

# Data Collaboration in New York City

## The Challenges of Linking HIGH SCHOOL and POST-SECONDARY DATA



prepared by the Annenberg Institute  
for School Reform at Brown University

## ABOUT GATES COLLEGE READINESS INDICATOR SYSTEMS

The Annenberg Institute for School Reform (AISR) at Brown University and the John W. Gardner Center (JGC) at Stanford University have each received three-year grants from the Bill & Melinda Gates Foundation to work together to select a network of sites and develop models for College Readiness Indicator Systems (CRIS). As part of this collaborative effort, AISR and JGC develop, test, and disseminate effective tools and resources that provide early diagnostic indications of what students need to become college ready. The two organizations serve complementary, but distinct roles. JGC develops and studies the implementation of a tri-level (individual, setting, and system) early warning system using a flexible, “design-build” approach with the partner districts. AISR focuses on cross-site learning; brokering expertise and supports for partner districts; understanding issues related to district, municipal, state, and federal contexts; and process documentation. The CRIS sites are Dallas, New Visions for Public Schools (New York City), Philadelphia, Pittsburgh, and San Jose, California.

<http://annenberginstitute.org/cris>

## ABOUT THE ANNENBERG INSTITUTE FOR SCHOOL REFORM

The Annenberg Institute for School Reform (AISR) is a national policy-research and reform-support organization, affiliated with Brown University, that focuses on improving conditions and outcomes for all students in urban public schools, especially those attended by traditionally underserved children. AISR’s vision is the transformation of traditional school systems into “smart education systems” that develop and integrate high-quality learning opportunities in all areas of students’ lives – at school, at home, and in the community. AISR conducts research; works with a variety of partners committed to educational improvement to build capacity in school districts and communities; and shares its work through print and Web publications.

Rather than providing a specific reform design or model to be implemented, AISR’s approach is to offer an array of tools and strategies to help districts and communities strengthen their local capacity to provide and sustain high-quality education for all students.

<http://annenberginstitute.org>

Prepared by:

*Shani Wilkes, Kathryn Brohawn, Pascale Mevs, and Jaemin Lee*

Much of the content of this publication on the Leaky Pipeline project is adapted with permission from the presentation “Linking High School Data and Post-Secondary Outcomes in NYC: A Researcher’s Perspective” Jennifer K. Bell-Ellwanger and Shani Wilkes of the New York City Department of Education Research and Policy Support Group, presented as part of the CRIS work at a December 2011 webinar sponsored by the Annenberg Institute for School Reform at Brown University.

Editing:

*Margaret Balch-Gonzalez*

Graphic Design:

*Haewon Kim*

This Publication is available for download at

<http://annenberginstitute.org/publication/data-collaboration-new-york-city>

© 2012 Annenberg Institute for School Reform at Brown University

## Contents

- 1 Sharing Data to Evaluate College Readiness
- 1 The Leaky Pipeline Project: Linking Secondary and Post-Secondary Data in New York City
- 2 Setting Up a Data Exchange and Collaboration
- 4 Results: New Ways of Analyzing Data
- 5 Lessons Learned: What to Consider When Institutions Collaborate
- 9 Recommendations for College Readiness Data Sharing across Institutions
- 10 Data Collaborations: A Powerful Tool to Inform Policy and Practice



# Data Collaboration in New York City

## The Challenges of Linking HIGH SCHOOL and POST-SECONDARY DATA

### Sharing Data to Evaluate College Readiness

Education leaders across the country confront a growing challenge: too many students are not college ready when they leave high school. Although indicators exist to identify students at risk of dropping out of high school, few indicators of students' college readiness are currently in place, and few districts have linked indicators to practices and policies in ways that would enable action to create meaningful, lasting change.

The College Readiness Indicator Systems (CRIS) initiative – a collaboration between the Annenberg Institute for School Reform (AISR) at Brown University and the John W. Gardner Center for Youth and their Communities at Stanford University, funded by Bill & Melinda Gates Foundation – aims to address this need for better indicators. Five sites receive support to develop and test college readiness indicators, use them to create effective interventions, and share knowledge and best practices with each other. AISR, in collaboration with the sites and national experts, is preparing a series of publications and webinars that aim to disseminate this emerging knowledge on college and career readiness early warning systems with a broad national audience.<sup>1</sup>

Data access and integration, in particular, emerged as key issues in early work with the sites. And combining secondary and post-secondary data to trace student outcomes through high school to college is one of the biggest challenges of data integration.

One approach to this challenge has been a data-sharing collaboration between the New York City Department of Education (NYCDOE) and the City University of New York (CUNY) to evaluate the college preparedness of their shared students. To support this work, the Bill & Melinda Gates Foundation provided funds to the NYCDOE in 2010 for what is known as the Leaky Pipeline project. In addition to the capacity to analyze data in new ways, the NYCDOE researchers gained many insights into the challenges and best practices of developing and operating a PreK–20 data-sharing collaborative. In this publication, developed in collaboration with AISR as part of the CRIS initiative, these researchers aim to share their insights with others who are engaged in or seek to engage in the work of sharing secondary and post-secondary data across institutions to support college readiness.



<sup>1</sup> For more on the CRIS initiative, go to [www.annenberginstitute.org/CRIS](http://www.annenberginstitute.org/CRIS).

## The Leaky Pipeline Project: Linking Secondary and Post-Secondary Data in New York City

As policymakers seek to understand the relationship between secondary outcomes and post-secondary success, data-sharing collaborations that track and analyze student data from prekindergarten through college are becoming increasingly common. As of 2010, forty states were developing preK–16 or preK–20 data systems that track students from prekindergarten through or beyond college (Data Quality Campaign 2010). These data-sharing collaboratives are helping to answer important questions: What percentage of a school district's high school graduates enroll in college within a certain timeframe following high school graduation? What percentage of students within a district require remediation upon entering college? What are the factors that influence whether students successfully enroll in college or transfer from two-year to four-year colleges? The data-sharing collaboratives also aim to determine whether the answers to these questions differ for different groups of students.



In New York City, which has both the largest urban school system in the United States, with 1.1 million students enrolled, as well as the largest public university system, such a collaborative proved to be ideal. A large overlap of students between the New York City Department of Education (NYCDOE) and the City University of New York (CUNY) provides a wealth of data with which to answer important questions that can hold implications for a large number of students. Roughly 40 percent of the cities' public school graduates enroll in the public university system within one year of graduating from high school, and roughly 70 percent of CUNY first-time freshman have graduated from the NYCDOE. Combining data from both sources aims to answer questions such as:

- What are the outcomes for NYCDOE students after they enroll at CUNY?
- What is the variation in college outcomes and trajectories of students among NYCDOE high schools?
- Which schools have the greatest success in preparing students for college?
- What are the college outcomes and trajectories of students with particular characteristics and achievement histories, such as students who have received a certain type of diploma, participated in Advanced Placement courses, and achieved different scores on standardized tests?

To answer these questions, among others, NYCDOE and CUNY started to share their data in 2008, and with the Leaky Pipeline grant from the Bill & Melinda Gates Foundation in 2010, NYCDOE began to further analyze the college outcomes of its students and the factors that lead to college readiness. Specific research goals included to directly inform policy, create helpful resources for schools, and generate knowledge to support college preparedness of New York City public school graduates. With the grant award, NYCDOE hired a dedicated researcher to work with CUNY and other partners to conduct analyses, establish a baseline set of college readiness indicators to share with secondary schools, and create a preliminary system to collect and track New York City students' post-secondary outcomes.

## Setting Up a Data Exchange and Collaboration

It took many initial steps to establish a data exchange and collaboration between NYCDOE and CUNY.

### Develop a Memorandum of Understanding.

NYCDOE and CUNY developed a Memorandum of Understanding (MOU) in August of 2008, two years before the Leaky Pipeline grant was awarded. The MOU established a two-way data-sharing agreement by which both institutions would send and receive data in order to conduct research regarding the predictors of post-secondary readiness and success. This partnership created an opportunity to develop common research goals and launched the early stages of a PreK–20 data-tracking system.

### Establish research goals.

With the MOU in place and non-disclosure agreements signed by all researchers working with the data to ensure confidentiality, the work of determining common research goals began. In October

2008, CUNY and NYCDOE formed the College Readiness & Success Working Group to develop research questions.

### Conduct initial data analysis.

The Leaky Pipeline Project provided NYCDOE, for the first time, with the ability to directly link its students' data to their outcomes at CUNY. Initial data analysis was used to answer basic questions about the NYCDOE-to-CUNY pipeline related to demographic characteristics, achievement histories, and college outcomes of NYCDOE students attending CUNY.

### Establish key partners at both institutions.

Several partners within NYCDOE and CUNY were identified early in the process of establishing the collaborative: the NYCDOE Research and Policy Support Group, the CUNY Office of Institutional Research and Assessment, and the CUNY Office of Policy Research.

Prior to beginning analyses, researchers established a data exchange between NYCDOE and CUNY and also sought out additional data sources. Student-level data from NYCDOE and CUNY, as well as from the National Student Clearinghouse (NSC) StudentTracker were used to answer the agreed-upon research questions. Figure 1 describes these data and their sources.

FIGURE 1. Data sources and elements

DATA SOURCE	POPULATION	DESCRIPTION
New York City Department of Education	All NYCDOE students in grades 9–12	<ul style="list-style-type: none"> <li>• demographics, including free and reduced-price lunch status</li> <li>• student transcript data</li> <li>• state test scores, including eight-grade ELA and math scores and Regents exam scores</li> </ul>
City University of New York	All NYCDOE students who applied to or enrolled in CUNY	<ul style="list-style-type: none"> <li>• demographics</li> <li>• test scores: SAT and assessment test results</li> <li>• enrollment in remedial courses</li> <li>• courses and grades</li> <li>• retention and graduation status</li> </ul>
National Student Clearinghouse StudentTracker	All NYCDOE students, arranged by high school cohort	<ul style="list-style-type: none"> <li>• post-secondary enrollment dates and status</li> <li>• school name and characteristics</li> <li>• college graduation status and date</li> <li>• college major</li> </ul>



## Share the results.

Bimonthly updates were provided to the Gates Foundation. Results were shared internally at NYCDOE, and NYCDOE held bimonthly meetings with CUNY to further knowledge of CUNY data and revisit research questions. NYCDOE developed “lessons learned” from the data exchange partnership between CUNY and NYCDOE. These lessons were shared with Gates and other partners of Gates Foundation. For a complete timeline, see Appendix 1.



## Results: New Ways of Analyzing Data

By matching the CUNY and NSC data, NYCDOE was able to create a new set of metrics to classify the characteristics of the NYCDOE student graduates who enrolled in college and analyze their college trajectories and outcomes. These findings were shared with all NYCDOE schools. NYCDOE also developed new accountability metrics to identify and refine the kinds of support schools need to provide to prepare their students for college.

The development of research questions and the identification of these metrics was a cyclical process. Many analyses were replicated with different cohorts of students to examine trends, and as new analyses were conducted, research questions were often revisited.

### New metrics for analyzing student characteristics and college outcomes

The following metrics were created for the use of schools, using secondary and post-secondary data, to describe NYCDOE graduates who enrolled in college and to trace their outcomes.

Enrollment at CUNY (using CUNY data):

- Readiness/need for remediation
- Secondary achievement histories of enrollees vs. non-enrollees
- Persistence
- Association between high school performance and CUNY outcomes/success
- Demographic differences
- Special populations (English language learners, students with disabilities)

Overall college enrollment (using NSC data):

- Readiness based on NYCDOE diploma status
- Secondary achievement histories of enrollees vs. non-enrollees
- Persistence based on consecutive enrollment by semesters



- Demographic differences
- Special populations (English language learners, students with disabilities)

## New NYCDOE accountability metrics

New metrics were also created for NYCDOE to hold schools accountable for college-ready supports for their students.

### *Where Are They Now? Reports: NYCDOE Graduates' Success at CUNY*

Based on the research findings, NYCDOE developed an interactive report for each NYCDOE high school and informed school principals on their students' outcomes after high school graduation and enrollment at CUNY. This report was provided to analyze trends in student progress and success at CUNY, with a particular emphasis on the outcomes of CUNY students needing remediation versus those who do not. Each report included the number of high school graduates currently enrolled in CUNY, students who took remediation courses at CUNY by subject, and the percentage of students still enrolled after remediation. These reports also highlight students' outcomes by demographic characteristics.

### *School Progress Reports*

NYCDOE also developed and added college-ready metrics to New York City's accountability system, which is a city-level system in addition to the state's No Child Left Behind accountability system. Using this city-level data, NYCDOE included three college-ready metrics in each school's progress report to help schools to refine their support for students to graduate college ready. Available to parents, teachers, principals, and school communities, NYCDOE progress reports highlight their school's strengths and

weaknesses by comparing the school with a peer group of up to forty schools with the most similar student population, and with all schools city-wide.<sup>2</sup> By including these college-ready metrics in the reports, now schools and parents can see how many of their graduates received college-level credits (e.g., Advance Placement courses), how many passed remediation according to CUNY standards, and how many enrolled in two- or four-year colleges.

Three additional college-ready behavior metrics will be included in future reports:

### *College Prep Course Index*

Percentage of students who have:

- taken/received a certain score on: Algebra II or Math B Regents exam, Chemistry Regents exam, Physics Regents exam, Advanced Placement exam, and/or International Baccalaureate exams; or
- earned a grade of "C" or higher in a college dual-enrollment course (e.g., College Now, Early College); or

### Types of Results That Can Be Generated by the NYC Data Collaboration

- The number of students graduating from an NYCDOE high school in four years and enrolling in a CUNY program the following fall
- Gender and ethnic differences in enrollment and persistence at CUNY
- The connection between eighth-grade state test scores in math and English and persistence in college
- The college enrollment rates of all high schools across the city

<sup>2</sup> Progress Reports grade each school with an A, B, C, D, or F and are based on student progress (60 percent), student performance (25 percent), and school environment (15 percent). See <http://schools.nyc.gov/Accountability/tools/report/default.htm>

- passed another course certified by NYCDOE as college- and career-ready.

***College Readiness Index***

Percentage of students who have passed out of remediation, according to CUNY’s standards (SAT and Regents scores) by August after their fourth year.

***College Enrollment Rate***

Percentage of students in the graduation cohort who enrolled in a two- or four-year post-secondary institution in the fall after graduating.

For more information on these additional college-ready behavior metrics, see Appendix 2.



**Lessons Learned: What to Consider When Institutions Collaborate**

Between January and April 2011, researchers reflected upon the data collaboration and developed lessons learned from the NYCDOE and CUNY data exchange partnership. This section describes their conclusions about what makes data-sharing collaboratives successful.

***A core set of researchers within and across institutions saves time and avoids duplication.***

Researchers at NYCDOE and CUNY began collaborating on a shared data-tracking system as early as August of 2008, roughly a year and a half prior to receiving the Leaky Pipeline grant. During this time, despite the establishment of the College Readiness & Success Working Group, a core set of researchers had not yet been defined. As one researcher noted,

Prior to the Leaky Pipeline grant, which allowed for a dedicated researcher on college readiness, both data and research were passed around several researchers. Many analyses were duplicated and time was not available for necessary documentation.

In addition to tying up time that could be used for data documentation, this duplication and lack of coordination caused delays. Researchers realized that to pursue this work requires a core set of researchers within and across institutions who communicate frequently and work in collaboration on their research questions, analyses, and agendas.

## Fostering collaboration requires good communication between institutions.

Fostering collaboration between agencies proved to be challenging due to unstructured and inconsistent communication. Researchers learned that setting regularly scheduled meetings with attainable, clear, and specific goals for each meeting was critical for effective communications. This was especially important at the start of the partnership. Researchers met weekly and focused on one data element at each meeting, rather than discussing multiple topics. For example, one meeting might be dedicated solely to interpreting state exam scores and another might revolve around defining remediation at the college level. Once the data exchange is established, bimonthly meetings are needed to continue conversation about research using the data.

Maintaining an accessible, diverse team is also important in improving communication. A team with diverse range of expertise, including programmers, data analysts, researchers, and directors can inform and advise specific concerns raised during the project development and implementation. Equally important is establishing clear roles of the dedicated research team. For instance:

- Programmers can answer specific questions on data structure and design.
- Data analysts and researchers can answer questions related to the best data fields.
- Directors can answer policy-related questions and inform the entire team about any policy changes in their respective institutions.

## Institutions must communicate about data exchange and hold one another accountable for timelines.

Once partners have established which institution will perform the data matching, the next critical step is to maintain communication and a schedule for this process. In the case of the Leaky Pipeline, CUNY performed the matches and provided NYCDOE with a data file twice annually. However, CUNY required a substantial amount of time and resources to conduct this match.

Researchers who are responsible for matching data must communicate constantly in order for all researchers to be aware of any challenges with the timeline as they arise. Challenges should be shared immediately in order to address them in a timely manner and avoid delays. As one example, CUNY had difficulty determining whether some students had attended NYCDOE and delayed their matching process, until they asked the NYCDOE researchers to perform a quality check. Once NYCDOE received the request, this was a very quick process, and both institutions could have benefited if these steps had been a part of the process from the outset. To match data and address challenges promptly, it is important that both institutions not only allocate the resources and time, but also hold each other accountable to adhering to data-sharing timelines.

## Differences in definitions of populations of interest and cohorts should be clarified and accounted for in findings.

When institutions collaborate to use data, differences in defining populations of interest and cohorts are likely to arise based on the way each institution typically views its own populations. For example, to answer the question of what percentage of the cohort are entering college in the first fall after high school graduation, researchers might be using unequal definitions for “cohort.” Figure 2 displays these possibilities.

FIGURE 2. Differences in definitions: NYCDOE and CUNY

	NYCDOE	CUNY
<b>College enrollment is based on:</b>	Ninth-grade entering cohorts (e.g., students who were ninth-graders in 2004)	Students' first fall entry at CUNY (regardless of the year in which they graduated from a New York City high school)
<b>Date of graduation from NYCDOE:</b>	June of the year of interest	Anytime, any year
<b>Date of enrollment at CUNY:</b>	September of the year of interest	September of the year of interest

Since institutions may define their populations of interest and cohorts based on various perspectives and requirements, different definitions are acceptable. However, these differences need to be clearly noted and understood when conducting research and presenting the findings on behalf of both institutions.

### Creating common identifiers and shared data warehouses increases the accuracy of data matching.

In a school district as large as New York City, sharing and matching data can be especially challenging. Matching students on last name, date of birth, and school, for example, would generate over 20,000 duplicates! Thus, correctly identifying NYCDOE students who enrolled in CUNY required understanding the best combination of identifying information. By exploring these combinations, researchers discovered that matching students on first name, last name, and date of birth uniquely identifies 99.95 percent of students enrolled in NYCDOE, and matching students on first name, last name, date of birth, and school uniquely identifies 99.99 percent of students enrolled in NYCDOE (see sidebar on page 9 for examples of the number of duplicates generated by different data combinations).

Combining identifiers or having common identifiers, as well as unique identifiers like the NYCDOE's student identification number, can increase

the accuracy of matching students. NYCDOE assigns a unique identifier to all students when they enter the school system; to both systems' advantage, CUNY collects this identifier on their enrollment applications, which allows for direct student matches.

In addition, creating and maintaining a shared data warehouse can help prevent duplication of work or different reports of findings from researchers at each institution. During the Leaky Pipeline project, research teams at both institutions were analyzing the data in similar ways, which led to duplication of work between institutions, and occasionally different findings were reported. A common data warehouse helps researchers at the various institutions to report consistent student outcome and achievement numbers. For CUNY and NYCDOE, developing a common data warehouse has been a long-term plan. While this shared data warehouse is being created, data exchanges from one institution to the other is acceptable, though creating shared datasets is ideal for shared research questions.

## Detailed data documentation avoids duplication and saves time and resources.

During the NYCDOE-CUNY collaboration, many conversations that were conducted via email or phone were not documented, leading to repetitive conversations, duplicate analyses, and excess time spent on helping new staff members use the data. When engaging in the collaboration process, partners should be sure to create and maintain a detailed data documentation process.

For example, seeking data for “ethnicity” with no further explanation could lead to ambiguity and challenges in sharing data from different sources. A better way to document ethnicity would be to assign a unique number to each race. Likewise, “time at college” could be interpreted in a number of conflicting ways. A clearer definition would be “time student was enrolled in a particular college for that semester (days), created by subtracting enrollment end dates and enrollment.” Also included should be a description, the source, and any clarifying notes. The more documentation on the data, the better! For examples of effective versus unclear data documentation, see Appendix 3.

## Careful reconciliation of discrepancies allows collaborations with other agencies/sources of post-secondary data.

While having access to multiple sources of post-secondary data is a good thing, using multiple sources of post-secondary data can lead to discrepancies among data elements. This proved to be a challenge when using data from the National Student Clearinghouse’s StudentTracker service,

whose enrollment records did not always align with CUNY’s. Students can be identified by the National Student Clearinghouse (NSC) as having enrolled in CUNY, but not identified by CUNY as having enrolled in CUNY, and vice-versa. For example, sometimes a student will be enrolled in CUNY according to NSC, but not according to CUNY, because s/he withdrew after CUNY submitted data to NSC. Other times a student will be enrolled in CUNY according to CUNY, but not according to NSC because NSC was unable to match the student’s enrollment record. Thus, researchers had to determine both how to reconcile discrepancies and what impact this would have on their analyses.

Because access to multiple post-secondary education data sources is a benefit, researchers chose to use the NSC data despite under-reporting of enrollment status, but did so with caution. This meant using footnotes to explain that enrollment records may not be accurate when using NSC data

### What’s in a name?

Matching on first and last name only: 283,446 duplicates

- There are 142 students named “Jose Rodriguez” in NYCDOE.
- There are 215 students named “Unique” enrolled in NYCDOE.

Matching on last name and date of birth only: 169,591 duplicates

- Eight NYCDOE students with the last name of Chen were born on the same day in 1995.

Matching on last name, date of birth, and school: 20,818 duplicates

- Assuming these are all siblings, there are at least 10,126 sets of twins, 182 sets of triplets, and 5 sets of quadruplets attending the same school within NYCDOE.

since the list of colleges that report data and the reporting schedule vary. In addition, researchers merged data sources to create new datasets. For the purpose of the analyses, students who were reported as being enrolled in CUNY based on CUNY data were considered enrolled, but not those students reported in the NSC data. Researchers came to this decision because they trusted the accuracy of the CUNY data: CUNY data systems are updated in real time and provide the most accurate student enrollment records. For more on the data merging process, see Appendix 4.

### Budgeting for the cost of collaboration is critical.

Collaborating on a data exchange can be a costly process, since receiving data from other agencies often requires a fee. To pursue this work, it is critical to budget accordingly and secure funding for accessing data.

In the case of the Leaky Pipeline Project, NYC-DOE and CUNY were able to avoid paying a price for exchanging data by drafting an MOU for a free-of-cost data exchange. However, the cost of accessing data from the NSC StudentTracker service could not be avoided. Obtaining student records from NSC ranged from \$1,000 for up to 1,000 records to \$38,000 for up to 100,000 records.





## Recommendations for College Readiness Data Sharing across Institutions

NYCDOE researchers recommend the following steps to districts and higher-education institutions interested in applying the lessons of the Leaky Pipeline project.

**Create a place to house all data for both internal and external audiences such as principals, teachers, school staff, and parents.**

While student-level data would only be available to internal researchers, aggregated reports created from the data could be available for school staff and parents. This centralized database would give

school staff access to data on post-secondary outcomes for their schools and students. Components can include data to support academic advisement, financial advisement, and awareness of post-secondary options.

**Conduct trainings for school staff.**

School staff should be trained on how to prepare students for college (both academically and financially) and on how to use data on students' post-secondary outcomes to support change at the secondary school level.

### Establishing New Data Exchanges: The FAFSA Completion Pilot Project

Many students who are eligible for financial aid do not receive it because they fail to file the Free Application for Federal Student Aid (FAFSA). While an estimated 1.7 million students do not file the FAFSA each year because they incorrectly believe they are ineligible, one study suggests that helping students and families complete the FAFSA can increase post-secondary enrollment by roughly 30 percent (see [www.nber.org/papers/w15361](http://www.nber.org/papers/w15361)). To address this discrepancy and increase post-secondary enrollment rates, in 2010 the U.S. Department of Education launched the FAFSA Completion Pilot Project, which "aims to provide FAFSA completion data to twenty pilot sites across the country so that each site can focus its resources on students who have not completed the FAFSA and make FAFSA completion one component of a comprehensive college and career ready strategy" (see U.S. Department of Education, "Education Secretary's Senior Advisor on College Access to Hold First Meeting of FAFSA Completion Pilot Project Sites," [www.ed.gov/news/media-advisories/education-secretary-s-senior-advisor-college-access-hold-first-meeting-fafsa-c](http://www.ed.gov/news/media-advisories/education-secretary-s-senior-advisor-college-access-hold-first-meeting-fafsa-c)).

In New York City, one of the FAFSA Completion Pilot Project sites, NYCDOE has been collaborating with the federal government to establish a FAFSA data exchange. This partnership entails biweekly data exchanges from the Federal Student Aid (FSA) office to the NYCDOE that update schools on students' FAFSA completion status. According to NYCDOE, the goal of the pilot is to provide current data to schools to use in assisting students in the FAFSA completion process. Beginning in May 2011, the NYCDOE received data consisting of student identifiers, including first and last name and date of birth, as well as FAFSA completion flags indicating where in the application completion and submission process each student is. The data is currently shared through the ARIS private community, a secure NYCDOE data portal.

The FAFSA data exchange represents another way in which school districts can collaborate with outside agencies to improve post-secondary access and completion rates. Using the up-to-date FAFSA data, school counselors and other college advisors can work to increase the number of students who take the crucial first step toward financial aid of filing the FAFSA, thereby increasing the number of students who enroll in college.



Establish additional data-exchange relationships or obtain other post-secondary-related data.

One example of an additional data-exchange relationship is the U.S. Department of Education's Free Application for Federal Student Aid (FAFSA) Completion Pilot Project, which provides student-level data to schools and districts on their students' FAFSA completion status (see sidebar on page 11). This data can expand work on post-secondary readiness. Districts can collaborate with additional higher-educational institutions in their states; for example, the NYCDOE could work with State University of New York (SUNY) colleges to receive their data, which would enhance the evaluation of students' educational outcomes from Pre-K through university.

## Data Collaborations: A Powerful Tool to Inform Policy and Practice

To ensure that secondary students graduate college ready, it is critical to understand how high school graduates are doing once they are enrolled in college. According to a recent report from the Data Quality Campaign (2011), all states in the country today have robust data for stakeholders to make informed decisions on education reform – and that data is crucial to improve student achievement. Colleges and universities, districts, the U.S. Department of Education, and other institutions also have a wealth of data.

However, while collecting data requires a lot of effort, the real potential for change comes when data can be shared and used to make informed decisions to improve education systems. Few states are actually using their data effectively, and there



are many challenges to sharing data among different institutions. The NYCDOE-CUNY data-sharing collaboration is an example of how cross-institutional systems can be set up that are designed to jointly evaluate, build, and effectively use data in order to improve their education systems and better prepare students for post-secondary success. Through the Leaky Pipeline project, NYCDOE now has a powerful tool to help analyze successful supports, refine ineffective interventions, and better allocate resources to help the city's students become college ready.

Data sharing is not an easy process. As the Leaky Pipeline project has demonstrated, data sharing does not happen by simply exchanging data. The collaboration required NYCDOE and CUNY to develop a shared research agenda; create a new coding system that both could agree on and employ in their respective fields; invest considerable money, time, and staff; establish detailed documentation procedures; and maintain frequent and structured communications.

As the use of data becomes increasingly critical in policy changes, more school systems will need to collaborate with higher-education systems as well as other institutions. The Leaky Pipeline project offers an example of what to consider in developing data-sharing collaborations. Now that states, districts, and other entities have collected invaluable data on their students, using and sharing those data will be an onerous but critical next step in developing effective college readiness indicators and support systems for all students.

## References

- Data Quality Campaign. 2010. *Data for Action 2010*, [www.dataqualitycampaign.org/resources/details/1247](http://www.dataqualitycampaign.org/resources/details/1247)
- Data Quality Campaign. 2011. *Data for Action 2011*, [www.dataqualitycampaign.org/stateanalysis/about](http://www.dataqualitycampaign.org/stateanalysis/about)

## APPENDIX 1

### Timeline

TIMELINE	ACTIVITY DESCRIPTION
<b>October 2008</b>	College Readiness & Success Working Group formed with CUNY and NYC-DOE; Development of research questions
<b>October – November 2008</b>	Data exchanged for purpose of college readiness analytics
<b>October – December 2008</b>	First preliminary analyses conducted (included analyses by demographics and Regents exam scores of NYCDOE Class of 2005 )
<b>February 2009</b>	Next set of analyses, which included more background data (including demographics, Regents, and 8th-grade test scores) and outcome data (such as “on-track” to graduating from CUNY)
<b>March 2009</b>	Results of early analyses presented to the NYCDOE leadership
<b>May – June 2009</b>	Similar analyses conducted, but using different high school cohorts to examine trends
<b>November – December 2009</b>	Analyses specific to the “pipeline” were conducted (this follows the enrollment of students throughout each semester to see the percentage “dropping off” the pipeline)
<b>January 2010</b>	Received National Student Clearinghouse data
<b>February 2010</b>	Received Leaky Pipeline grant, which allowed for dedicated researcher
<b>March – June 2010</b>	Analytics focused on the background characteristics and outcomes of students who enroll in CUNY vs. Non-CUNY schools
<b>March – December 2010</b>	Bi-monthly updates provided to Gates Foundation. Results shared internally at NYCDOE. NYCDOE held bi-monthly meetings with CUNY to further knowledge of CUNY data and revisit research questions.
<b>January 2011 – April 2011</b>	NYCDOE developed “lessons learned” from the data exchange partnership between CUNY and NYCDOE. These lessons were shared with Gates and other Gates partners

## APPENDIX 2

### Additional Metrics on NYCDOE Progress Reports

PHASE-IN METRIC	DESCRIPTION OF METRIC
<p><b>College Preparatory Course Index</b></p>	<p>This metric is based on the percentage of students in the class of 2011 (cohort M) who have:</p> <ul style="list-style-type: none"> <li>• Scored 65+ on the Algebra II or Math B Regents exam, or</li> <li>• Scored 65+ on the Chemistry Regents exam, or</li> <li>• Scored 65+ on the Physics Regents exam, or</li> <li>• Scored 3+ on any Advanced Placement (AP) exam, or</li> <li>• Scored 4+ on any International Baccalaureate (IB) exam, or</li> <li>• Earned a grade of “C” or higher in a college dual enrollment course (e.g., College Now, Early College), or</li> <li>• Passed another course certified by the NYCDOE as college- and career-ready</li> </ul> <p>Students meeting more than one of the requirements above will only be counted once in the numerator.</p>
<p><b>College Readiness Index</b></p>	<p>This metric is based on the percentage of students in the class of 2011 (cohort M) who have graduated and passed out of remediation according to the standards of City University of New York (CUNY) by August after their 4th year. To contribute, a student must:</p> <ul style="list-style-type: none"> <li>• Graduate with a Regents diploma, and</li> <li>• Earn a 75 or higher on the English Regents or score 480 or higher on the Critical Reading SAT, and</li> <li>• Earn an 80 or higher on one Math Regents and demonstrate completion of coursework in Algebra II/Trigonometry or a higher-level math subject, or score 480 or higher on the Math SAT.             <ul style="list-style-type: none"> <li>o A student can demonstrate completion of math coursework by (1) passing a course in Algebra II/Trigonometry or higher and taking one of the following exams: the Math B Regents, Algebra II/Trigonometry Regents, AP Calculus, AP Statistics, or IB Math exam, or (2) by passing the Math B or Algebra II/Trigonometry Regents.</li> </ul> </li> </ul> <p>CUNY is in the process of transitioning to a new standard for math – an interim standard will be in place for 2011, and the new standard will take effect in 2012. For the Progress Report, we will apply the standard for 2012 (the standard described above) in this year’s unscored phase-in metric, to better inform schools about how they are likely to perform in 2012, when the metric will be scored.</p>
<p><b>College Enrollment Rate</b></p>	<p>This metric is based on the percentage of students in the class of 2010 (cohort L) who enrolled in a two- or four-year college or university by December 31, 2010 (the fall after graduating)</p>

## APPENDIX 3

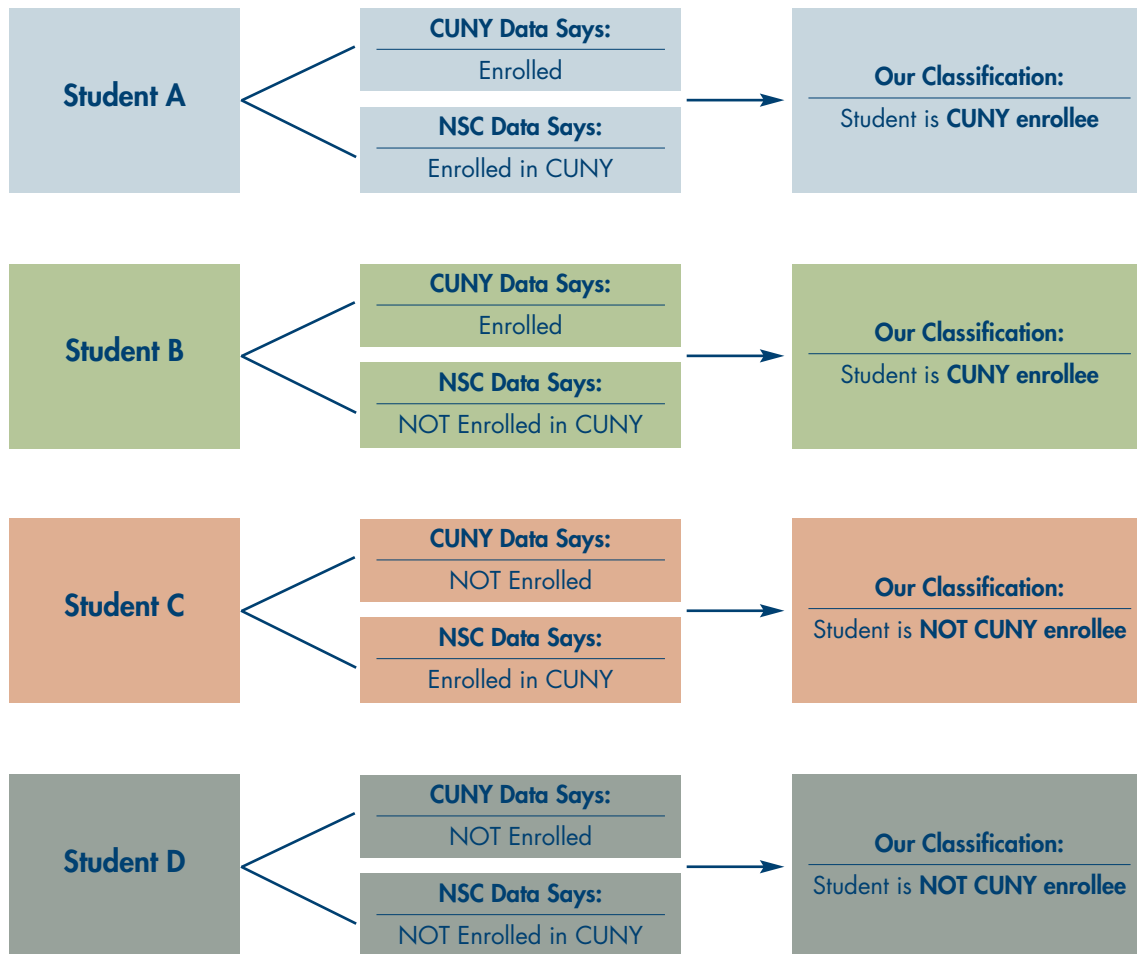
### Effective Data Documentation

EXAMPLE OF CLEAR DATA DOCUMENTATION					
VARIABLE	TYPE	WIDTH	DESCRIPTION	SOURCE	NOTES
GENDER	String	1	Gender	DOE	F=Female; M=Male
ETHNIC	Numeric	1	Ethnicity	DOE	1=Native American; 2=Asian 3=Hispanic; 4=Black; 5=White
TIME_AT_COLLEGE	Numeric	3	Time student was enrolled in particular college for that semester (Days)	Created by DOE Researcher	Created by subtracting enrollment end dates and enrollment begin dates

EXAMPLE OF DOCUMENTATION THAT COULD RESULT IN AMBIGUITIES	
VARIABLE	DESCRIPTION
GENDER	Gender
ETHNIC	Ethnicity
TIME_AT_COLLEGE	Time student was enrolled in particular college for that semester

## APPENDIX 4

### Data-Merging Process





prepared by



Annenberg  
Institute for  
School Reform

AT BROWN UNIVERSITY

- Box 1985  
Providence, Rhode Island 02912
- 233 Broadway, Suite 720  
New York, New York 10279

[AISR\\_Info@brown.edu](mailto:AISR_Info@brown.edu)

[www.annenberginstitute.org](http://www.annenberginstitute.org)

Twitter: @AnnenbergInst