Risk Factors that Significantly Increased the Odds of Dropping Out of High School*

Type of Risk Factor	Chicago[1]	Philadelphia[2]	Fall River[3]
Academic Performance	*Receiving more than one grade of F in core academic courses or not earning enough credits to be promoted after 9th grade	*Earning an F in English or mathematics during 6th grade *Failing one or more courses during 8th grade *Entering 9th grade with math or reading scores below 8th grade level *Being retained in 9th grade	*Significant drop in grade point average from 8th to 9th grade *Being retained in any grade during K–8 or in high school
Educational Engagement	n/a	*Low attendance (80 percent or lower) during 6th grade *Receiving a failing classroom behavior mark during 6th grade *Low attendance during 8th grade *Low attendance during the first 30 days of 9th grade	*Significant drop in attendance beginning in 6th grade and worsening in subsequent years

---

[1] http://ccsr.uchicago.edu/publications./07WhatMattersFinal.pdf
[2] http://www.projectuturn.net/downloads/pdf/Unfulfilled_Promise_Project_U-turn.pdf
[3] Roderick, M. (1993). *The Path to Dropping Out: Evidence for Intervention.* Westport, CT: Auburn House.

HOW TO CONDUCT A LONGITUDINAL COHORT STUDY

The studies described above provide a critical foundation for states or systems working to develop their own early warning system, yet there are many other considerations states must consider as they begin to collect and analyze data to accurately predict which students are on track to graduation and which aren't. Below, we explore the various steps states need to take – as well as the related implications– to successfully build a longitudinal cohort study.

Step 1: Obtain top-priority student-level data elements for initial analysis

Longitudinal data, information collected and accumulated over time, permit analysts to follow the progress of individual students from grade to grade as members of a "cohort," or a group of students who start out in the same grade at the same time. Such data make it possible to observe what happens to students who develop risk factors *at any point* along the way, and thereby to paint a more detailed, nuanced portrait of the patterns and pathways students tend to follow as they move toward dropping out or graduating from high school.

To keep track of which students develop risk factors, when they develop them, and how they fare throughout the pipeline, school systems will have to build electronic recordkeeping systems, or data warehouses, that maintain information on individual students and schools over time.

The Data Quality Campaign has generated a robust and comprehensive set of resources that articulate the importance of building P-20 longitudinal data systems, identify the critical components and highlight the necessary state actions to maximize use of the data fully.

Step 2: Analyze the information to identify high-yield risk factors

The goal of this step is to identify a group of high-yield risk factors, ones that do the best job — and a good enough job — predicting which students will drop out (e.g., 75 percent or more of the students who had the risk factor did not graduate on time). Analysts ought to follow students' time, examining outcomes for subgroups within the cohort that exhibit various characteristics, determining what proportion drop out or graduate, as well as what proportion of students who do *not* exhibit a given risk factor drops out or graduates. However, if that method does not identify high-yield risk factors, analysts can then begin to look for them in more complex configurations of data (and, to hone the analysis, they might want to examine more complex configurations anyway):

a) Combinations of indicators. For example, analysts can examine what happens to students who exhibit different *combinations* of multiple risk factors. At this point, analysts should examine not just whether educational characteristics (academic performance and educational engagement) are higher yield in combination, but also whether certain educational risk factors are more predictive for males than for females or for some racial/ethnic groups more than others. Such combinations might include:
   - Failing grade plus low attendance
   - Failing grade plus male
   - Failing grade plus low attendance, male and Hispanic

b) Changes in indicators. Since several studies have found steep declines in academic performance or educational engagement to be good predictors of falling off course for graduation, particularly during transition years, analysts might want to examine changes in student characteristics from one year to the next. For example:

- Difference in GPA (or some kind of mathematical average reflecting overall academic performance) between 5th and 6th grades
- Difference in attendance between 8th and 9th grades

Note that the bulleted examples above are just that — examples of possible configurations of data to examine. In other words, there is not one formulaic set of steps for examining student characteristics to identify high-yield risk factors. Rather, analysts must take a more exploratory approach.

Step 3: Conduct a pipeline analysis based on the high-yield risk factors Identified in Step 2

Analysts should observe the pathways that students with high-yield early warning signs follow as they progress through the pipeline to determine:

- Whether there are any especially troublesome grade levels — such as the transition years — where indicators of academic performance or educational engagement begin to plummet.
- What the impact of high-yield risk factors is in determining intermediate outcomes along the pipeline before students drop out, such as staying on track or failing the 9th grade.

Analysts also should look for patterns in the grade level and year that students drop out of high school to answer questions such as:

- Are there distinct subgroups of early leaving and late leaving dropouts?
- How large is each group?
- How do their educational pathways and warning signs differ?

Step 4: Assess how effective and efficient the highest-yield indicators are and determine the return on investment for utilizing them to trigger specific interventions

Any given risk factor or combination of risk factors can only measure the likelihood that a given student will drop out. A risk factor is never a true predictor, but rather an assessment of how much risk an *individual* student faces based on his or her membership in a *group* whose educational outcomes we can predict much more accurately based on past cohort data. Therefore, policymakers and system leaders can use hypothetical scenarios to assess what various high-yield risk factors might produce if used as predictors to trigger interventions.

This process begins with knowing the answers to several questions about a risk factor:

- What percentage of students with the risk factor dropped out?
- What percentage graduated?

Using that information, system leaders can assess what proportion of dropouts would have been correctly targeted for interventions if a given risk factor or combination of risk factors had been used as the trigger. This is especially important for planning purposes because it helps leaders assess both the potential effectiveness and potential efficiency of various risk factors. The lower the number of "false negatives" (students a risk factor did *not* identify but who *did* drop out), the more *effective* the risk factor is. Zero false negatives means a district would have identified and provided interventions to *all* students who eventually would have become dropouts. While it's not actually possible to get the number of false negatives down to zero in practice, the higher the dropout rate for students that have a risk factor, the more effective the risk factor.

Similarly, the lower the number of "false positives" (students a risk factor *would* have identified but who did *not* drop out), the more *efficient* the risk factor is. Zero false positives means a district would not

have provided interventions to any students who would have graduated anyway, thus saving dollars. Of course, whether policymakers would have "wasted" dollars providing interventions to such students is another question. Providing extra assistance to students who would otherwise graduate with low grades and skills – and likely unprepared for college and careers – might not be a waste of resources.

Reducing false positives is the key to saving money, because it means not spending dollars on students who would graduate on their own without additional help. Reducing false negatives is the key to raising graduation rates, because it means identifying more students who are on the path to dropping out and would benefit from extra help.

Step 5: Conduct a school-level analysis to better understand where individual risk factors are most concentrated and which schools put students at an even greater risk of dropping out

Once the most powerful student-level predictors of dropping out are known for a particular school system, the district can conduct a school-level analysis. For example:

a) Concentration of risk factors: Analysts should determine which middle schools and high schools receive especially high numbers of students with particular high yield risk factors (and identify the most common feeder elementary and middle schools for those middle and high schools.)

b) Contribution to risk factors: Analysts should identify middle and high schools where transition years are most problematic. For example, they can identify schools where:
     * High numbers of students exhibit declines in academic performance or educational engagement during transition years;
     * High numbers of students who previously showed no warning signs exhibit such declines during transition years; and
     * Students in general experience especially steep declines.

c) Impact on outcomes: Analysts should evaluate how much individual high schools themselves contribute to the dropout problem independent of student-level risk factors. They can do that by using regression analyses to "control for" the risk factors that students carry with them when they enter 9th grade to determine which high schools have higher-than-predicted or lower-than-predicted outcomes. They also can calculate adjusted rates on important outcomes for each individual high school including:
     * Adjusted rate of freshmen staying on track (the "on-track indicator");
     * Adjusted 10th grade promotion rate;
     * Adjusted rate of recovery among students who fall off track in 9th grade; and
     * Adjusted four- and five-year graduation rates.

The results of the initial analysis should inform which warning signs, and specific grade levels, the system monitors. In general, systems should target intervention to the grade levels that are the weakest points in their secondary pipeline, or the "crisis spots" that set students on the road to dropping out. If feasible, districts should monitor and report on warning signs in every middle and high school grade level on an annual basis. But they should consider more frequent monitoring and reporting if certain grade levels turn out to be particularly important sticking points on the road to graduation. However, the decision about when and how often to report and act on warning signs should be driven primarily by what a system finds when it conducts its own initial pipeline analysis. An early warning data system can produce a range of report with information helpful for reducing dropout rates such as:

- Reports on which *individual* students exhibit risk factors to ensure school and program staff are targeting intervention efforts;
- Reports of *aggregate* risk factors at the school or district level designed to inform planning, budgetary and evaluation conversations about the design and scope of interventions; and
- Reports of adjusted outcome measures for school-level committees working on improvement plans and for district officials evaluating the impact of investing in particular school-wide interventions and reforms.

Data produced by the system can have uses beyond direct intervention. For example, system leaders should consider investing the resources to provide middle school educators with reports showing how their graduates fare in high school — how many stayed on track as they progressed through high school and how many dropped out.

EXAMPLES OF UTILIZING EARLY WARNING DATA FOR DECISION MAKING

Intervention: Extra help for certain *groups* of students who share a particular risk factor. If many students exhibit risk factors related to low academic performance, for example, policymakers might decide to provide accelerated instruction to such students in the form of catch-up courses.

Intervention: Especially intensive or personalized help for *individual* students. Students who exhibit warning signs related to educational engagement — for example, very low attendance or very poor behavior — might require one-on-one counseling. Schools can target students for counseling based on data from the early warning system. However, if high proportions of students exhibit such warning signs in some schools, the district might consider school-wide interventions instead or in addition to one-on-one counseling. Some school-wide reforms make it much easier to ensure that adults can provide individual support to students instead of simply expecting often overworked counselors to do so.

Prevention: School-level reforms that can *reduce* student risk factors and dropping out. Providing interventions to students who develop risk factors can help improve graduation rates. But district leaders can also intervene on a school-wide level to create conditions that help prevent students from developing risk factors in the first place, and that reduce the negative impact of some schools on graduation rates. For example, are there middle schools and high schools where transition years are especially difficult for students? Do students who enter such schools exhibit big declines in academic performance, educational engagement or both? If academic performance, school or district leaders might consider curriculum changes, professional development or carving out more time for math and literacy. If educational engagement, leaders might consider restructuring those grade levels into small learning communities, instituting adult advocate or mentor programs or restructuring schedules to allow teachers more time to interact in supportive ways with individual students.

HOW STATES CAN HELP

The cost of building an accurate early warning system is relatively small compared with the cost of providing programmatic interventions or system-wide reforms meant to increase graduation rates. But the payoff of basing interventions on accurate data can be huge. A large school system that invests in better data to support dropout prevention can obtain much better results for hundreds of thousands or even millions of dollars less than a similar system whose leaders decide to skip that step.

States can provide a range of resources to help districts conduct cohort analyses—from start-up funding to technical expertise. States can broker collaborative arrangements among geographically- and demographically-similar districts that wish to save time and money by pooling their research and development efforts to identify and employ better risk factors. States can also provide leadership and incentives to inspire districts to become more data-driven in their dropout prevention efforts. Even though most decisions about what kinds of interventions to provide to which students will be made at the local level, states can use their influence to encourage districts to make such decisions based on solid data and sound judgment.

Analyzing which precise risk factors are most predictive of dropping out is relatively quick, easy, and — best of all — inexpensive. Districts that cannot undertake such an analysis should still collect information to analyze student- and school-level needs before spending significant resources on programs meant to improve graduation rates. The major factors that put students at risk of dropping out — low academic performance and educational engagement — are well established, and a district can hedge its bets by collecting multiple kinds of information to measure each of those factors (for example, classroom grades and standardized test scores).

Finally, school system leaders should consider analyzing other high school outcome variables in addition to promotion and graduation. For example, another goal of high school reform is to increase the proportion of graduates who complete a sequence of rigorous college prep courses and score at college-ready levels on state assessments. What factors predict whether students are more or less likely to do so? Can school systems create risk factors and on-track indicators related to college-and-career readiness, too? As states continue to build out their longitudinal data systems and collect indicators of college and career readiness, districts and schools should encourage and leverage those efforts to improve their own capacity for ensuring all students graduate, and all graduate ready for the next step.