

Unlocking New York City's High School Progress Report

Sean P. Corcoran
Grace Pai

A report commissioned by



Foreword

Among the reform strategies adopted by the New York City Department of Education during the administration of Mayor Michael R. Bloomberg, the push towards school accountability has been the most wide-ranging, influential and, at times, controversial.

In a pioneering effort to improve school accountability and move it beyond a simple recitation of high-stakes test scores, New York City developed the Progress Report (PR), a set of metrics that includes the overall “A” through “F” letter grade that the DOE began issuing annually to most public schools beginning in 2007.

In 2011, in an effort to gain a better understanding of New York City’s accountability system, New Visions for Public Schools commissioned New York University economist Sean Corcoran and his colleague, Grace Pai, to assess various components of the high school PR. The result, *Unlocking New York City’s High School Progress Report*, represents the most comprehensive analysis of the PR to date.

The study attempts to explore how the PR’s methodology can more accurately assess the impact that each of New York City’s more than 400 high schools has on the diverse student population it serves. Our goal is to contribute to the rich and ongoing conversations that the DOE has welcomed since the PR’s first iteration and which have been critical to its significant improvements over time.

By their very nature, accountability systems have many goals and, as a result, are challenging to understand and to interpret. Such systems, the authors write, “focus attention on—and delineate expectations for—desired student outcomes. They provide metrics that the district and its stakeholders can use to monitor school performance. They serve as a learning tool for educators, highlighting success areas or areas in need of improvement. They draw attention to schools that are performing well relative to their peers, providing models of good practice, and provide a basis for intervention or closure when a school is performing poorly.”

One of the central challenges of the PR grade is to fairly account for the differences that schools confront as they serve the city’s famously diverse student population. The DOE’s “Educator’s Guide to the Progress Report” reflects this concern: “The overall Progress Report Grade is designed to reflect each school’s contribution to student achievement, no matter where each child begins his or her journey to career and college readiness. The methods are designed to control for demographic characteristics of students so that the final score for each school has as little correlation as possible with incoming student characteristics such as poverty, ethnicity, disabilities, and English learner status.”

The New York City high school PR evaluates schools on a range of metrics including attendance, credit accumulation, graduation rates, and scores on standardized assessments. Schools are compared to every other school in the city as well as to a group of forty “peer” schools whose incoming students share certain characteristics. In order to account for variation in incoming student populations, the comparison to peer schools is weighted three times as heavily as the citywide comparison.

Using data from the 2010-2011 school year, Corcoran and Pai elaborate on the key features of the PR methodology. Among their findings:

- A school’s performance in each subcategory depends on how it performed relative to the range of outcomes of its peers. Peer groupings are formed using an index that is based on average proficiency of a school’s incoming students, with adjustments for the percent of students who receive special education or are over-age. Importantly, peers only affect a school’s score to the extent they affect the range of outcomes to which its performance is benchmarked.
- While the use of peers moderates the correlation between incoming student characteristics and scores, schools’ overall PR scores remain associated with many pre-existing risk factors, suggesting that a school’s score can be influenced by factors outside of its control. Since schools in peer groups vary by size, location, admission method, poverty rate, and percent of students with disabilities, a wide range of outcomes is likely to be observed with a more diverse peer group. The wider the range, the less a peer group represents a sharp comparison of similar schools.
- The Peer Index has only a modest effect on the overall grade assigned to high schools in large part because of the diversity of peer groups. When the authors calculated PR scores using a formula that ignored peers entirely in favor of a citywide comparison, about two-thirds of high schools received the same grade.
- The weighted Regents passing rate—a subcategory that accounts for a sizable portion of a school’s overall score—is treated differently from other categories, such as high school graduation rate. It is benchmarked, first, against expectations (based on 8th grade test scores) and second, against the peer group and citywide average. The implication of this double benchmarking is that schools with high-achieving students may be penalized for failing to achieve mathematically impossible growth targets.

Despite its limitations, the PR takes important steps toward an accountability system which recognizes schools that substantially improve the performance of highly challenged student populations. This system allows such schools to be judged successful regardless of whether they match the absolute performance of schools where students arrive better prepared. The PR’s current methods take into account the starting point of incoming students, not just through the use of peer groups, but also through quasi-growth metrics such as the weighted Regents passing rate and the ability to earn extra credit for gains with struggling student populations.

We believe that, its limitations aside, the PR has been made progressively stronger with each iteration and can further be improved upon. As Corcoran and Pai point out, “even an imperfect accountability tool can incentivize changes in behavior and improve school performance.”

This study clarifies not only the impact of the current version of the PR, but also points to several key revisions that can improve the PR (specifically, the Peer Index’s methodology) to increase its accuracy. Most significantly, the authors suggest that the peer group approach should be either significantly modified or abandoned altogether. Among the options they propose is comparing actual student performance to predicted performance based on a broad range of student and school characteristics. If the peer group approach is maintained, they suggest that the calculation of peer index should be based on a wider range of incoming student characteristics.

Improving the PR’s effectiveness is important because an accurate assessment helps educators target school improvement efforts. Moreover, New Visions recognizes from years of experience working with a diverse set of high schools that those schools that use the tool most effectively disregard the overall letter grade. Educators instead look closely at performance in each component of the PR to understand where there is need for improvement and how to prioritize and manage change. For this reason, New Visions recommends the department consider emphasizing grades of component parts of the report, such as the environmental survey, student academic growth, and student graduation rates, and attendance. The department should de-emphasize use of a single grade as an indicator of the school’s overall success or failure.

We believe that Corcoran and Pai’s report contributes important insights into the broader conversation about how to improve New York City’s accountability system. In the end, the fundamental premise of New York City’s accountability system—that we must continue to hold ourselves accountable to rigorous, measurable outcomes—is critical to serving this city’s diverse student population effectively. With thoughtful modification, the Progress Report can do a better job of measuring school impact.

New Visions for Public Schools
March 28, 2013

About the Authors

Sean P. Corcoran, Ph.D. is an associate professor of educational economics at New York University's Steinhardt School of Culture, Education and Human Development, and an affiliated faculty of NYU's Robert F. Wagner Graduate School of Public Service. He has been a research associate of the Economic Policy Institute in Washington, D.C. since 2004, and was a visiting scholar in residence at the Russell Sage Foundation in 2005-06. His research includes long-run trends in the quality of teachers, the impact of income inequality and court-ordered school finance reform on the level and equity of education funding in the United States, the properties of value-added measures of teacher effectiveness, and the political economy of school choice reforms. In 2009, he led the first evaluation of the Aspiring Principals alternative certification program in New York. He is currently a visiting scholar at the Center for Education Policy Analysis at Stanford University, where he is completing several studies on the high school choice behavior of middle school students in New York City.

Grace Pai, B.S., M.Ed., is a former Peace Corps volunteer and high school math teacher with New York City's Teaching Fellows program. A current Ph.D. student in international education at NYU's Steinhardt School of Culture, Education and Human Development, she researches teacher development, education and the labor market, marginalized youth populations, and education policy, both domestically and in developing countries.

About New Visions for Public Schools

Since 1989, New Visions for Public Schools has been a leader in fostering innovation and reform to improve the quality of education children receive in New York City's public schools. Our work has reached tens of thousands of students, parents, educators, and community members across the city. Working with the public and private sectors, New Visions develops programs, solutions and strategies to energize teaching and learning and to raise the level of student achievement. Partnering with the New York City Department of Education, New Visions is accountable for improving student achievement in 73 New York City public schools serving more than 40,000 students. As a charter network, New Visions is opening up to 18 charter high schools with the first four now running in the Bronx, and two to open in Brooklyn in fall 2013. As a laboratory, New Visions is researching and developing novel solutions for the needs of schools, teachers, and students. Our goal is to graduate all students ready and successful for college, career, and life.

 NewVisionsforPublicSchools

 @NewVisionsNYC



320 West 13th Street, 6th Floor
New York, NY 10014
Phone: 212 645 5110
www.newvisions.org

Unlocking New York City's High School Progress Report

Sean P. Corcoran
Grace Pai

Institute for Education and Social Policy
New York University
665 Broadway, Suite 805
New York, NY 10012-2300

March 28, 2013

Acknowledgements:

The authors would like to thank Jim Kemple, Aaron Pallas, Lori Nathanson, Mark Dunetz, and Elizabeth Debraggio for helpful feedback and suggestions. Our paper also benefited from a careful review by the Office of Assessment and Accountability at the NYC Department of Education, and in particular Martin Kurzweil, Simone DeSouza, and Adina Lopatin. Funding for this research was provided by New Visions for Public Schools, who is not responsible for the content or recommendations expressed in this report. All remaining errors are solely our own.

Unlocking New York City’s High School Progress Report

Sean P. Corcoran
Grace Pai

New York University

I. Introduction

The Progress Report is the backbone of New York City’s school accountability system. Bringing together a rich set of metrics ranging from student progress on standardized tests to perceptions of the school environment, the report offers parents, educators, and the public a detailed look at how their schools are performing. Parents are encouraged to use the Progress Report in choosing schools; schools are expected to use the report to find areas in need of improvement and to learn from higher-achieving peers; and district leaders rely on the Progress Report to identify underperforming schools. Taken together, it is hoped that by holding schools publicly and administratively accountable, they will improve.¹

The end product of the annual Progress Report is straightforward: a numeric score and letter grade ranging from “A” to “F”. For many of its stakeholders, however, the calculations behind this measure are poorly understood. In this paper, we aim to shed light on the methodology behind the Progress Report and highlight some of its implications for evaluating school quality. We focus exclusively on the high school Progress Report, relying primarily on data from 2010-11. Many of the concepts discussed here, however, can be easily applied to other years and to the elementary/middle school report.

The Progress Report is similar to other accountability tools found in practice in that it serves many purposes. Among other things, it is intended to set expectations about student outcomes; to monitor performance; to encourage educator effort; to provide information to the public; and to serve as a basis for rewards or sanctions, such as bonuses and school closure (Linn, 2006). Designing such a multi-purposed system necessarily requires a number of trade-offs. For example, tying accountability to the *level* of achievement—“status” measures such as test scores or graduation rates—focuses attention on the outcomes of interest and holds all schools to the same standards. Such measures are strongly influenced by out-of-school factors, however, and tend to hold schools accountable for factors outside of their control (Hamilton & Koretz, 2002). Tying accountability to *growth* or *improvement* ameliorates this problem, but raises other questions about the sufficiency and comparability of growth across schools.

A related question is whether outcomes—either status or growth—should be benchmarked against a fixed standard (e.g., *No Child Left Behind’s* proficiency target) or against other schools (“norm-referencing”). Here again, a fixed target allows all to be held to the same high standard, but may set unrealistic expectations for schools serving more disadvantaged populations. Benchmarking to similar schools puts outcomes in a more appropriate context, but runs the risk of institutionalizing low expectations for low-performing schools. It also begs the question of how “similar schools” should be defined. Some accountability systems go to great lengths to statistically adjust outcomes for differences in student populations, while others resist, citing a

loss in transparency and concerns about lowering the bar for schools serving low-achieving students. As will become clear, the NYC Progress Report uses a combination of status and growth measures, and benchmarks results against multiple populations in order to account for differences in student populations across schools.

When high stakes are attached to an accountability system, it is vital that it separate—to the extent possible—the unique impact of the *school* from pre-existing risk factors that also affect achievement (Clotfelter & Ladd, 1996; Raudenbush, 2004). A system that systematically advantages or disadvantages schools according to factors outside their control would be considered by most to be unfair. This concern is reflected prominently in the *Educator Guide to the NYC Progress Report* from the NYC Department of Education:

“The overall Progress Report Grade is designed to reflect each school’s contribution to student achievement, no matter where each child begins his or her journey . . . The methods are designed to be demographically neutral so that the final score for each school has as little correlation as possible with incoming student characteristics such as poverty, ethnicity, disabilities, and English learner status. To achieve this, the Progress Report *emphasizes year-to-year progress, compares schools mostly to peers matched based on incoming student characteristics, and awards additional credit based on exemplary progress with high-need student groups.*” (p. 1-2, emphasis added).

A recurring question in this paper is the extent to which the Progress Report does adequately account for differences in student populations across schools. The high school Progress Report contains mostly status measures (e.g., four-year graduation and Regents exam passing rates), and each is benchmarked against the performance of other schools. In an attempt to compare similar schools, the report uses a *peer index* to match schools to a group of peers based on student characteristics. A school’s overall score relies on a weighted average of *peer* and *citywide* comparisons in which peer comparisons are given the majority of the weight.

When schools vary in the populations they serve, scores benchmarked against the performance of peers should generally differ from those that ignore this variation. (This is the reason for making peer comparisons in the first place). In fact, our analysis finds the use of peer groups has only modest effects on the overall grade assigned to high schools in New York City. For example, in 2010-11 more than 2/3 would have received the same letter grade under a formula that ignored peers and simply compared the performance of schools citywide. We find the average number of points on the Progress Report changes only 7-10% (up or down) with the use of peer groups, relative to simple citywide comparison of schools that ignored peers.

This somewhat surprising result becomes clear when considering two features of the Progress Report that we describe in Sections 3 and 4. First, the index used to assign schools to peer groups relies almost exclusively on the mean ELA and math proficiency of incoming 8th graders. While this is the best indicator of the prior performance of a school’s students and a strong predictor of high school outcomes, it ignores other potentially important factors explaining school performance. As a result, peer groups are often quite dissimilar. Second, peers only affect a school’s Progress Report score by influencing the “peer range,” the *range* of outcomes observed among peer schools over the past four years. Because a school’s score depends on its

position in the peer range, one might believe its score would be highly sensitive to its selection of peers. In fact for most schools it is not, as peer groups are heterogeneous enough that a school's position among peers is often not much different from its position citywide.

What this means in practice is that the current use of peer groups has only modest effects on most schools' Progress Report score. This is particularly true for those in the middle of the peer index (where most schools are), but less so for those with very low or very high-achieving students. While making adjustments for very disadvantaged or advantaged schools, the current system does not go far to ensure others are contrasted with otherwise similar schools. The use of peers moderates the correlation between incoming student characteristics and scores, but as we show in Section 5, Progress Report scores remain associated with many existing risk factors such as poverty, 8th grade achievement, and the percent of students who are English language learners. We find these student and school characteristics jointly "explain" about 30% of the variation in the final Progress Report score. This finding echoes that of a recent study from the New York City Independent Budget Office (2012) that posed similar questions about the association between demographics and Progress Report scores.²

In the next section, we begin by introducing the main components of the high school Progress Report, and the "percent of range" calculation that is used to produce each subcomponent score. Sections 3 and 4 take a closer look at the construction of peer groups and the extent to which peer groups "matter" to the school's Progress Report score. (That is, how much they adjust a school's score relative to a simple citywide ranking). Section 5 examines how Progress Report scores and grades are related to other student and school characteristics, and Section 6 unpacks the complex "weighted Regents exam passing rate" measures, which differ in important ways from the others on the Progress Report. Finally, Section 7 concludes with a discussion and some recommendations for improving the design of the Progress Report. Our hope is that this report and its recommendations serve as the beginning of a thoughtful conversation about the goals of the Progress Report, and how the design of the report can best meet its objectives.

2. A Look at the Bottom Line: The Overall Progress Report Score³

Prior to 2011-12, the Progress Report consisted of three weighted categories, *Student Progress* (60 points), *Student Performance* (25 points), and *School Environment* (15 points), with an additional credit component called “closing the achievement gap.”⁴ In 2011-12, a fourth category, *College and Career Readiness*, was added. (In this report we rely on data from 2010-11). Each category consists of four to twelve subcategories, worth anywhere from 2.5 to 6.25 points (Table 2.1). *Student Progress*, for example, includes twelve separate subcategories representing credit accumulation and Regents Exam passing rates.

Table 2.1: High School Progress Report Subcategories, 2010-11

	Points
Student Progress:	60
Percent earning 10+ credits in Year 1	5
Percent earning 10+ credits in Year 2	5
Percent earning 10+ credits in Year 3	5
Percent earning 10+ credits in Year 1 (school’s lowest third)	5
Percent earning 10+ credits in Year 2 (school’s lowest third)	5
Percent earning 10+ credits in Year 3 (school’s lowest third)	5
Average Regents Exam passing rate	5
Weighted Regents Exam pass rate: English	5
Weighted Regents Exam pass rate: mathematics	5
Weighted Regents Exam pass rate: science	5
Weighted Regents Exam pass rate: global history	5
Weighted Regents Exam pass rate: U.S. history	5
Student Performance:	25
4-year graduation rate	6.25
6-year graduation rate	6.25
Weighted diploma rate (4-year)	6.25
Weighted diploma rate (6-year)	6.25
School Environment:	15
<u>School Survey scores:</u>	
Academic expectations	2.5
Communication	2.5
Engagement	2.5
Safety and Respect	2.5
Attendance rate	5.0
Overall score	100
Additional Credit (“closing the achievement gap”)	Up to 14

Notes: points were identical in 2007-08, 2008-09, and 2009-10 (Childress et al., 2011), but changed in 2011-12.

Most of the subcategories on the high school Progress Report are “status” rather than “growth” measures, as traditionally defined. In other words, they represent a snapshot of the achievement or attainment of a school’s students at a fixed point in time, not the *change* in achievement from a prior period, as student growth on an English Language Arts exam from 3rd to 4th grade might indicate. The *Student Progress* subcategories capture progression through the

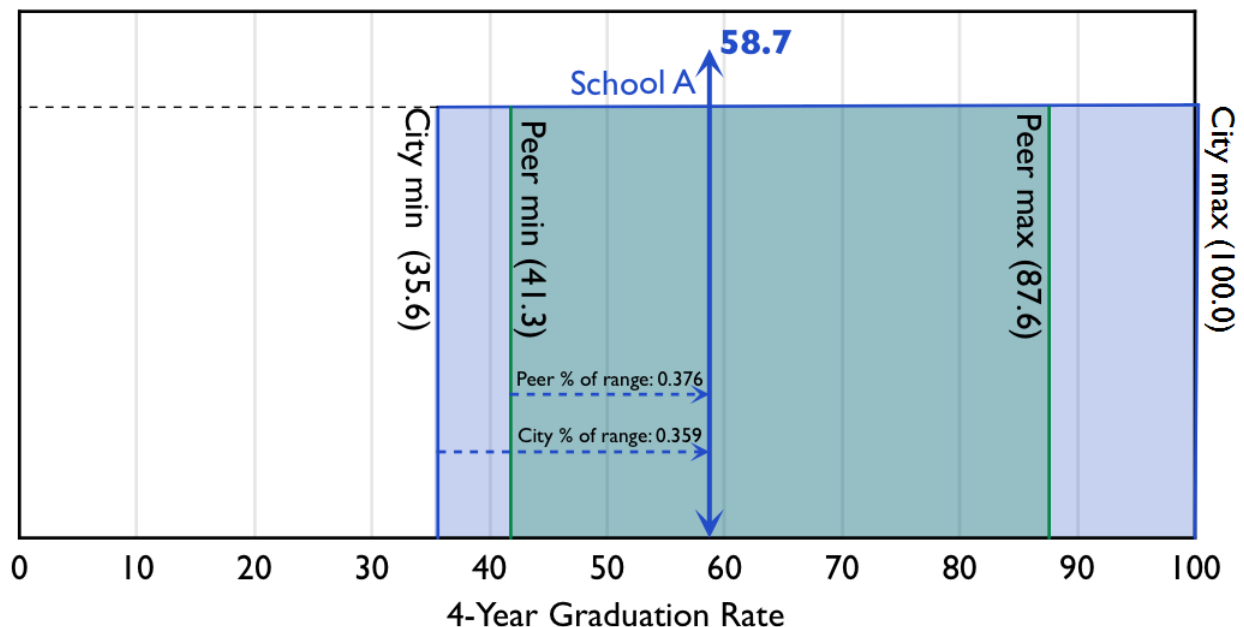
requirements of high school, but do not account for differences in starting points by measuring year-to-year growth. The Weighted Regents pass rates account for differences in starting points by measuring how well students perform relative to expectations (see Section 6), and thus they are closer in spirit to a “growth” measure. This distinction is unimportant for the discussion that follows, but is worth pointing out as one of the major differences between the high school and elementary/middle Progress Reports (and some other accountability tools used elsewhere).

Letter grades are assigned to schools based on the total score, inclusive of additional credit. Since 2009-10, schools earning 70 or more points receive an “A,” while 58 – 69.9 points receive a “B,” 47 – 57.9 points receive a “C,” 40 – 46.9 points receive a “D,” and 39.9 or fewer points receive an “F.” Cut points were initially set by the NYCDOE Office of Accountability with a particular “curve,” or distribution, of grades in mind.

It is important to note that each subcategory score depends on how a school performs relative to the range of outcomes observed in its peer group and citywide, and thus each are norm-referenced measures. Performance relative to peers is weighted at 75% (the “peer percent of range”), while performance relative to schools citywide is weighted at 25% (the “city percent of range”). The heavy weight on peers reflects the NYCDOE’s desire to compare outcomes in similar schools.

As a simple example, suppose School A had a 58.7% 4-year graduation rate in 2010-10. In this subcategory School A can receive up to 6.25 points. The higher is its 4-year graduation rate in the peer and city ranges, the more of these points it earns, and the lower is its graduation rate in these ranges the fewer points it earns. To determine its subcategory score, we need to know how School A performed relative to its peer and city range, as pictured in Figure 2.1.

Figure 2.1: School A’s Peer and Citywide Range: 4-Year Graduation Rate



The peer range for 4-year graduation rates is determined using all 4-year graduation rates observed in peer schools during the prior four years (2007 through 2010).⁵ Rather than use the actual minimum and maximum of these values, however, the peer minimum and peer maximum are calculated as two standard deviations below and above the peer mean, respectively. This limits the effect of outliers on the peer range, but at the same time introduces the possibility that the minimum and maximum extend beyond what is actually observed in the data.⁶

Between 2007 and 2010 School A’s peer range extended from **41.3%** (the peer minimum) to **87.6%** (the peer maximum). School A’s peer percent of range is its graduation rate (58.7%) as a percent of the total distance from the peer minimum to the peer maximum (**37.6%**).⁷ An identical procedure is used to find the city range and city percent of range, with the only difference being the city range uses all 4-year graduation rates observed citywide. School A’s city percent of range for 2010-11 was **35.9%**. Both are pictured in Figure 2.1.

The two “percent of range” values are weighted according to the 75% / 25% peer and city weights as shown below, and applied to the 6.25 possible subcategory points. In 2010-11, School A earned **2.32** of the 6.25 points:

((0.376 x 75%) + (0.359 x 25%)) x 6.25 = 2.32			
Peer Percent of Range x 75%	City Percent of Range x 25%	Points possible	Points Earned

It is instructive to point out that the only way a school could earn all of the total possible points in this subcategory is to match the *maximum* 4-year graduation rate attained among all schools citywide in the past four years (in this example, 100%).

Although this example focuses on 4-year graduation rates, the same procedure is used to determine points earned in all other *Student Progress*, *Student Performance*, and *School Environment* subcategories. It is clear, then, the peer and city range and “percent of range” calculations are the keys to understanding a school’s Progress Report score.

Given their disproportionate weight (75% of each subcategory score) the peer percent of range has the greatest influence on a school’s Progress Report result. However, as we show in Section 4, peer groups are only important to the extent they move the peer range away from the city range (usually, making it narrower). If the peer range differs little from the city range, a school would receive approximately the same score whether the peer or citywide comparison is made. This would not be of concern if schools are relatively homogeneous—making peer comparisons less necessary—but could be problematic if the peer groups are not adequately grouping similar schools. Incidentally, in the above example there was only a modest difference between the peer and city percent of ranges (37.6 and 35.9).

In the next section, we describe how Progress Report peer groups are determined, and the dimensions on which these groups do and do not represent similar schools.

3. Forming Peer Groups: The Peer Index

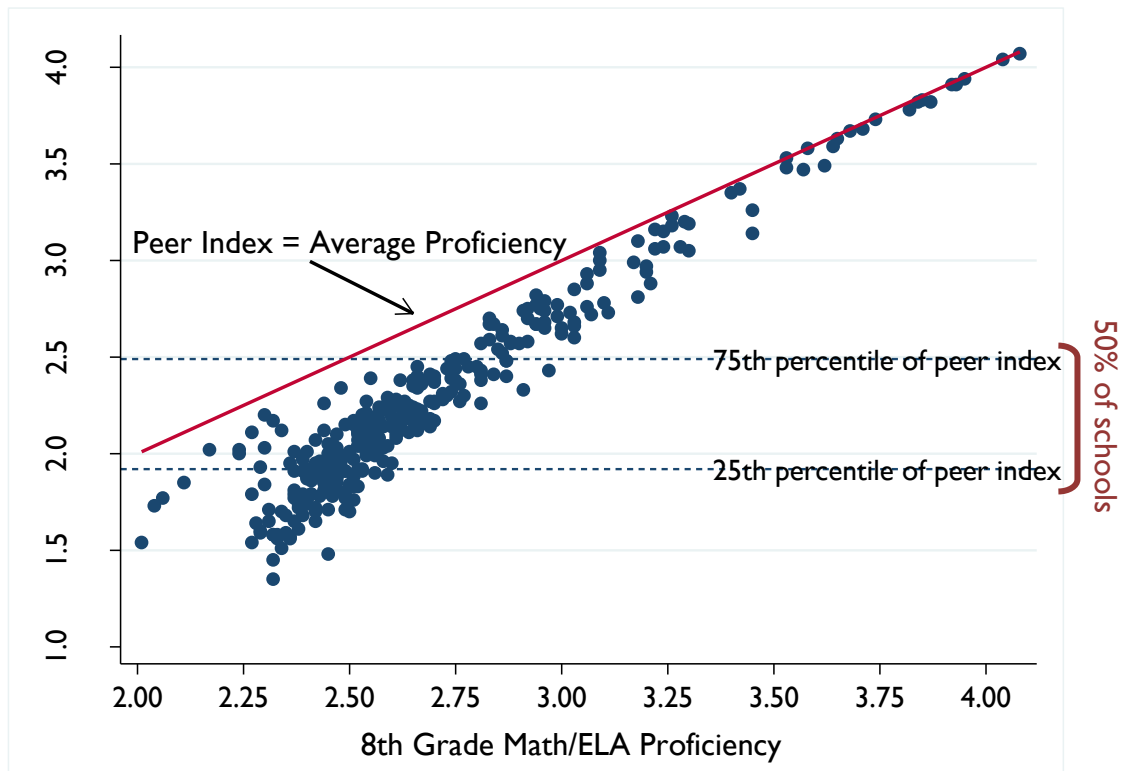
As noted in Section 2, a school's peer group is potentially quite important to its Progress Report score. This section describes how a school's peer group is determined. Schools are assigned peers based on a peer index, calculated for every school and ranging from 1.00 to 4.50, with lower values representing more academically disadvantaged schools, and higher values representing more advantaged schools.⁸ Effectively, the index is the average math and ELA proficiency of a school's incoming 8th graders, with a few adjustments for special populations as illustrated below for example School A:⁹

	<u>School A</u>
Mean 8th grade proficiency	2.44
- 2 × (% of students with disabilities/100)	- 2 × (0.246)
- 2 × (% of students in self-contained classes/100)	- 2 × (0.021)
- (% of students overage at entry/100)	- 0.096
Peer Index	1.81

The three adjustments for student characteristics can only lower a school's peer index from its average proficiency level. Figure 3.1 shows each school's average proficiency (on the horizontal axis) and its corresponding peer index (on the vertical axis). Each school is a point, and the diagonal represents cases in which the proficiency score and peer index are identical. The peer index is always less than or equal to average proficiency; the only difference between the two is the adjustment for special populations. (In School A's case, its relatively high percentage of students with disabilities reduces its index from **2.44** to **1.81**).

As might be expected, average proficiency and the peer index are most similar in schools with high levels of incoming proficiency (towards the right of Figure 3.1). These schools tend to have the fewest overage students and students with disabilities. Schools with low levels of proficiency (towards the left of Figure 3.1) tend to have more of these students and thus receive greater adjustments. For the average school, the peer index is roughly 0.41 points lower than average proficiency as a result of the disability and overage adjustments. This is a fairly large adjustment, relative to the overall standard deviation of proficiency scores across schools (0.37).

Figure 3.1: Relationship Between Peer Index and Average 8th Grade Proficiency



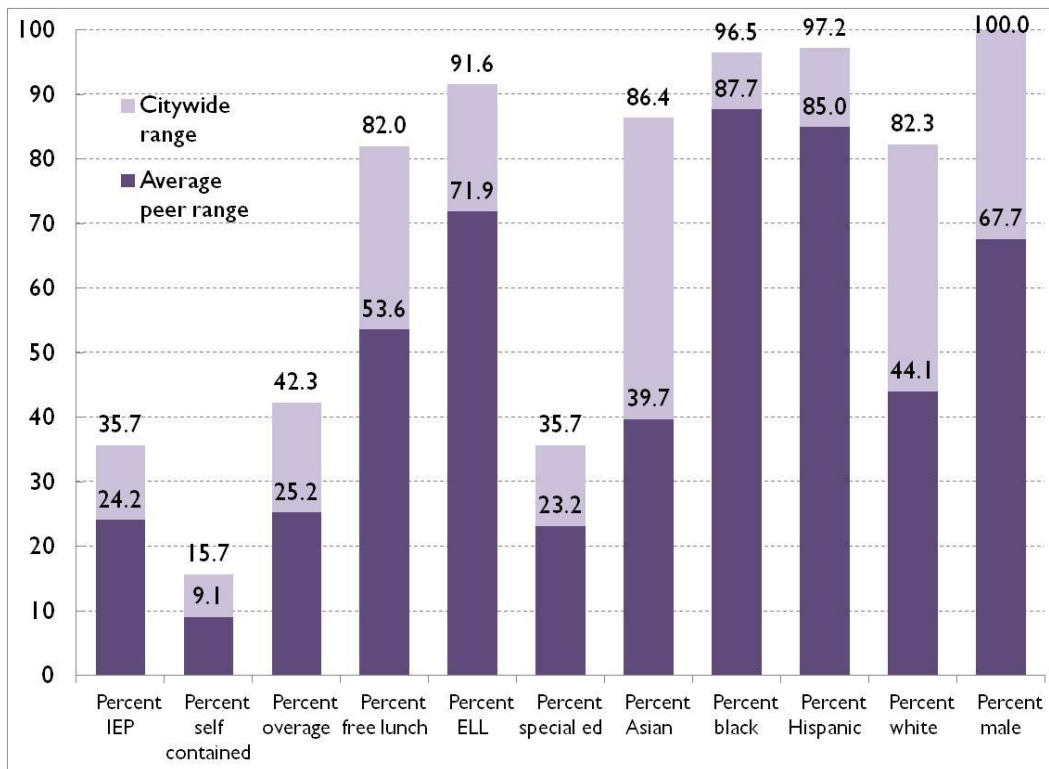
Note: each point represents a high school in 2010-11. The diagonal shows cases in which the peer index and average proficiency level are identical. The dashed lines indicate the 75th and 25th percentiles of the peer index; half of the city’s high schools fall between these values. Only the 332 schools in 2010-11 with an overall Progress Report score are included in this figure. New schools with insufficient data and schools slated for closure may be assigned a peer index but not have a Progress Report score.

Figure 3.1 also indicates the 75th and 25th percentiles of the peer index (1.9 and 2.5). Half of the city’s high schools fall within this relatively narrow band (the mass of points between the two dotted lines). A long tail of high peer index schools extends above 2.5, while a more concentrated group of low peer index schools falls below 1.9. School A’s relatively low peer index of 1.81 would place it at roughly the 19th percentile of the peer index citywide.

To assign peer groups, schools are sorted from lowest to highest peer index value and the 20 schools just above and 20 schools just below a given school constitute its peers.¹⁰ Because the peer index is based almost exclusively on the average proficiency of incoming students (with an adjustment for disabled and overage students), it ignores other potentially important factors associated with student performance. To get a sense of the extent of heterogeneity within peer groups for 2010-11, we summarized student and school characteristics for every school’s peer group, describing the peer group of the average, or “typical” school. As a measure of diversity within peer groups we focused on the *range* of school characteristics within peers. While other measures of variability—such as the standard deviation—are usually preferable, the range is appropriate in this case, given that the range of outcomes is used to benchmark scores.

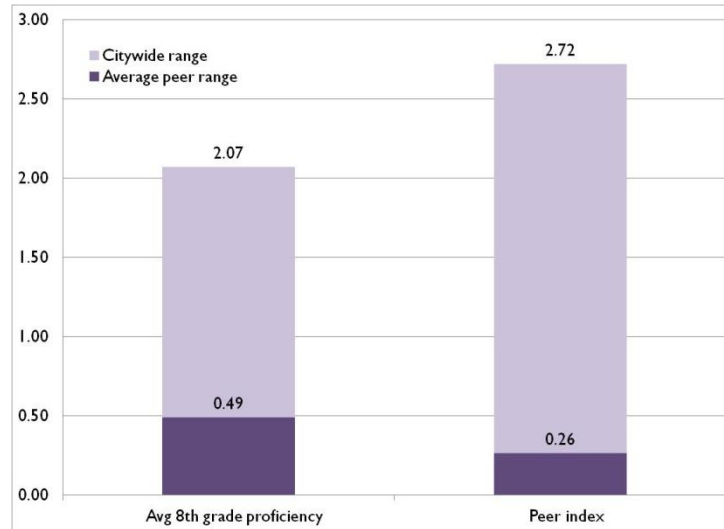
Figure 3.2 illustrates the typical (average) range of selected school characteristics for peer groups in 2010-11.¹¹ Each average is shown against the range of the same characteristic citywide (i.e., the largest possible range a peer group could have). As an example, the average school's peer group exhibited a range of 53.6 points in the percent of students eligible for free lunch, a wide range but narrower than the 82-point range observed citywide. By most measures, peer groups are very diverse. For example, the average school's peer group exhibited a range of 71.9 points in the percent of students who are English language learners (ELL), 25.2 points in the percent of students who are overage, and 85-88 points in the percent of students who are black or Hispanic. In many cases, these ranges are not much smaller than the range exhibited citywide.

Figure 3.2: Range of Selected School Characteristics for the Average Peer Group and City



By design, the range of peer indices and incoming student proficiency is much narrower for the average peer group, as shown in Figure 3.3. Whereas the citywide range in mean 8th grade proficiency was 2.07 in 2010-11, the average peer group exhibited a range of only 0.49. Because the peer index itself is used as the grouping variable, the average peer group had a range of peer indices that was much narrower, at 0.26.

Figure 3.3: Range of Mean 8th Grade Proficiency and Peer Index for the Average Peer Group and City



A look at other high school characteristics, such as location, size, and admissions method, reveals similar heterogeneity within peer groups. For example, the average school's peers include both small and large high schools, with an enrollment range of 3,272 students. (By comparison, the range between smallest and largest high school citywide is 5,041). What this means is that it is not unusual for a school to have, for example, one peer with an enrollment of 250 students and another with an enrollment of 3,500.¹² We also find less than a third of a school's peers are located in the same borough, on average.¹³ (While in principle there is no reason one cannot compare schools across boroughs, in some cases this may involve schools in very different contexts).

NYC high schools also vary in the extent in which they can screen their incoming students. *Screened* and *audition* schools rank applicants according to their own criteria, such as grades, attendance, or a portfolio of work, while *unscreened* and *zoned* schools have few requirements for admission. *Limited unscreened* schools give priority to students who attend an information session, and often have residential priorities. *Educational option* schools are a hybrid of screened and unscreened, and attempt a balance of low-, middle-, and high-achieving students.¹⁴

For the average school, we find 38% of its peers (about 15 of the 40 schools) use the same selection method, although this ranges from a low of 9.6% for audition schools to a high of 54% for limited unscreened schools. Most school types have both screened and unscreened peers. For the average audition or screened school, about half (47 – 53%) of peers are also audition or screened schools. But even limited unscreened and educational option schools have an average of 1 in 5 peers that screen their incoming students. By definition, schools in the same peer group are comparable with respect to their peer index. However, to the extent students in

screened and unscreened schools differ in ways not captured by the peer index, it may be inappropriate to attribute differences in their outcomes to the quality of the school.

These findings are merely a description of the typical peer group, and do not demonstrate that that peer groups used in the Progress Report are “excessively dissimilar.” Put another way, it is not *a priori* clear “how similar” peer groups would have to be along any of these dimensions before they would be considered appropriate. Peer groups by definition are similar with respect to the peer index, largely the average proficiency of incoming students. Whether or not the remaining heterogeneity in peer groups is important depends on how these factors relate to achievement, aside from their association with 8th grade proficiency, and how they affect the width of the peer range, which is ultimately how peers influence a school’s scores.

What these results do illustrate is that the typical peer group is a highly diverse set of schools, differing in size, location, admissions method, poverty rates, and proportions of students with special educational needs. This fact plays an important role in the next section, which investigates the extent to which peer groups influence a school’s Progress Report score.

As a final point, the fact that the peer index relies almost entirely on average 8th grade proficiency begs the question of how missing information about 8th grade test scores affects Progress Report scores.¹⁵ In the average high school, about 17-19% of students are missing 8th grade achievement scores, and thus are not reflected in the peer index. About 1 in 10 schools have 30% or more students without 8th grade scores. They may be new to the district or were otherwise absent or excluded from the test.

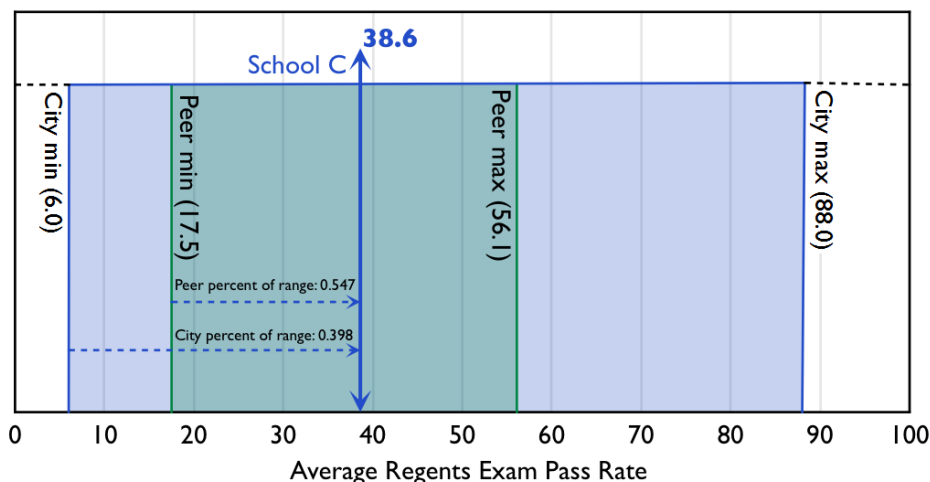
How missing test score information affects a school’s score is a difficult question that depends on how students with missing data compare to those with test scores. We examined the relationship between Progress Report scores and the percent of students missing 8th grade test scores and found that, holding the peer index constant, schools with more missing scores tended to perform *better* in most subcategories, suggesting that the peer index understates, on average, the incoming student achievement in schools where more data is missing. This need not hold for all schools, however; in some schools with missing test score data the peer index likely overstates student achievement in that school. A deeper discussion of this issue is provided in Supplementary Appendix B.

4. How Much Do Peer Groups “Matter” in Practice?

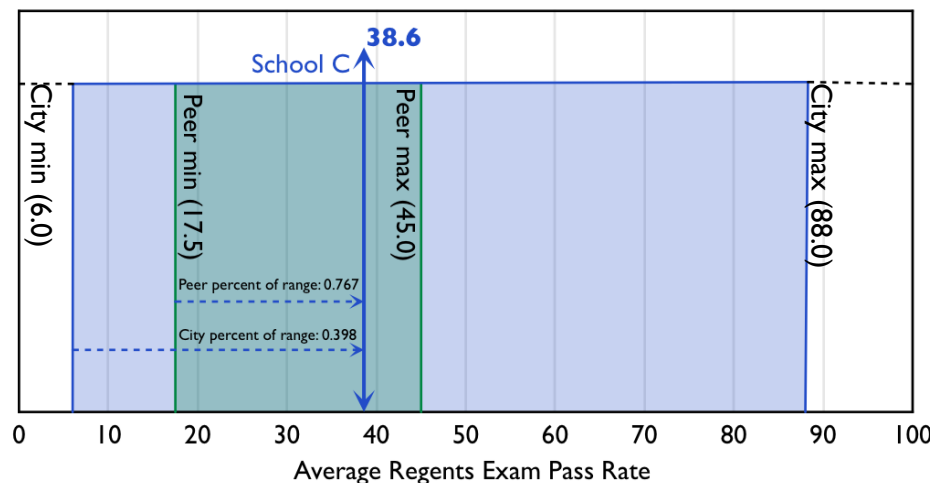
The Progress Report emphasizes peer comparisons in order to fairly contrast the performance of schools serving similar populations. Our analysis in Section 3, however, suggested the peer index primarily differentiates schools based on incoming students’ ELA and math proficiency. Along other dimensions, the typical peer group represents a more dissimilar set of schools. It is logical to ask, then, how these peer groups influence Progress Report scores, and how well these groups accomplish their goal of contrasting similar schools.

To understand how peer groups affect Progress Report scores, it is important to recognize the following: *peer groups “matter” to the extent they influence a school’s peer range and the school’s position in that range.* For example, a school would fare better with peer group A than peer group B if its position in A’s range were higher than its position in B’s range. Similarly, a school would fare better with peer group A than in a simple *citywide* comparison (that ignored peers) if its position in A’s range were higher than its position in the *citywide* range.

Figure 4.1: Peer Versus Citywide Comparisons



(a) School C fares better in a peer comparison (0.547) than in a citywide comparison (0.398)



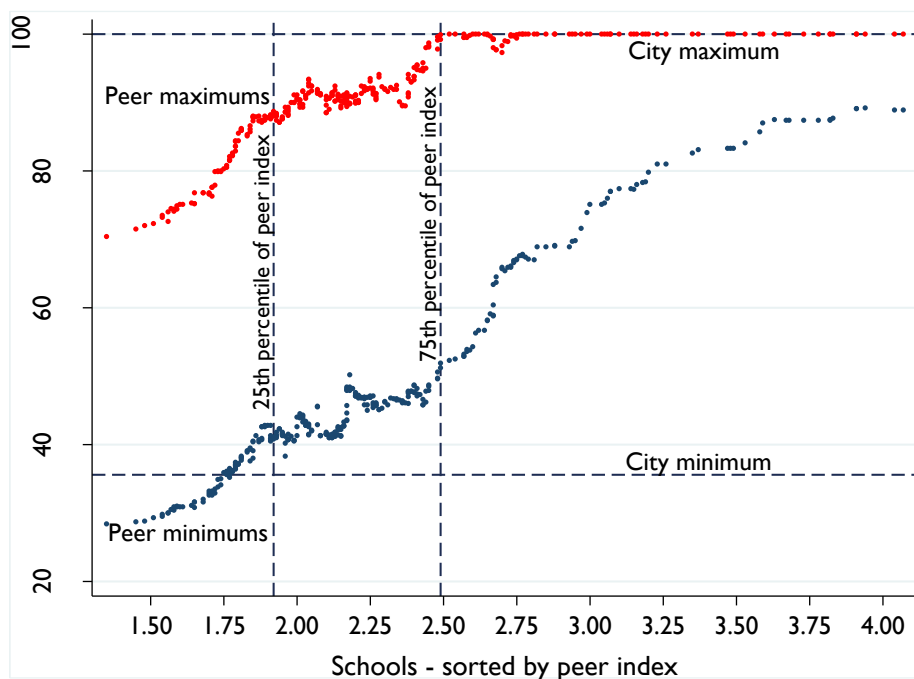
(b) School C fares even better in a peer group with a lower maximum (0.767).

This is illustrated in panel (a) of Figure 4.1, which shows School C's peer and citywide range for Average Regents Exam passing rates. In 2010-11, pass rates citywide ranged from 6% to 88%. School C's passing rate was 38.6%, which was 39.8% of the distance between the city minimum and maximum (its city percent of range.) Its peers exhibited a more restricted range of passing rates, from 17.5% to 56.1%. School C's pass rate was 54.7% of the distance between the peer minimum and maximum (its peer percent of range.)

School C fares better in a peer, rather than citywide, comparison because its peer group exhibited a narrower range of passing rates (a higher min and a lower max), and because C scored comparably higher in this range versus the city range (the max dropped by much more than the min rose).¹⁶ This is effectively how peer groups "adjust" for differences in school characteristics. Panel (b) of Figure 4.1 shows a case in which School C has the same peer minimum, but a lower peer maximum; School C fares even better with this peer group.

These examples are cases in which the peer percent of range is higher than the city percent of range, but for other schools and subcategories this will be reversed; it depends on how the peer and city ranges differ. Figure 4.2 shows how peer ranges compare with the city range for 4-year graduation rates for all schools in the city. For a school with a given peer index (on the horizontal axis) the two points on the vertical dimension are the upper and lower bounds of its peer range. A comparison of peer ranges with the city range suggests schools near the top of the peer index (e.g., above the 75th percentile) are likely to have a lower peer percent of range than city percent of range, while those near the bottom (e.g., below the 25th percentile) are likely to have a higher peer percent of range than city percent of range.

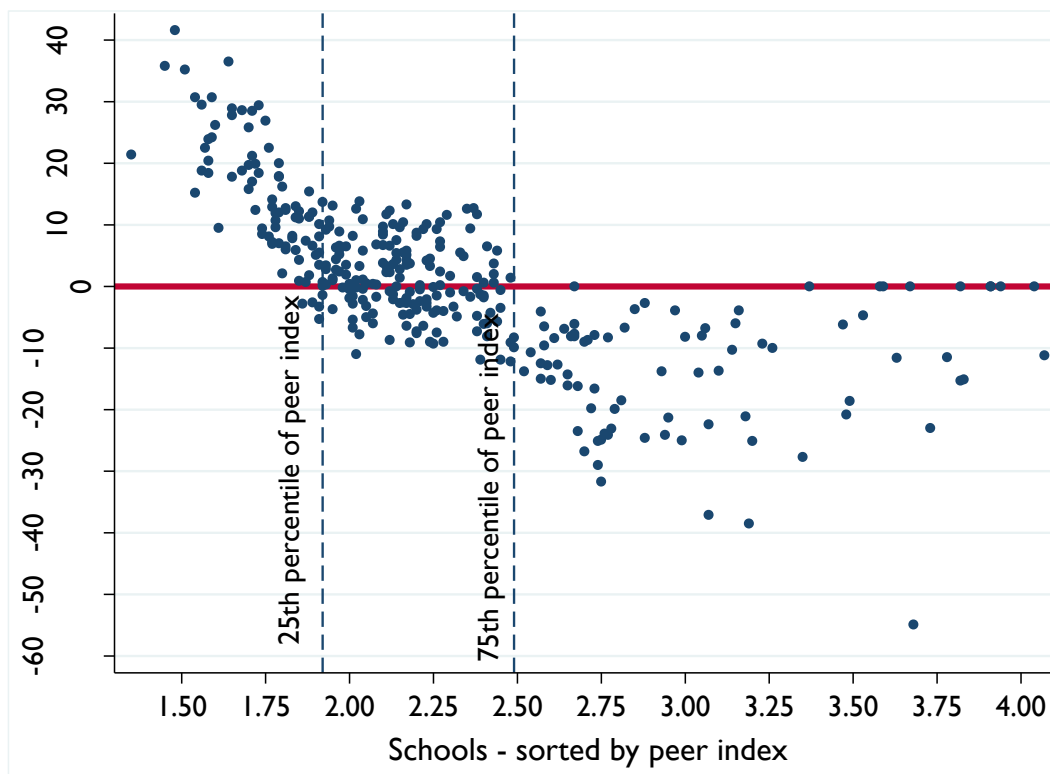
Figure 4.2: Peer and City Maximum and Minimum, 4-Year Graduation Rates



In this example, schools near the bottom are “curved up” by the peer comparison, relative to a simple citywide comparison, because their peer maximum is much lower than the city maximum (and their peer minimum is the same or lower than the city minimum).¹⁷ Those near the top are “curved down” relative to a simple citywide comparison, because their peer minimum is much higher than the city minimum (and their maximum is the same as the city). It is less obvious what happens to the schools in the middle, where the peer range is modestly narrower than the city range.

While Figure 4.2 shows each school’s peer range, Figure 4.3 shows how each school’s *peer percent of range* differs from its *city percent of range*. (Recall that a weighted average of the two determines the school’s score). For a school with a given peer index (on the horizontal axis), the difference between the peer and city percent of range is shown as a point on the vertical dimension. Differences above zero indicate that a school fared better under the peer comparison versus the city comparison, and values below zero indicate the school fared worse under the peer comparison.

Figure 4.3: Difference Between Peer and City Score (Percent of Range), 4-Year Graduation Rates



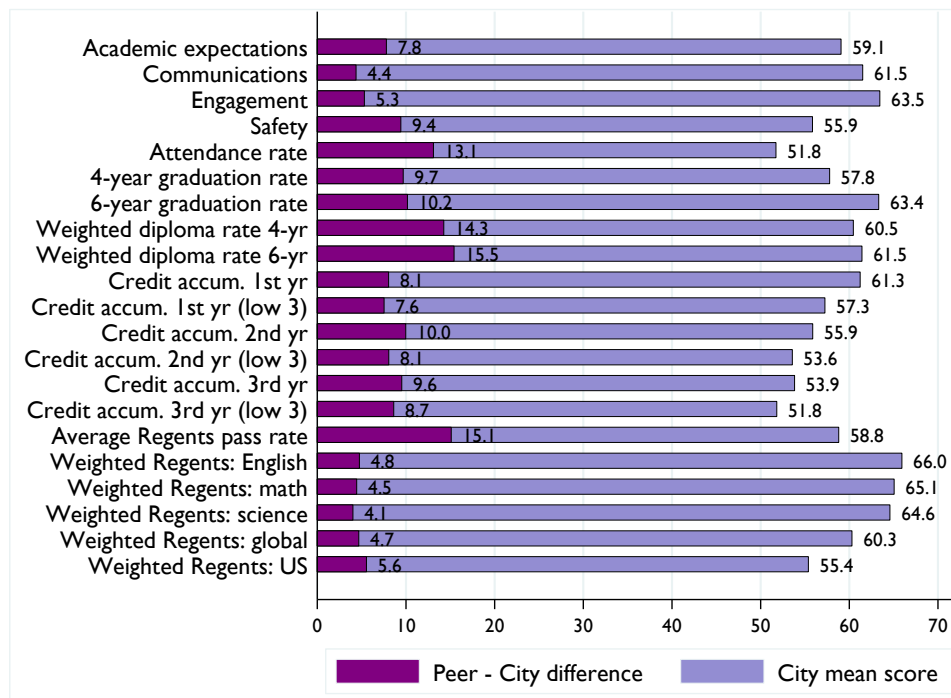
As suggested by Figure 4.2, academically disadvantaged schools tend to be “curved up” by the peer comparison (relative to the city) while more advantaged schools tend to be “curved down” by the peer comparison. It is also worth noting that schools in the middle generally see modest differences between their peer and city percent of range. For the average school

between the 25th and 75th percentile of the peer index, the difference was about +/- 5.0 on a baseline city percent of range of 54. The median difference was about +/- 4.3.

The example in Figures 4.2 and 4.3 were based on the 4-year graduation rate. But the Progress Report consists of 21 different subcategories. Motivated by the above examples, we can see how much peer groups “matter” for each subcategory by examining the extent to which its peer and city percent of range differ. If they differ little, then peer comparisons offer little more than what a simple citywide comparison would. If the difference is more substantial, then peer groups are more influential. Before proceeding, it is important to emphasize that neither measure is necessarily the “right” one. A simple citywide comparison is generally undesirable, as it fails to account for differences in student populations. At the same time, a peer comparison that has large effects on a school’s ranking is not *a fortiori* a good one. For now we are only interested in how different the current peer and city percent of range are.

Figure 4.4 summarizes the differences between peer and city percent of range, by subcategory.¹⁸ In this figure, the long bar is the average *city* percent of range in the indicated subcategory in 2010-11, used as a reference. The short bar is the average *difference* between a school’s peer and city percent of range in the indicated subcategory. For example, in Academic Expectations the average school received a 7.8 point higher (or lower) peer percent of range than their city percent of range. To assess whether this is a meaningful difference, we can compare this to the average city percent of range for Academic Expectations of 59.1.¹⁹

Figure 4.4: Mean City Percent of Range and Peer-City Difference, by Subcategory



It is apparent that in some subcategories peer comparisons generate quite different scores, on average, than a citywide comparison would. These include, for example, the Average Regents

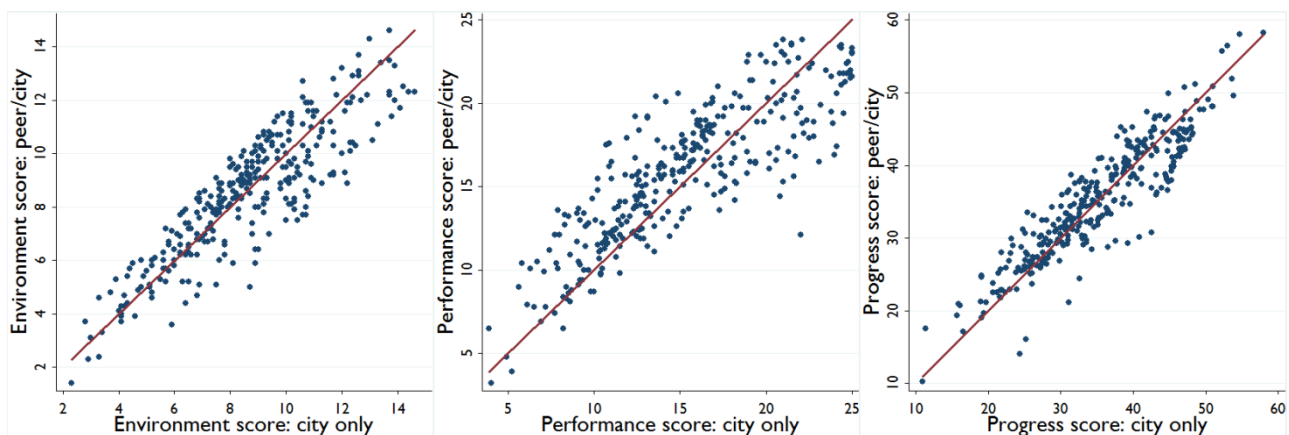
Pass Rate, Attendance Rate, and Weighted Diploma Rates (4 and 6 year). In other categories, the peer group has a small effect on the score relative to a citywide comparison. These include the Weighted Regents Pass Rates (math, science, global history, and English)—which are examined in greater detail in Section 6—and Communication measures.²⁰

At first glance it is surprising that peer comparisons sometimes differ little from citywide comparisons that do not use peer groups at all. But the explanation follows from the key point made earlier: peers only matter to the extent they influence the peer range. And as it turns out, peer ranges are often not substantially different from the citywide range. We compared the reported peer and city ranges for all schools and subcategories and found that in some categories—such as Academic Expectations, Communication, Engagement, Safety, and 1st year credit accumulation—the peer and city ranges were frequently quite similar.²¹ As a result, the *percent of range* tend to be similar in these categories. In others, such as the Average Regents pass rates, attendance rate, and Weighted Diploma rates, the peer and city ranges tend to be less similar, and thus the percent of range tend to be least similar in these categories.

How do peer groups affect overall Progress Report scores? This depends on their net effect across each of the above subcategories and the weight each subcategory receives in the final score. In some subcategories, a school may receive a higher peer percent of range than they would in a citywide comparison, while in other subcategories the school may receive the same or lower percent of range.

We calculated the total number of points each school would receive under a simple citywide comparison, ignoring peers, and compared these with the points the school received using the 75%/25% peer/city weights. The results are shown in Figure 4.6, broken out by the three main categories (*Environment*, *Performance*, and *Progress*). Each point in these figures is a high school; the horizontal axis shows the total points received under a pure citywide comparison while the vertical axis shows the peer-city weighted score. The diagonal represents cases in which the scores are the same.

Figure 4.6: Reported Category Scores versus Scores under a Pure Citywide System

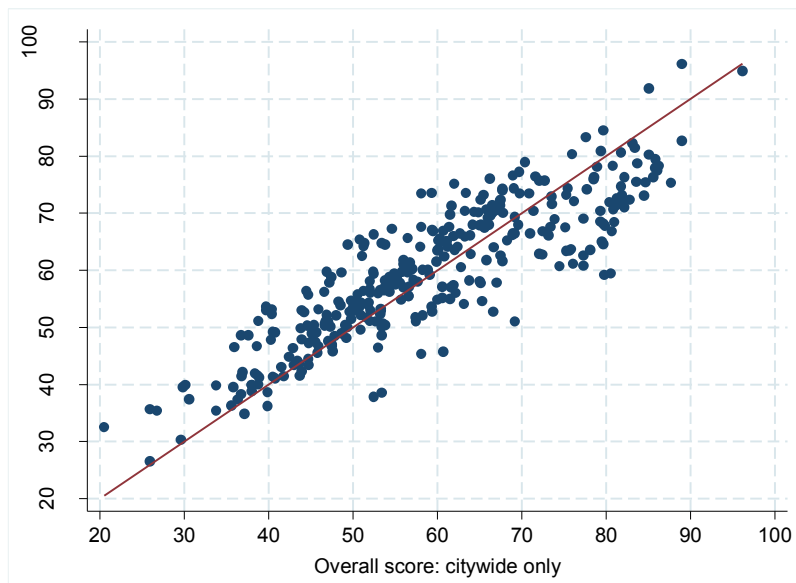


What stands out in this figure is the strong consistency in scores. Most schools lie close to the diagonal, meaning their scores differ little from what they would be in a pure citywide comparison that ignored peers altogether. This is most true for the Progress category (where the correlation is 0.91) and least true for Performance (where the correlation is 0.84).²²

Not only is the Progress category seemingly the least affected by peer group comparisons, it is also the most heavily weighted on the Progress Report. Based on this, one might predict that the final score would be minimally affected by peer groups. In fact this is the case, as shown in Figure 4.7. This figure is comparable to those in Figure 4.6, but represents the final Progress Report score, ignoring “additional credit.” (For now we ignore additional credit, as it is largely “off formula”).

Again there are strong similarities between the actual Progress Report score—which combines peer and citywide results—and the score that would result from a simple citywide comparison. The difference in scores for the average school is +/-5.7 points; the difference for the median school is +/-4.6 points. (The mean is higher due to a small number of schools with large differences). These differentials are small, on a baseline average citywide score of 59 points. Overall, the correlation between the two scores calculated using different methods is 0.88 (where 1.0 is a perfect positive correlation).

Figure 4.7: Actual Progress Report Score vs. Score Under a Simple Citywide Comparison



Note: excludes “additional credit” points.

These results hint that the letter grade on the Progress Report is unlikely to differ much from the grade that would result from a citywide comparison alone. To look at this, we added the 2010-11 additional credit equally to each score in Figure 4.7 (citywide only, and weighted

peer/city) and then applied the NYCDOE cut points for grades.²³ Table 4.1 provides the results, where the rows indicate the grade a school would have received in a citywide system and the columns indicate the letter grade the school actually received.

Reading across the rows, **77%** of schools who would have received an “A” in a pure citywide comparison received an “A” on the actual Progress Report; 22% received a “B”; less than 1% received a “C”, and none received a “D” or “F.” Of schools who would have received a “B” in a pure citywide system, **61.3%** received a “B” on the actual Progress Report, 23.8% received an “A,” and 13.8% received a “C.” Most “D” and “F” schools (**53.3%** and **48.2%**) under a citywide comparison received the same grade on the actual report, although the use of peer groups tended to bump “D” and “F” schools to the next higher grade.

Table 4.1: Actual Progress Report Grade vs. Grade under a Pure Citywide System

		Actual letter grade on Progress Report (row percentages)					# of schools	% of total
		A	B	C	D	F		
Letter grade based on pure citywide comparison	A	76.9%	22.2%	0.9%	0.0%	0.0%	117	35.2%
	B	23.8	61.3	13.8	1.3	0.0	80	24.1
	C	1.3	38.5	56.4	2.6	1.3	78	23.5
	D	0.0	0.0	43.3	53.3	3.3	30	9.0
	F	0.0	0.0	18.5	33.3	48.2	27	8.1
# of schools		110	105	74	28	15	332	100.0
% of schools		33.1%	31.6%	22.3%	8.4%	4.5%	100.0	
% of students		30.5%	31.9%	22.0%	8.6%	6.9%		

Notes: Includes only schools with an overall score reported in 2010-11. “Additional credit” points are added equally to the “pure citywide” score and original formula.

Together, 212 of 332 schools, or about 2/3 (63.9%) received the same letter grade they would have received under a simple citywide comparison. 77 schools (23.2%) received a higher grade than they would have under a citywide comparison, and 43 schools (13.0%) received a lower grade. As emphasized earlier, our analysis does not tell us whether one letter grade is more “correct” than the other. Rather, what is worth noting is the high level of consistency between the existing system and one in which peer groups are not used at all.

An accountability system that largely reflects a school’s position in the citywide distribution of outcomes is at risk for not meaningfully separating the school’s contribution to student achievement from existing risk factors. In the next section, we look explicitly at the Progress Report’s ability to mitigate the correlation between its scores and student risk factors.

5. How do Progress Report Scores Vary with Student and School Characteristics?

If an accountability system is separating the unique impact of the school from outside factors related to achievement, then it follows its school ratings should not vary systematically with pre-existing student or school characteristics associated with outcomes. This is also an explicit goal of the Progress Report, as noted in the NYCDOE *Educator's Guide* ("The methods are designed to be demographically neutral so that the final score for each school has as little correlation as possible with incoming student characteristics," see Section 1).

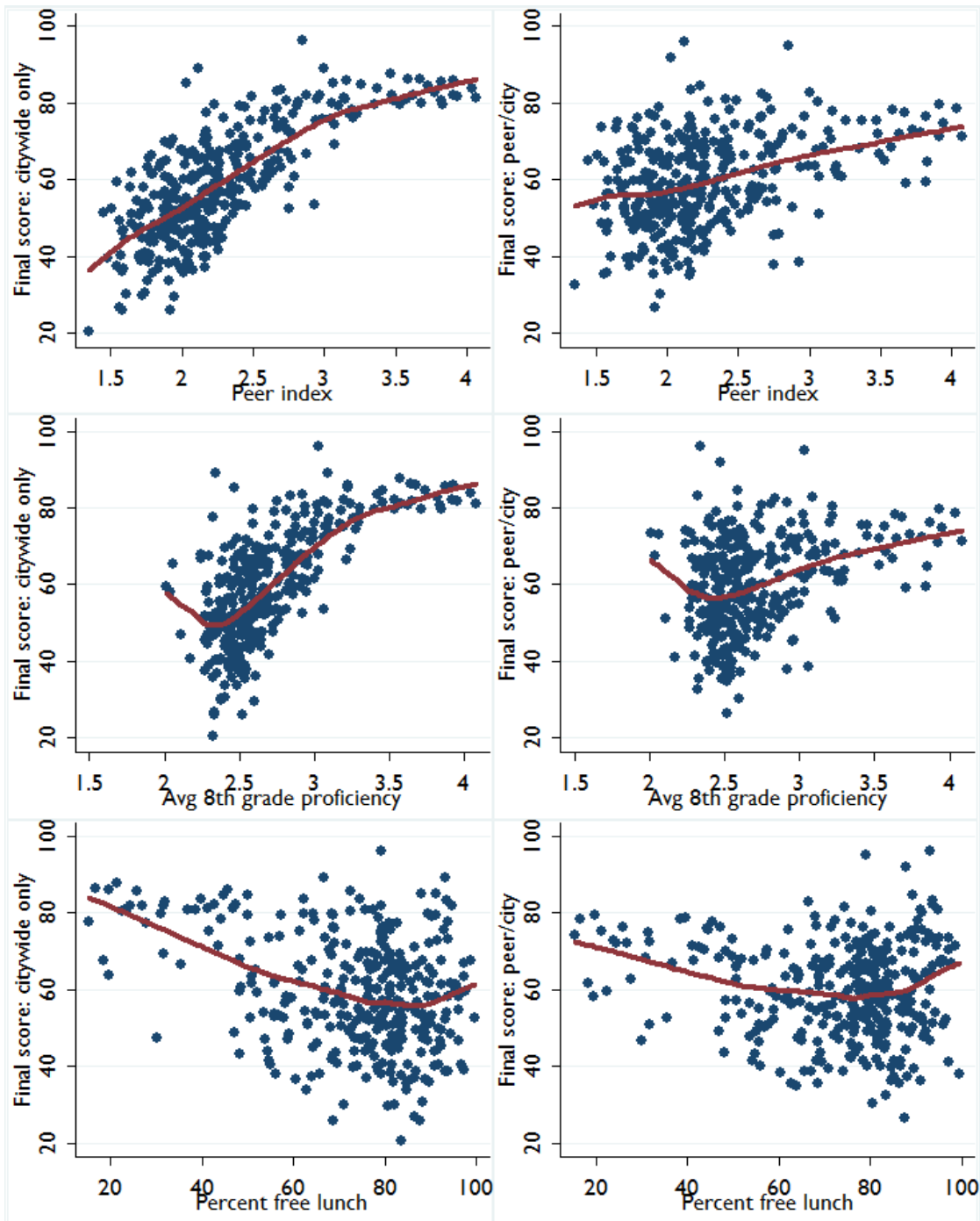
In this section we briefly look at how final Progress Report scores are related to selected student and school characteristics under both a simple citywide comparison (as in Section 4) and the official peer/city formula. The expectation is that the use of peer groups will noticeably weaken the correlation between student characteristics and final scores. Figure 5.1a and 5.1b show these relationships visually. In these pictures, each dot is a school; the horizontal axis provides the level of one school characteristic (e.g., percent eligible for free lunch), while the vertical axis provides the school's Progress Report score. For comparison, the graphs on the left use the citywide-only score and the graphs on the right use the weighted peer/city score. For now we exclude additional credit from both scores, since these are largely "off formula" adjustments. We add in additional credit in Table 5.1.

In these figures, a strong (linear) correlation would appear as points clustered around a line, while a lack of correlation would appear as a non-directional "cloud" of points. (A relationship can also be non-linear, as the fitted lines in these figures indicate). While the correlation with student characteristics weakens with the use of peer groups, in some cases the correlation remains moderate. Table 5.1 reports correlation coefficients with each school characteristic.²⁴ These values range between -1 and +1, from a perfect negative correlation to a perfect positive correlation. Values close to zero indicate weak or no correlation.

Excluding additional credit, there remains a modest correlation between Progress Report scores and average 8th grade proficiency, the percent of students in special education, the racial/ethnic composition of the school (Asian, white, black, or Hispanic), and to a lesser extent, school size. However, each of these correlations is—as intended—lower than they would be if peers were ignored and only citywide comparisons were made. Still, when considering all of these characteristics together, they "explain" about 30% of the variation in the actual Progress Report score, or 22% when additional credit is included. (This is the multiple regression R-squared reported in Table 5.1).

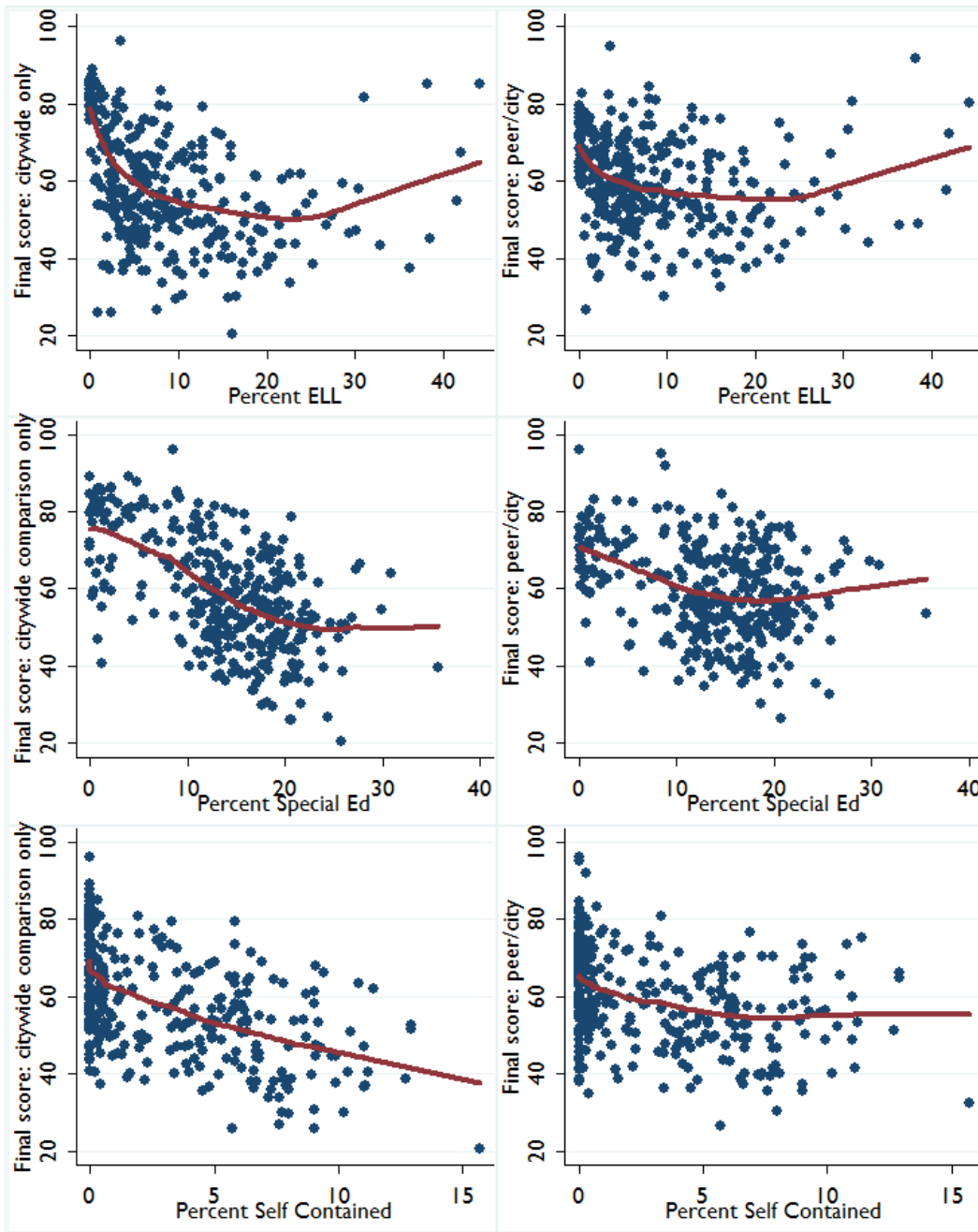
We note that the remaining correlation between Progress Report scores and student characteristics is not simply due to the 25% weight on citywide comparisons (which do tend to be strongly correlated with student risk factors). The second column of Table 5.1 shows the correlation between overall scores relying only on peer comparisons and student and school characteristics. Many of the modest correlations seen using the weighted peer/city scores (such as those with 8th grade proficiency, percent in special education, and racial composition) are also apparent in the peer-only scores.

Figure 5.1a: Relationship between Overall Progress Report Score and Student Characteristics (Peer Index, Average 8th Grade Proficiency, and Percent Eligible for Free Lunch): Citywide Only and Combined City/Peer Score



Notes: includes only the 332 high schools with an overall score reported in 2010-11. "Additional credit" points are excluded from the overall score in these figures. Solid lines are best-fit lowess lines.

Figure 5.1b: Relationship between Overall Progress Report Score and Student Characteristics (ELL, Special Education, and Self-Contained Special Education): Citywide Only and Combined City/Peer Score



Notes: includes only the 332 high schools with an overall score reported in 2010-11. "Additional credit" points are excluded from the overall score in these figures. Solid lines are best-fit lowess lines.

Table 5.1: Correlation between Overall Scores and Student/School Characteristics

	Overall Score: City Only	Overall Score: Peer Only	Overall Score: City & Peer	Overall Score + Add'l Credit
Peer index	0.747	0.264	0.380	0.318
Average 8th gr. proficiency	0.692	0.242	0.348	0.290
Enrollment	0.001	-0.179	-0.147	-0.112
Percent free lunch eligible	-0.345	-0.047	-0.112	-0.077
Percent ELL	-0.082	0.115	0.081	0.061
Percent special education	-0.582	-0.225	-0.315	-0.246
Percent self-contained	-0.528	-0.279	-0.346	-0.324
Percent overage	-0.351	-0.0002	-0.075	-0.076
Percent Asian	0.421	0.171	0.233	0.199
Percent white	0.430	0.204	0.258	0.227
Percent black	-0.254	-0.202	-0.221	-0.210
Percent Hispanic	-0.228	-0.007	-0.055	-0.028
Percent missing scores	0.014	0.175	0.144	0.119
R ² from multiple regression	0.628	0.262	0.300	0.228

Notes: includes only the 332 high schools with an overall score reported in 2010-11. “Additional credit” points are excluded from the score in columns 1-3. Multiple regression includes all of the listed variables except the peer index, to avoid double counting (the index uses 8th grade proficiency, percent special ed, percent in self-contained classrooms, and percent overage; see Section 2).

An alternative way to look at the relationship between student characteristics and Progress Report results is to summarize these characteristics by final letter grade. This is provided in Table 5.2, which uses the letter grade as reported on the Progress Report (including additional credit). As suggested by the above correlations, letter grades also vary systematically with school characteristics, including the peer index, average 8th grade proficiency, special education share, and racial composition. For example, the average math and ELA proficiency of “A” and “B” schools was 2.84 and 2.71, respectively, while in “D” and “F” schools the average was 2.48 and 2.47, about one standard deviation lower. In “A” and “B” schools, 25.8% and 18.6% of students was Asian or white, versus 8.9% and 4.2% in “D” and “F” schools.

We also looked at how letter grades varied with school *admissions methods*, shown at the bottom of Table 5.2. For example, 40% of “A” schools used a screened or audition method to admit students, as compared with 15.4% and 14.3% of “D” and “F” schools, respectively. Similarly, “educational option” schools—which aim to admit fixed proportions of low, middle, and high-achieving students—comprise a much higher proportion of “C” – “F” schools than “A” and “B” schools.

How one should interpret these associations with school admissions method depends on how this school characteristic is viewed. If the admission of students through a screened process or not (or some hybrid) is a school practice that can impact student achievement and is something the principal should be held accountable for, then it should not be considered a “pre-existing” condition. However, if admissions method are persistent over time and shape the population

served by the school—which is likely more accurate—then the Progress Report system may be unfairly disadvantaging non-selective schools with less influence over their student population.

Table 5.2: Progress Report Letter Grade and Average School Characteristics

	A	B	C	D	F
Peer index	2.50	2.31	2.12	2.00	1.85
Average 8th grade proficiency	2.84	2.71	2.57	2.48	2.47
Enrollment	753	825	777	893	1345
Percent eligible for free lunch	71.4	72.3	76.5	75.3	77.9
Percent ELL	13.8	12.0	12.4	14.6	10.4
Percent with IEP	12.1	14.2	15.9	15.5	19.4
Percent self-contained	1.9	2.7	3.4	4.4	7.0
Percent overage	6.0	6.7	6.9	8.4	8.7
Percent Asian	13.2	9.7	5.9	6.4	2.5
Percent white	12.6	8.9	4.9	2.5	1.7
Percent black	29.4	40.7	40.8	47.2	46.6
Percent Hispanic	44.0	39.9	47.8	43.0	48.7
Percent missing 8th grade scores	20.0	19.1	15.9	20.3	13.7
Percent screened admissions	36.4	20.2	20.8	11.5	14.3
Percent audition admissions	3.6	9.6	5.2	3.9	0.0
Percent limited unscreened admissions	38.2	42.3	33.8	38.5	21.4
Percent educational option admissions	11.8	14.4	31.2	38.5	42.9
Percent charter school	1.8	1.9	2.6	0.0	0.0
Percent specialized school	5.5	1.9	0.0	0.0	0.0
Percent other/multiple admission type	2.7	9.7	6.5	7.8	21.4

Notes: includes only the 332 high schools with an overall score reported in 2010-11. These letter grades *do* reflect additional credit.

Finally, we note that while a reduction in correlation between school rankings and pre-existing risk factors is a desirable and expected property of a school accountability system (not to mention a stated goal of the NYC Progress Report), it cannot be used *alone* to establish whether such a system is well-designed. To take a simple example, suppose an accountability system simply “curved up” schools with high poverty rates by adding 20 points to their score, and “curved down” schools with low poverty rates by subtracting 20 points from their score. This ad hoc adjustment would substantially reduce the correlation between final scores and poverty, but would do nothing to ensure that a school’s performance was being judged against comparable schools.

6. A Closer Look at the Weighted Regents Subcategories

Weighted Regents Pass Rates represent a sizable proportion of the overall high school Progress Report score (in 2010-11, 25 out of the total 100 points). The calculation behind these subcategory scores is also less transparent than the others. In this section, we take a closer look at how these scores are calculated, and at some of their key implications.

To calculate a Weighted Regents Pass Rate, students who successfully pass a Regents exam are assigned a weight in inverse proportion to their *likelihood* of passing that exam.²⁵ For example, if half of a particular population typically passes Integrated Algebra, then any student who passes from this population is assigned a weight of two ($1 / 0.50$). If 75% typically pass, the passing student is assigned a weight of 1.5 ($1 / 0.75$), and if 100% typically pass, the passing student is assigned a weight of one ($1 / 1.00$). A student who takes and *fails* the exam is assigned a weight of zero. A school's weighted pass rate for an exam is the simple average of its student weights.

As an illustration, consider a hypothetical school with 20 students taking the Chemistry Regents Exam. Based on past results for the Chemistry Regents, one could predict each student's likelihood of passing the exam and determine what weight they should receive if they pass. Some hypothetical weights and exam results for four groups of students are shown in Table 6.1.

Table 6.1: Hypothetical Weighted Regents Pass Rate for Chemistry

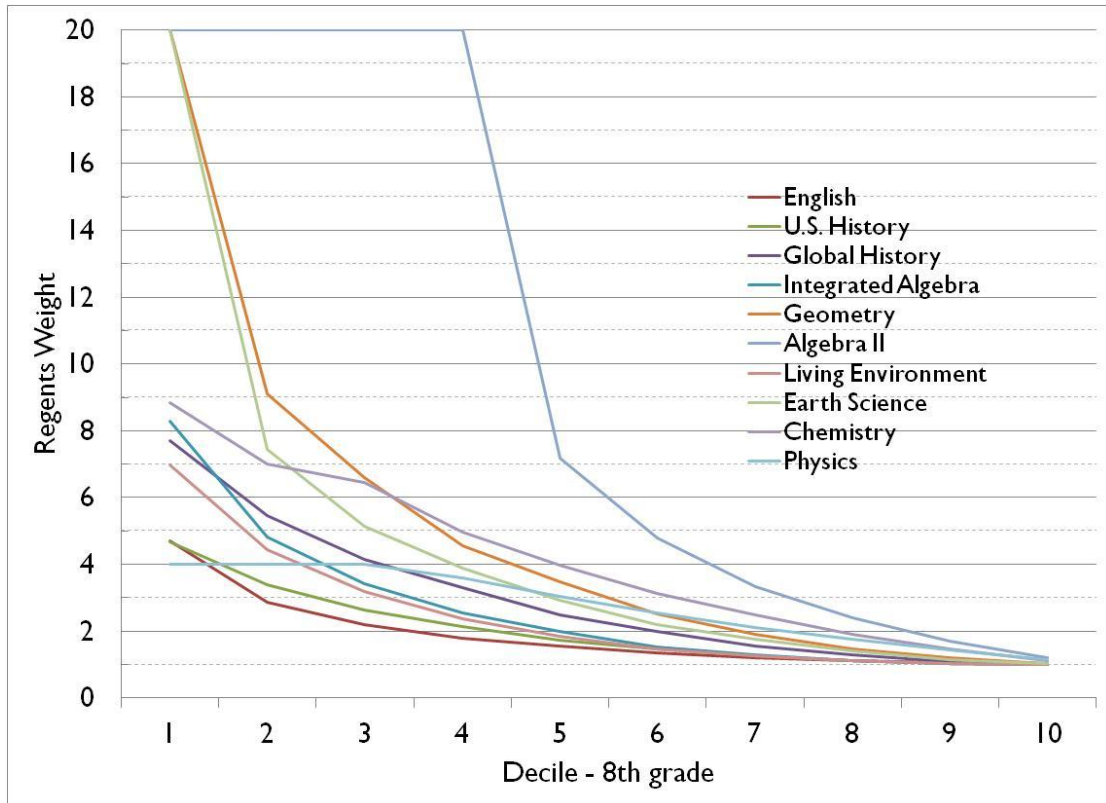
Group	# Total	# Passed	% Passed	% Likely to pass	Weight	Weighted students
1	5	3	60%	40%	$(1 / 0.40) = 2.5$	$(2.5 \times 3) = 7.5$
2	8	4	50%	60%	$(1 / 0.60) = 1.67$	$(1.67 \times 4) = 6.67$
3	5	5	100%	75%	$(1 / 1.75) = 1.33$	$(1.33 \times 5) = 6.67$
4	2	2	100%	100%	$(1 / 1.00) = 1.0$	$(1.0 \times 2) = 2.0$
Total	20	14	70%			22.83
Average (Weighted Regents Pass Rate)						$(22.83 / 20) = 1.14$

In this school some groups performed better than predicted, while others performed worse. In Group 1, 3/5 passed (60%) instead of the predicted 40%. In Group 2, 4/8 passed (50%) instead of the predicted 60%. If every group performed *exactly* as predicted, the Weighted Regents Pass Rate would equal 1.0. As groups perform better or worse than predicted, the pass rate moves above or below 1.0. Moreover, a group's influence on the weighted pass rate depends on the *share* of total enrollment in that group. So, a school may have higher-than-predicted success in passing students in Group 1, but if Group 1 students make up a relatively small share of enrollment in that school, their success will have a modest impact on the weighted pass rate.²⁶

In practice, a student's predicted likelihood of passing a Regents—and thus the weight they receive—is based on their 8th grade proficiency in a particular subject. The 2010-11 weights used ($1/\text{the likelihood of passing}$) are pictured in Figure 6.1. For example, on the U.S. History

Regents, students who scored in the 2nd decile in 8th grade history (between the 10th and 20th percentiles) are assigned a weight of 3.39 upon passing. Those in the 8th decile who pass are assigned a weight of 1.12. Clearly the highest weights are assigned to initially low-performing students, and to students passing Physics, Chemistry, Geometry, and Earth Science. Initially high-achieving students and students passing the English, U.S. History, Living Environment, and Integrated Algebra exams are given relatively lower weights.

Figure 6.1: Weights Used in Weighted Regents Pass Rates, 2010-11



Students *missing* 8th grade proficiency scores are imputed weights according to certain demographic characteristics (black/Hispanic, free lunch eligible, special education, ELL, and “interrupted formal education”), as described in the NYCDOE *Educator’s Guide*.

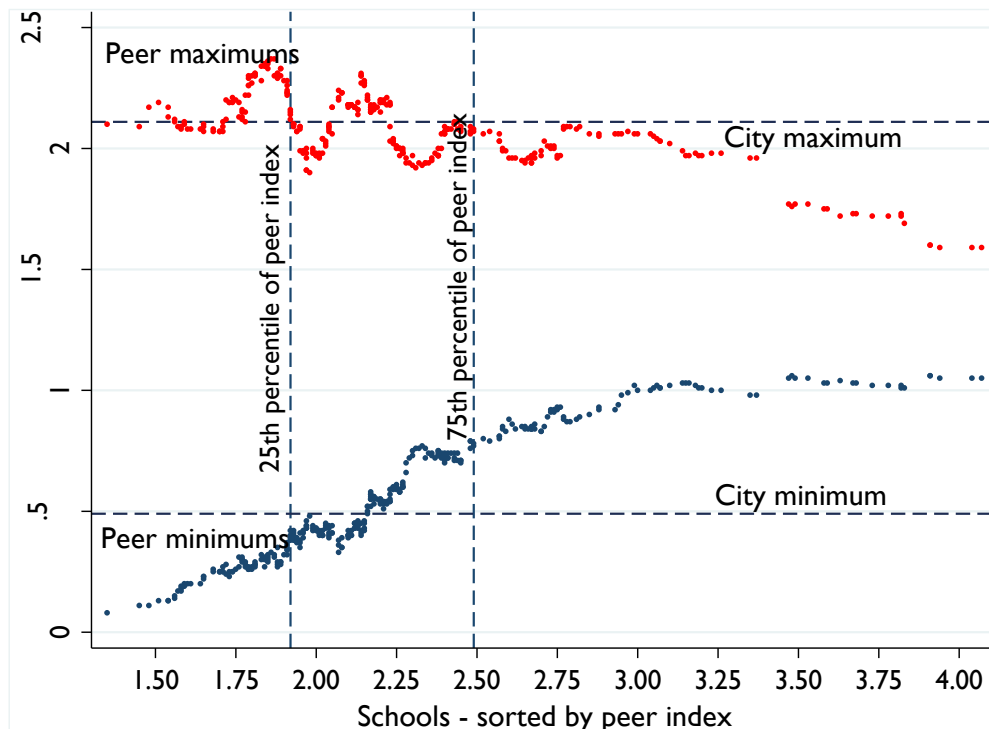
Once a school’s Weighted Pass Rate is determined, its peer and city scores are calculated using the “percent of range” method described in Sections 2 and 4. As an example, School A’s weighted pass rates are reported in Table 6.2. An interesting pattern seen here is that the peer ranges for School A are uniformly *wider* than the city ranges. This is the *opposite* of the usual pattern, where peer ranges tend to be narrower than those observed citywide (e.g., Figure 4.2). The reason is as follows: schools low on the peer index (like School A) are by definition most likely to enroll students with high weights, low-achieving students and those with special needs. Thus, the *potential variability*, or range in weighted pass rates, is much higher for schools serving

initially low-achieving students than for more advantaged schools with fewer highly-weighted students. The minimum and maximum observed for schools with high-weight students would be considered outliers citywide, and thus are excluded from the citywide range. Figure 6.2 shows this clearly, for Weighted Regents Pass Rate in mathematics. As in Figure 4.2, each pair of dots on the vertical dimension is the upper and lower bounds of the peer range for a school with a given peer index (on the horizontal axis). The peer range is much wider for low peer index schools than for high peer index schools.

Table 6.2: Weighted Regents Pass Rates for School A, 2010-11

	English	Math	Science	Global History	U.S. History
Weighted Regents Pass Rate	0.92	1.64	0.83	1.25	1.18
Peer Percent of Range	33.30	66.80	26.50	47.00	50.60
Peer Min	0.50	0.29	0.22	0.14	0.34
Peer Max	1.76	2.31	2.52	2.50	2.00
City Percent of Range	28.30	71.00	21.40	50.00	53.10
City Min	0.62	0.49	0.47	0.30	0.50
City Max	1.68	2.11	2.15	2.20	1.78
Width of Peer range (Max – Min)	1.26	2.02	2.30	2.36	1.66
Width of City range (Max – Min)	1.06	1.62	1.68	1.90	1.28

Figure 6.2: Peer and City Maximum and Minimum, Weighted Math Regents



Based on the above, it is clear the Weighted Regents Pass Rate calculations are fundamentally different than for the other subcategories, such as the 4-year graduation rate. In the latter case, each school's graduation rate is simply compared to those attained by its peer group and other schools citywide to compute peer and city scores. For the Weighted Regents Pass Rates, student results are *first benchmarked against expectations*—with higher scores given to schools whose students perform better than predicted. Then, this adjusted score is compared to the scores attained by its peer group and schools citywide. Effectively, it is “benchmarked” twice.

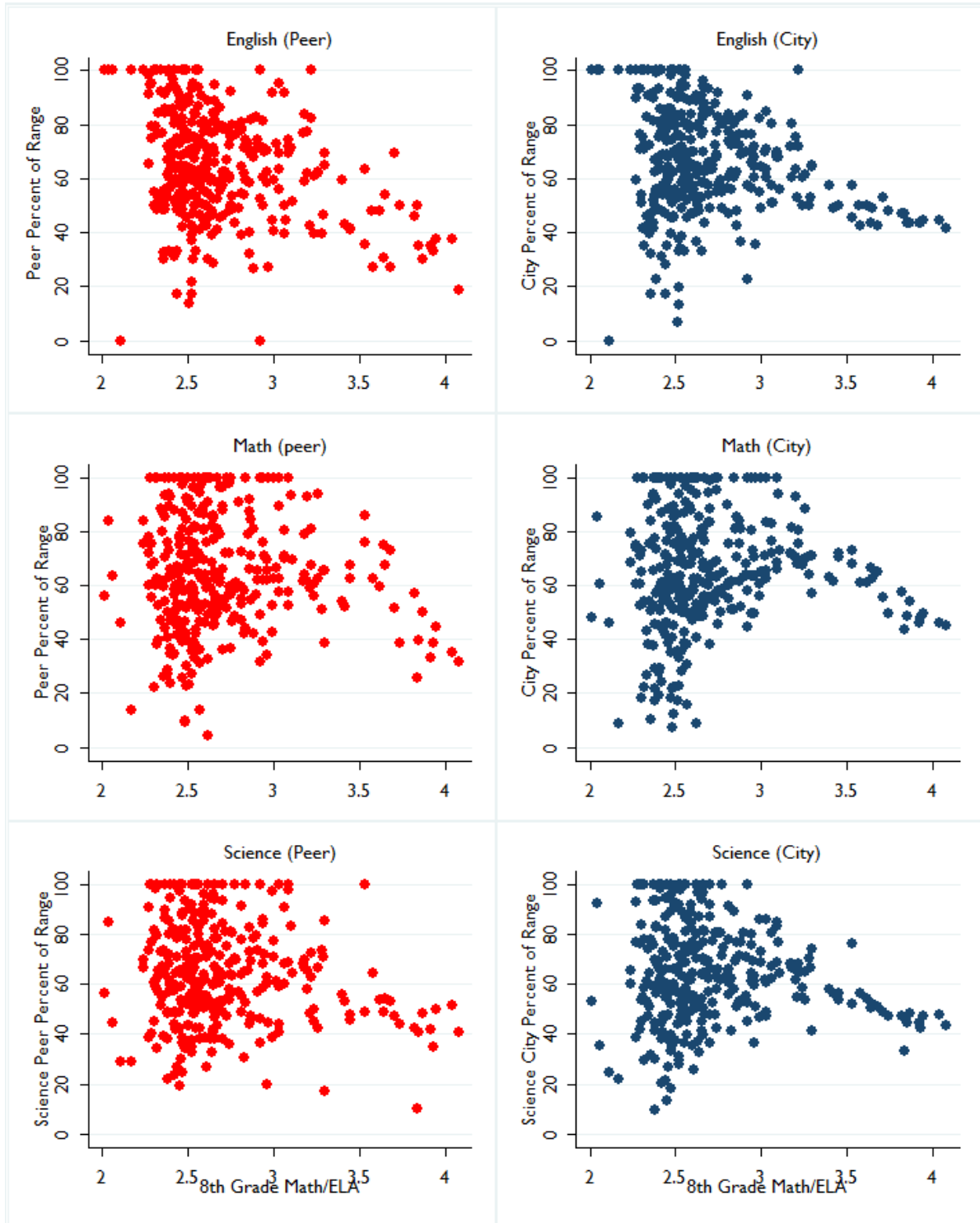
We see several implications of this method. First, there is a theoretical limit to the weighted pass rate a school can receive that depends on the composition of students enrolled there.²⁷ For example, if a school is made up entirely of top decile students with a Regents pass rate weight of 1.0, the school can do no better than a 1.0 weighted pass rate. If a school is made up entirely of bottom decile students with a weight of 4.69 for passing (e.g. as on the English Regents), it can receive no greater than a 4.69 weighted pass rate. Obviously, most schools have a mix of students that determines what their maximum “potential” score will be.

The issue arises when comparing schools' Weighted Regents Pass Rate to their peers and the city (using the “percent of range”). It is possible a school could perform at or near its maximum possible score, yet still score low relative to the peer or city range. This is most likely to apply to academically advantaged schools whose peers (and other schools citywide) have higher *potential* scores. Peer comparisons are most likely to be misleading in this case when peer groups are dissimilar (as we found them to be on average in Section 3).

Figure 6.3 shows how peer and city Weighted Regents Passing Rate scores (the “percent of range”) relate to the average proficiency level of a school's incoming students. Each point is a high school, with its average proficiency on the horizontal axis and peer (or city) score on the vertical axis. Three subject areas are shown: math, English, and science. It is clear schools with higher average proficiency among their incoming students tend to score lower on *both* peer and city Weighted Regents comparisons, with almost no schools scoring relatively high by these measures. This likely reflects the theoretical limit faced by these schools and their inability to improve relative to peers (and other schools citywide) with higher potential scores.

We took a closer look at the cluster of scores at 100%, not only because these schools appear to be top performers, relative to expectations, but also because extreme cases define the set of plausible outcomes for their peers (i.e., the peer range). Notably, we found that high schools scoring at 100% on both the peer and city Weighted English measures were disproportionately international schools with a high proportion of recent immigrants and missing scores on the 8th grade ELA test. (These included the Manhattan International High School, the International Community High School, the Academy for Language and Technology, and others). While these schools may indeed be performing at a high level, their over-representation at the maximum raises the question of whether their Weighted Regents provides an appropriate benchmark for other schools in their peer group, or citywide. Their high Weighted Regents score may be high in part because the expected pass rates on which they were based were too low.

Figure 6.3: Relationship between Weighted Regents Scores (Percent of Range) and Schools' Average 8th Grade Proficiency: English, Mathematics, and Science



7. Discussion and Recommendations

Accountability systems have many goals. They focus attention on—and delineate expectations for—desired student outcomes. They provide metrics that the district and its stakeholders can use to monitor school performance. They serve as a learning tool for educators, highlighting success areas or areas in need of improvement. They draw attention to schools that are performing well relative to their peers, providing models of good practice, and provide a basis for intervention or closure when a school is performing poorly.

Designing an accountability system that meets these diverse goals, that is rigorous and fair and at the same time transparent and accessible, is not an easy task. Difficult choices must be made regarding which outcomes to include and how they should be measured—e.g., status vs. growth, criterion vs. norm-referenced—and the extent to which outcomes should be adjusted to account for differences in student populations. Separating the impact of the school and its practices from pre-existing risk factors becomes particularly important when high stakes are involved, and there are a number of alternative, if imperfect, ways to do so (Raudenbush, 2004).

The NYCDOE has been attentive to these design issues in creating its accountability system, one of the most comprehensive and sophisticated in the nation. The Progress Report provides a kind of “balanced scorecard” that incorporates both intermediate (e.g., credit accumulation, Regents Exam passing rates) and long-run indicators (graduation rates, *College and Career Readiness*). Each outcome is norm-referenced in that it is benchmarked against the historic performance of other schools. Differences in student background are addressed using a school peer index and peer groups. The Progress Report remains an evolving tool, regularly refined by the NYCDOE while maintaining continuity with prior years.

However, some of the design choices on the high school Progress Report result in an opaque picture of how schools are performing relative to similar schools, in some cases systematically disadvantaging or advantaging schools with certain characteristics. This paper sought to shed light on the methodology behind the high school Progress Report and highlight some of these design issues. Its key points and conclusions can be briefly summarized as follows:

- The score a school receives in each subcategory depends on how it performed relative to the *range* of outcomes observed in other schools over the preceding four years. Its performance is benchmarked against outcomes of schools in its peer group, and against all schools citywide.
- Peer groups are formed using a *peer index* that is based largely on average proficiency of a school’s incoming students, with adjustments for the percent who are receiving special education, or are over age. Importantly, peers only affect a school’s score to the extent they alter the range of outcomes to which its performance is benchmarked.
- By definition, the range is sensitive to extremes—the highest and lowest outcomes observed in other schools. In practice, it is limited to two standard deviations below and above the mean, which narrows the range for some schools. In others, however, the

peer minimum or maximum lies beyond what is actually observed in the data. This means: (1) peer ranges, meant to identify a narrower set of plausible outcomes for each school, often remain quite wide; (2) the peer range for neighboring schools on the peer index can vary in random ways depending on which schools are in their peer group; and (3) atypical schools play a disproportionate role in defining schools' peer range.

- Because the peer index is based primarily on 8th grade test scores, peer groups are very diverse in practice, differing in size, location, admissions method, poverty rate, and the percent of students with special needs. The more diverse the peer group, the wider the range of outcomes one is likely to observe among those schools. The wider the range of outcomes, the less a peer comparison represents a sharp comparison of similar schools.
- Peer ranges are wide enough that scores on the Progress Report are often not much different from what they would be if peer comparisons were ignored and a school's outcomes were simply compared to all others citywide. This is most true for schools in the middle of the peer index (where most schools are), and less so for schools serving very high shares of low- or high-performing students.
- Progress Report scores remain correlated with many pre-existing risk factors, including poverty, 8th grade achievement, the percent of students who are ELLs, and the school's admissions method (i.e., selective or non-selective).
- Because a school's peer index is based primarily on the proficiency of its incoming students, it is less accurate for schools where a large fraction of students are missing test data. (About 1 in 10 schools have 30% or more students without 8th grade scores). This can bias a school's Progress Report score upward or downward, depending on the profile of students missing scores. As importantly, this can bias *other* schools' scores, when their peer range is distorted by schools improperly included in their peer group.
- The Weighted Regents scores differ in that they are "benchmarked" twice. The original weighted passing rate reflects how students performed on the Regents relative to predicted (based on 8th grade scores). But this measure is benchmarked a second time against the weighted pass rates of peers and all schools citywide. This can penalize schools for not achieving results that are mathematically impossible to achieve with their population of students. Schools serving high-achieving students appear to be most affected by this.

Two major themes emerge from these findings. First, a peer index dependent on a single school characteristic (the 8th grade scores of incoming students) puts the system at risk of ignoring other important student and school factors that affect achievement. Second, the use of *ranges* to benchmark performance makes scores susceptible to extreme values, and sensitive to diverse peer groups or atypical peers. Limiting the range to two standard deviations from the mean helps, but the standard deviation is also sensitive to extreme values, and in practice peer ranges remain quite wide.

The tendency for peer ranges to be wide means that, for the average school, peer comparisons do not differ substantially from a citywide ranking, and thus do not go far to ensure similar schools are compared. *On average*, the selection of peers does not appear to be very important. At the same time, the sensitivity of the peer range to atypical peers introduces some random variation, where two schools with similar performance and a similar peer index can have different Progress Report scores simply due to which specific peers are in their peer group. (An example is provided in Supplementary Appendix F). Finally, the wider the peer range, the more challenging it is for a school to improve its score.²⁸

Based on these findings, we have several recommendations for improving the design and validity of the high school Progress Report. We view these recommendations as the beginning of a thoughtful conversation about the goals of the Progress Report, and how its design can best meet these objectives.

1. Peers should be defined based on more than a single-dimensional peer index.

While the current peer index has the advantage of being relatively transparent—schools always know to whom they are being compared—it also runs the risk of rewarding or punishing schools that are advantaged or disadvantaged along other dimensions associated with outcomes. Its accuracy is also sensitive to missing data on middle school test scores. We recommend considering one of two alternatives:

- Abandon the use of 40-school peer groups based on test scores, and instead benchmark high school outcomes against those *predicted* given schools' baseline characteristics. Predictions could be based on *all* available information, including past achievement, mobility, poverty and immigrant status, prior attendance, special needs, school size and admissions method, and so on. Schools performing above expectations would score higher on this measure, while those performing below expectations would score lower. This approach, which is similar to that being considered by the New York State Education Department to evaluate school principals, would render peer and city ranges unnecessary.
- Maintain peer groups, but use an alternative method to assign peers, such as grouping schools according to predicted performance in a subcategory (where predictions come from pre-existing student and school characteristics). Under this system, the peer and city ranges and “percent of range” score could be preserved (although see Recommendation #2).

2. The use of ranges to benchmark outcomes should be modified or abandoned.

As noted above, the use of ranges introduces multiple problems. They are particularly sensitive to outliers, diverse peer groups, and atypical peers. For the average school, the peer range is wide relative to the city range, limiting its usefulness as a tool for comparing similar schools. Again, we recommend one of several alternatives:

- As suggested in Recommendation #1, move to a system that benchmarks actual performance against predicted, which renders peer groups—and peer and city ranges—unnecessary.

- Maintain peer groups but adopt an alternative measure of peer performance (i.e., other than “percent of range.”) For example, a school’s outcome—such as its 4-year graduation rate—could be compared to the peer group mean, with points assigned according to how far above or below this mean the school performed.
 - Maintain peer groups, but adopt a method of *smoothing* the peer minimum and maximum, to avoid random fluctuations in the peer range for schools with a similar peer index. (See the example provided in Supplementary Appendix F).
3. **The Weighted Regents Pass Rate should not be benchmarked against the peer or city range.** Schools can be penalized on the Weighted Regents measures by not attaining scores that are mathematically impossible, given their student population. An alternative method to the “percent of range” should be found to assign scores to schools based on their Weighted Regents Passing Rates.
 4. **The effects of missing 8th grade test scores needs to be addressed.** The current peer index relies heavily on the average proficiency of incoming 8th graders. This raises the possibility that the peer index for schools with a high fraction of missing test scores over- or understates their students’ true proficiency (depending on the profile of students missing data). An improved system could impute proficiency for students missing 8th grade test scores, using other characteristics predictive of proficiency, or explicitly account for the percent missing test scores in a regression approach.
 5. **The report should be more transparent about how peer groups affect scores.** The current use of 40-school peer groups offers an impression that each school is being compared with a narrowly tailored group of similar schools. In fact, the use of peer groups has a modest effect on the score of the average school, and introduces random differences in the scores of neighboring schools on the peer index. The Progress Report documentation should be more transparent about the influence a peer group can have.

It is important to keep in mind that the limitations identified in this report do not necessarily render the high school Progress Report invalid and that even an imperfect accountability tool can incentivize changes in behavior and improve school performance (e.g., Rockoff & Turner, 2008; Winters & Cowen, 2012). Further research is needed in order to fully assess the ways in which these proposed alternatives improve upon the current system, and the extent to which these findings apply to the elementary/middle school edition of the Progress Report. We hope, however, that this report and the above recommendations will spark a productive conversation about how the high school Progress Report can be improved to better meet its objectives.

References

Clotfelter, C. T., & Ladd, H. F. (1996). Recognizing and Rewarding Success in Public Schools. In Clotfelter, C. T., & Ladd, H. F., Eds. *Holding Schools Accountable*. Washington, D.C.: The Brookings Institution. (pp. 23–63).

Hamilton, L. S., & Koretz D. M. (2002). Tests and Their Use in Test-Based Accountability Systems. In Hamilton, L. S., Stecher, B. M., & Klein, S. P., *Making Sense of Test-Based Accountability in Education*. Santa Monica, CA: RAND Corporation.

Linn, R. L. (2005). Issues in the Design of Accountability Systems. CSE Technical Report 650. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from <http://www.cse.ucla.edu/products/reports/r650.pdf>

Linn, R. L. (2006). Educational Accountability Systems. CSE Technical Report 687. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from <http://www.cse.ucla.edu/products/reports/R687.pdf>

Raudenbush, S. W. (2004). *Schooling, Statistics, and Poverty: Can We Measure School Improvement?* Princeton, NJ: Educational Testing Service.

Rockoff, J., & Turner, L. J. (2010). Short-Run Impacts of Accountability on School Quality. *American Economic Journal: Economic Policy*, 2(4), 119–147.

Winters, M. A., & Cowen, J. M. (2012). Grading New York. *Educational Evaluation and Policy Analysis*, 34(3), 313–327.

¹ At least one study has found that schools in New York City responded to the elementary/middle school Progress Report in ways that produced gains in math and ELA scores (Rockoff & Turner, 2008; see also Winters & Cowen, 2012).

² This Independent Budget Office report was released at about the same time the first draft of this paper was given to the NYCDOE to review. See also “On Schools’ Performance, Invisible Line Between ‘A’ and ‘F’”, *The New York Times*, April 30, 2012, p. A22.

³ We rely heavily on various editions of the NYC DOE publication “Educator Guide: The New York City Progress Report,” especially the 2009-10 (November 3, 2010 edition; a newer version is available from August 30, 2011), and 2010-11 editions (November 28, 2011). See <http://schools.nyc.gov/Accountability/tools/report/default.htm>

⁴ The additional credit component awards extra points to schools with particularly strong performance with disadvantaged populations.

⁵ This is presumably to minimize year-to-year variability in the range of results that might be observed in a 40-school peer group (perhaps due to “cohort effects”). Rather than forming a peer range with only 40 values, this approach uses 160 (40 x 4) values, assuming a result was reported for every school in each of the four years.

⁶ I thank Mark Dunetz for clarifying this. It appears the NYCDOE may have changed its method of calculating the peer range in 2010-11. The 2009-10 Educator’s Guide defines the peer minimum and maximum as follows: “The peer range ‘minimum’ is the *lowest non-outlier score* and the peer range ‘maximum’ is the *highest non-outlier score* [emphasis added],” where outliers are defined as values more than two standard deviations away from the mean. This definition would suggest that the minimum is the observed minimum *or* the mean minus two standard deviations, whichever is greater, and the maximum is the observed maximum *or* the mean plus two standard deviations, whichever is smaller. A replication using 2010-11 data suggests that the minimum and maximum are defined as the mean +/- two standard deviations.

⁷ $(\text{School Result} - \text{Peer Minimum}) / (\text{Peer Maximum} - \text{Peer Minimum}) = (58.7 - 41.3) / (87.6 - 41.3) = 0.376$.

⁸ The maximum peer index observed in 2010-11 was 4.07 while the minimum was 1.35.

⁹ Combining a test score measure (average proficiency) with a percentage (the percentage of students meeting some criteria) is peculiar, and it is unclear where the multiplying factors come from. (A 1 point increase in the percent of students with special needs has twice the impact on the peer index as a 0.01 point lower average proficiency level. Presumably, only 8th graders with prior test scores are used in the peer index. Those new to the district or otherwise missing a score would not play a role. See Supplemental Appendix B.

¹⁰ A look at School A’s peers in 2010-11 reveals a faithful application of this rule, with some exceptions. Of School A’s 40 closest schools on the peer index, only 34 were actually included in the peer group; the remaining six were schools slated for closure. To accommodate the loss of schools, the lower and upper bounds of School A’s peer group were extended to include six additional schools.

¹¹ These are again based on the 332 schools in 2010-11 with an overall Progress Report score. In most cases, it makes little difference whether one examines the *median* range of peer characteristics or the mean. The latter has the potential to be influenced by schools with unusually wide or narrow ranges of peer characteristics.

¹² This is not driven by outliers with a particularly large enrollment range. The median range is 3,781 students, and the 25th percentile is 2,070.

¹³ A table of these results can be found in Supplemental Appendix E.

¹⁴ See Corcoran and Levin (2010). High schools can have more than one *program*, and these programs can have different admissions methods. In order to assign a single admissions method to a school, we did the following: (1) if the school has only one program, use that admissions method (about 75% of schools); (2) if the school has multiple programs of the same type, use that admissions method; (3) if the school has multiple programs and a modal admissions method, use that method; (4) hand inspect the remaining cases.

¹⁵ 8th grade proficiency also plays a critical role in the Weighted Regents Pass Rates, as described in Section 6.

¹⁶ In Supplemental Appendix A we provide a general rule to determine whether a school’s peer percent of range will be higher or lower than its city percent of range.

¹⁷ Peer minimums can fall below the city minimum if they are low enough to be considered “outliers” in the city distribution.

¹⁸ Detailed statistics are provided in the Supplemental Appendix Table A.1. As that table shows, the *median* difference in peer and city percent of range is typically smaller than the mean difference. This is due to the small number of schools in the tails with large differences.

¹⁹ An alternative benchmark would be the standard deviation in the city percent of range, a measure of variation across schools in that category. This ranges from about 16.8 (Engagement) to 24.4 (Credit accumulation Year 2, lowest third).

²⁰ The Weighted Regents Pass Rates differ from many of the other subcategories in that its initial calculation already reflects differences in student populations. We address this in Section 6.

²¹ These results are shown visually in the Supplemental Appendix, Figure E.1.

²² The correlation for the *Environment* scores is 0.88. The Spearman rank correlations (as opposed to Pearson) are very similar: 0.91, 0.85, and 0.86 for *Progress*, *Performance*, and *Environment*, respectively). As seen in Figure 4.6, the combined peer/city score is more likely to be higher than the citywide score than lower. What is more important, however, is that the rankings of schools remain largely intact.

²³ As an alternative to using the same cut scores, we also applied the same *percent* distribution of grades. These results are given in the Supplemental Appendix Table E.3.

²⁴ A complete table of correlations including all category and subcategory scores is provided in Supplemental Appendix E.

²⁵ A score of 65 or higher is considered passing.

²⁶ An alternate presentation of the Weighted Regents Pass Rates is provided in Supplemental Appendix C.

²⁷ See Supplemental Appendix C for a more detailed discussion.

²⁸ The extent to which improvements translate into higher scores on the Progress Report depends entirely on the width of the peer and city ranges. The wider the range, the larger the improvement required to raise one's score. As an example, the average 4-year graduation rate across schools in 2010-11 was 71.8. The mean improvement necessary to achieve a half-point gain in this subcategory was 3.7 percentage points, a relatively large increase (about $\frac{1}{4}$ standard deviation in the distribution of 4-year graduation rates). The total number of points possible in the 4-year graduation rate subcategory is 5, the average points earned in 2010-11 was 2.9, with a standard deviation of 1.0. The improvement in graduation rates required for a half-point score gain was largest for schools in the middle of the peer index, at around 4.0 percentage points.

Appendix A: How the Peer Percent of Range Differs from the City Percent of Range

A general rule can be used to determine whether a school's peer percent of range will be higher or lower than its city percent of range. Suppose **a** and **b** are the minimum and maximum of the citywide range, Δ_1 is the difference between the peer and city minimum (usually, $\Delta_1 \geq 0$), and Δ_2 is the difference between the peer and city maximum (usually, $\Delta_2 \leq 0$). If **x** is a school's score, it is easy to show that the peer percent of range will be greater than the citywide percent of range whenever:

$$x > \frac{\Delta_1 b - \Delta_2 a}{\Delta_1 - \Delta_2}$$

To illustrate this rule's application, consider these cases:

- If the peer range has a higher *minimum* but the same *maximum* ($\Delta_1 > 0$ and $\Delta_2 = 0$), then the peer percent of range will always be lower than the city percent of range. This is more likely to apply to schools near the top of the city distribution.
- If the peer range has the same *minimum* but a lower *maximum* ($\Delta_1 = 0$ and $\Delta_2 \leq 0$), then the peer percent of range will always be higher than the city percent of range. This is more likely to apply to schools near the bottom of the city distribution.
- If the peer minimum and maximum differ from **a** and **b** by the *same amount* ($\Delta_1 = |\Delta_2|$), then the peer percent of range will be higher than the city percent of range whenever **x** is greater than the midpoint of **a** and **b** ($(a+b)/2$). Otherwise, the peer percent of range will be lower than the city percent of range.
- If the peer minimum and maximum differ from **a** and **b** by *different amounts*, then conditions under which the peer percent of range exceeds the city percent of range depend on the relative change of the minimum and maximum. If the minimum differs by more than the maximum ($\Delta_1 > |\Delta_2|$) then the values of **x** for which the peer percent of range will be higher than the city percent of range are higher than the midpoint. If the maximum differs by more than the minimum ($\Delta_1 < |\Delta_2|$) then the values of **x** for which the peer percent of range will be higher than the city percent of range is lower than the midpoint.

Using the numbers from the School C example in Section 4, the city range is 6.0 – 88.0, while the peer range is 17.5 – 56.1. The minimum increased by 11.5 while the maximum fell by 31.9. Applying the above formula, the peer percent of range will be higher for a school with a first-year credit accumulation rate of 27.7 or higher. (As it happens, this holds for *all* schools in School C's peer group in 2010-11.) This example is a case in which the maximum changes by much more than the minimum.

Thus, as shown in Section 4, the impact of peer grouping from the standpoint of an individual school depends on the extent to which its peer range differs from the city range. Schools that

are relatively *low* on the peer index tend to see their maximum fall by much more than their minimum rises. This tends to increase their peer percent of range relative to their city percent of range. Schools relatively *high* on the peer index tend to see their minimum rise by much more than their maximum falls. This tends to decrease their peer percent of range relative to their city percent of range. Schools near the middle of the peer index tend to see little difference in their peer range, or the drop in their maximum is roughly equivalent to the rise in their minimum. For these schools, the peer percent of range is not likely to differ much from the citywide percent of range.

Appendix Table A.1 summarizes how the weighted peer and city results differ from those based purely on citywide comparisons, by category and subcategory. For example, the average school in 2010-11 would have earned 8.59 points (out of 15) on the School Environment component, based on citywide comparisons alone. The average change between this city score and one that weights both peer and city results is 0.95 (+/-) and the median change is 0.80 (+/-). Ultimately, the correlation between the citywide-only School Environment score and the city/ peer weighted score is 0.879.

When combining all of the main categories (*Environment, Progress, and Performance*) into an final score, the average school would have earned 58.97 points (out of 100) based on citywide comparisons alone (ignoring “additional credit”). The average change between this citywide comparison and one that weights both peer and city scores is 5.69 points (+/-). The median change is 4.65 points (+/-). Ultimately, the correlation between the overall score based only on city comparison and the city/peer weighted score is 0.875.

Appendix Table A.1: Difference between Citywide Only Score and Weighted Peer/City Progress Report Score, by Category and Subcategory

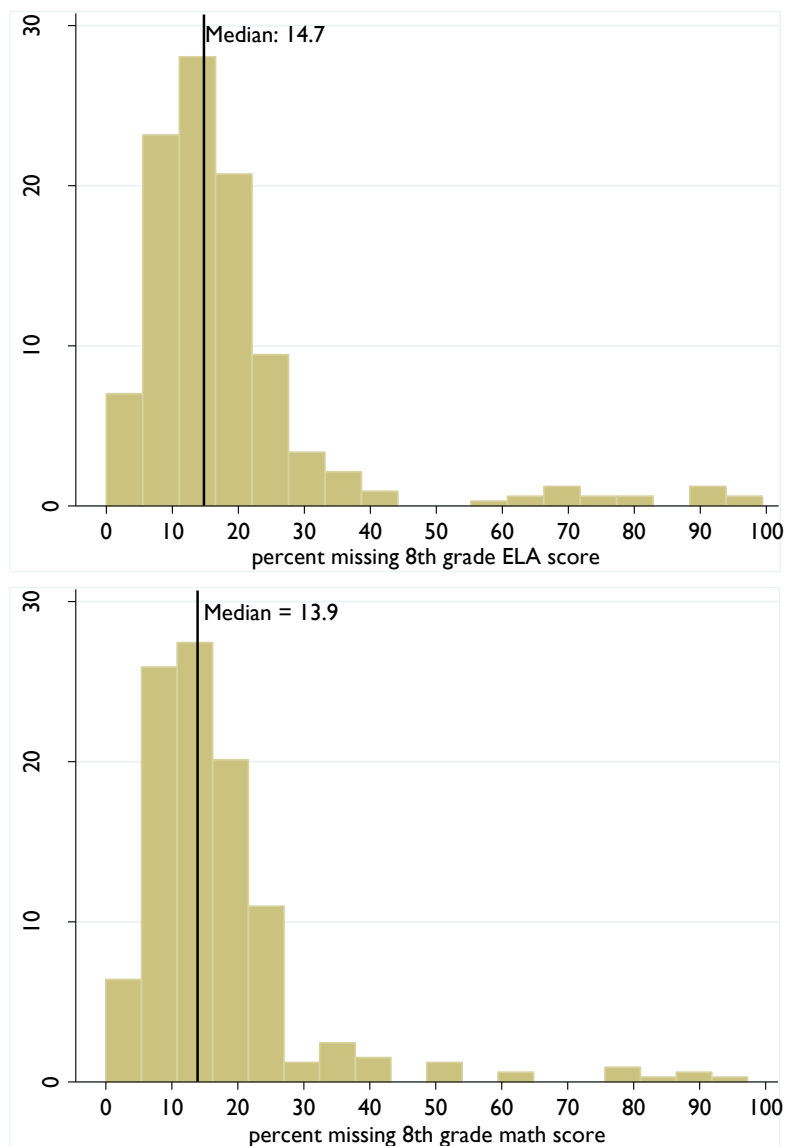
	City only		City & peer		Difference (absolute value)		Correlation
	Mean	SD	Mean	SD	Mean	Median	
School environment (15)	8.59	2.49	8.54	2.34	0.95	0.80	0.879
Safety	1.40	0.52	1.42	0.52	0.17	0.20	0.902
Academic expectations	1.47	0.46	1.49	0.51	0.15	0.10	0.927
Communication	1.53	0.48	1.51	0.50	0.09	0.10	0.971
Engagement	1.59	0.42	1.60	0.44	0.10	0.10	0.957
Attendance rate	2.59	1.17	2.52	1.07	0.49	0.40	0.837
Student progress (60)	35.23	8.60	35.30	7.92	2.70	2.10	0.914
Credit accum. 1st yr	3.06	1.07	2.95	1.04	0.30	0.20	0.911
Credit accum. 1st yr (low 3)	2.86	1.09	2.91	1.09	0.28	0.20	0.933
Credit accum. 2nd yr	2.80	1.17	2.79	1.14	0.37	0.30	0.903
Credit accum. 2nd yr (low 3)	2.68	1.22	2.71	1.21	0.30	0.20	0.942
Credit accum. 3rd yr	2.69	1.16	2.72	1.12	0.36	0.30	0.923
Credit accum. 3rd yr (low 3)	2.59	1.20	2.61	1.18	0.33	0.30	0.937
Average Regents pass rate	2.94	1.08	3.20	0.89	0.57	0.50	0.792
Weighted Regents: English	3.30	0.97	3.24	0.99	0.18	0.20	0.973
Weighted Regents: math	3.25	1.07	3.23	1.05	0.17	0.10	0.978
Weighted Regents: science	3.23	1.00	3.21	1.00	0.15	0.10	0.978
Weighted Regents: global	3.02	1.09	2.97	1.10	0.18	0.10	0.978
Weighted Regents: US	2.77	1.11	2.71	1.10	0.21	0.20	0.968
Student performance (25)	15.15	5.0	15.79	4.13	2.21	1.90	0.848
4-year graduation rate	3.61	1.46	3.66	1.35	0.46	0.40	0.906
6-year graduation rate	3.96	1.36	4.06	1.27	0.48	0.40	0.899
Weighted diploma rate 4-yr	3.79	1.28	3.98	1.16	0.67	0.50	0.773
Weighted diploma rate 6-yr	3.85	1.26	4.17	1.15	0.72	0.70	0.776
Overall score	58.97	14.61	59.63	12.18	5.69	4.65	0.875

Notes: N=332 (only schools with an overall score reported in 2010-11). "Additional credit" not considered here.

Appendix B: Effects of Missing 8th Grade Test Score Data

Because the peer index relies primarily on average 8th grade proficiency, it is natural to ask how schools with high fractions of students missing 8th grade scores are affected by the peer index system. As seen in Appendix Figure B.1, some New York City high schools enroll substantial shares of students with missing 8th grade scores.¹ The average high school in 2010-11 had 17 to 19% of students missing 8th grade scores, and about 1 in 10 had 30% or more of students without 8th grade scores. Students may lack scores if they are new the district in 9th grade, or were otherwise absent or excluded from the 8th grade tests.

Appendix Figure B.1: Missing 8th Grade Test Score Data in ELA and Math



How missing test score data is likely to affect a school's score depends on how students with missing data compare to those with test scores. To see this consider two possibilities:

- In School A, students missing 8th grade test scores are more *disadvantaged* than the population of students in A who *do* have test scores. This may be the case if students missing scores are relatively disadvantaged migrants from other districts, or if they are recent immigrants at risk for low achievement. The reported peer index in School A is likely to *overstate* the proficiency of students in that school, and thus A's peers may be more advantaged than if their peer index were a more accurate representation of the school. The result is that A performs comparatively poorly in its peer group.
- In School B, students missing 8th grade test scores are more *advantaged* than the population of students in B who have test scores. This may be the case if students missing scores are transfers from a private or charter school or high-achieving recent immigrants who are likely to perform better than students coming from public middle schools. The reported peer index in School B is likely to *understate* the proficiency of students in that school and thus B's peers may be more disadvantaged than if their peer index were a more accurate representation of the school. The result is that B performs comparatively well in its peer group.

A casual look at the data suggests that New York City schools include both of these school types. For example, many of the schools with the highest proportion of missing 8th grade scores are schools serving recent immigrants, including Newcomers High School (99.5%), Kingsbridge International High School (89.8%), and the International High School at LaGuardia Community College (71.7%).² Additionally, many desirable specialized and screened schools are missing higher-than-average proportions of scores, including Bard High School Early College (23.6%), Staten Island Tech (22.2%), Beacon (21.9%), and Bronx Latin (21.8%). These schools presumably enroll a large number of transfers from private middle schools.

Because missing test data can cause schools to look better *or* worse than they would if their "true" peer index were known (depending on whether it is more like School A or B), it is not obvious whether one effect or the other dominates. As a first pass we used multiple regression to look at the relationship between accountability measures (attendance rate, safety score, graduation rates, etc.) and the percent of students missing test scores, holding constant the peer index and other school characteristics. The question is whether—for a given peer index—schools with more missing data tend to perform better or worse than predicted. On most measures we find the answer is "better," suggesting that for the average school the peer index *understates* a school's incoming student proficiency when its proportion of missing scores is high.³ Of course, how missing data affects the Progress Report score of an *individual* school depends on how its students with missing data compare to its general population.

Appendix C: Alternate Presentation of Weighted Regents Pass Rate Calculation

Suppose students in a school are divided into G groups (indexed g). α_g is the predicted probability of passing for students in group g (e.g. 0.50), and p_g is the actual proportion of students in group g who passed (e.g. 0.60). n_g is the number of students in group g and n is the total number of students in the school. Then the Weighted Regents Pass Rate can be written as:

$$\sum_{g=1}^G \left(\frac{p_g}{\alpha_g} \right) \left(\frac{n_g}{n} \right)$$

This is illustrated in Appendix Table C.1 using the same hypothetical values from Table 7.1. This expression shows that the effect of increasing p_g (relative to α_g) depends on the share of enrollment n that is in group g .

A useful way to think about the effect of increasing pass rates is to ask what the marginal impact of another student passing the Regents would be on the weighted pass rate. This impact is $(1/n\alpha_g)$, which depends on both n and α_g . The score increases by more the lower is the likelihood of the student passing the test, and increases less the larger is n . The marginal impact of improving a *groups'* passing rate is $(n_g/n\alpha_g)$ which depends on α_g and the share of enrollment in group g .

Appendix Table C.1: Alternative Calculation of the Weighted Regents Passing Rate

Group	# Total	# Passed	% Passed	Probability of passing	Product in summation	
1	5	3	60%	40%	$(0.60 / 0.40) * (5 / 20)$	$1.5 * 0.25$
2	8	4	50%	60%	$(0.50 / 0.60) * (8 / 20)$	$0.833 * 0.40$
3	5	5	100%	75%	$(1.00 / 0.75) * (5 / 20)$	$1.333 * 0.25$
4	2	2	100%	100%	$(1.00 / 1.00) * (2 / 20)$	$1.0 * 0.10$
Total	20	14			22.83	1.14

Appendix D: The Impact of Improvements on Progress Report Scores

As described in Section 2, each subcategory score is calculated as follows, where X_i is the subcategory measure for school i (graduation rate, attendance rate, safety score, etc.)

$$\left[\left(\frac{X_i - PMIN}{PMAX - PMIN} \right) * 0.75 * M \right] + \left[\left(\frac{X_i - CMIN}{CMAX - CMIN} \right) * 0.25 * M \right]$$

The first term is the peer percent of range, weighted at 75% and multiplied by the number of possible subcategory points M . ($PMIN$ and $PMAX$ are the peer minimum and maximum). The second term is the city percent of range, weighted at 25% and multiplied by M . ($CMIN$ and $CMAX$ are the city minimum and maximum). How much of an impact will an improvement in X_i have on the subcategory score? It is easy to show that the resulting change in score for a change in X_i (ΔX_i) will be:

$$(\Delta X_i)M \left[\frac{0.75}{PMAX - PMIN} + \frac{0.25}{CMAX - CMIN} \right]$$

To achieve a *one point* increase in the subcategory score requires a change in X (ΔX_i) of:

$$\Delta X_i = \left(\frac{1}{M} \right) \frac{1}{\left[\frac{0.75}{PMAX - PMIN} + \frac{0.25}{CMAX - CMIN} \right]}$$

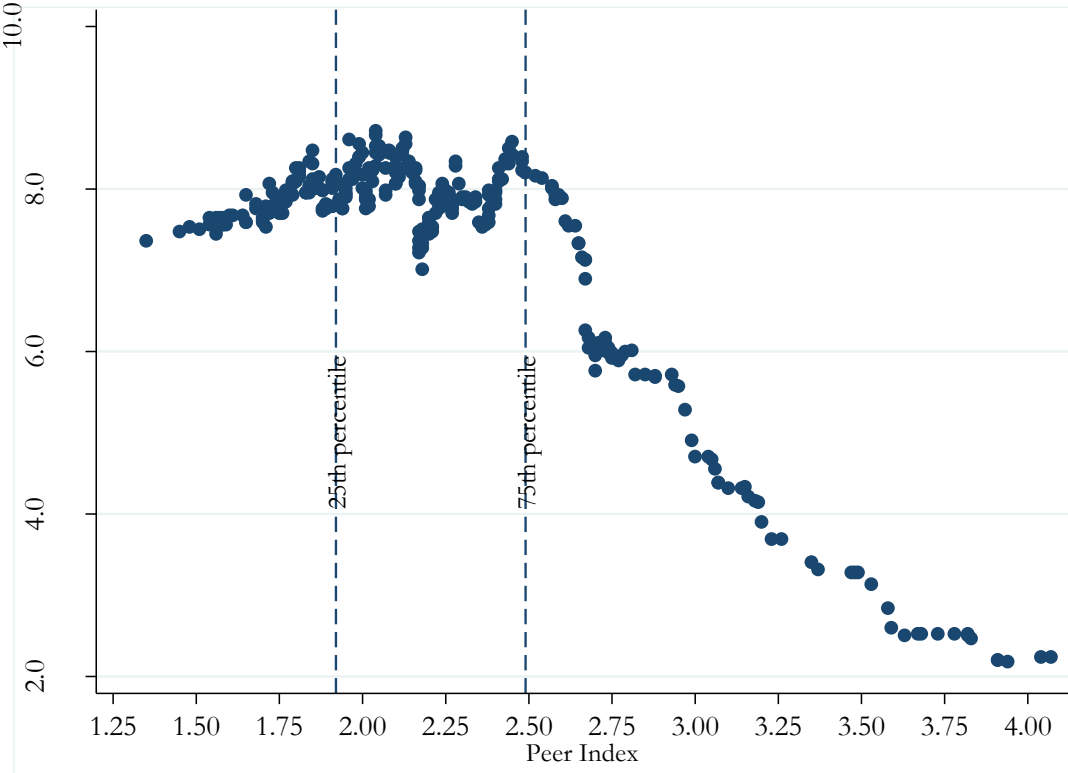
Notice if the peer and city ranges are the same this reduces to:

$$\Delta X_i = \left(\frac{CMAX - CMIN}{M} \right)$$

So, the improvement in X_i required to increase the subcategory score by one point is increasing in the size of the city range (the bigger the range, the further the school has to go to increase its score) and falls in the number of possible points M (this is just a matter of scaling). The same logic applies when the peer and city ranges are not identical; the wider is the peer and/or city range, the larger the improvement in X_i required to increase the subcategory score by one point, with the peer range mattering more than the city range.

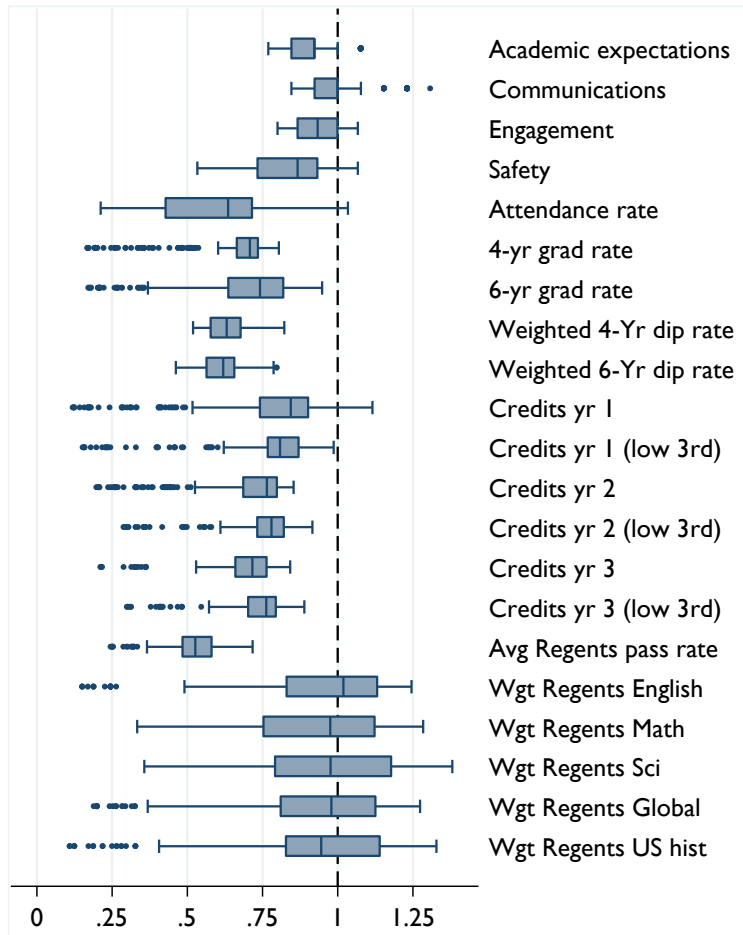
Using the above formula we computed the change in four-year graduation rates required for *each school* to improve the school's Progress Report score by one. (The number varies, since the peer minimum and maximum varies by school). The result is shown in Appendix Figure D.1, which organizes these changes in ascending order by peer index. For the schools in the middle 50% of the peer index, the improvement in graduation rates needed is approximately 8 percentage points.

Appendix Figure D.1: Change in 4-Year Graduation Rate Required for a One-Point Increase the Progress Report Graduation Subscore



Appendix E: Supplemental Figures and Tables

Appendix Figure E.1: Ratios of Peer Range to City Range, by Subcategory, 2010-11



Appendix Table E.1: Distribution of Peer Index Values, 2010-11

	All schools		Schools with an overall score	
	N	%	N	%
Peer Index <1.2	4	0.94		
1.2 <= Peer Index < 1.4	5	1.18	1	0.30
1.4 <= Peer Index < 1.6	18	4.24	13	3.92
1.6 <= Peer Index < 1.8	50	11.76	27	11.14
1.8 <= Peer Index < 2.0	69	16.24	56	16.87
2.0 <= Peer Index < 2.2	81	19.06	71	21.39
2.2 <= Peer Index < 2.4	62	14.59	51	15.36
2.4 <= Peer Index < 2.6	47	11.06	30	9.04
2.6 <= Peer Index < 2.8	31	7.29	27	8.13
2.8 <= Peer Index < 3.0	15	3.53	10	3.01
3.0 <= Peer Index < 3.2	16	3.76	12	3.61
3.2 <= Peer Index < 3.4	6	1.41	5	1.51
3.4 <= Peer Index < 3.6	8	1.88	6	1.81
3.6 <= Peer Index < 3.8	5	1.18	5	1.51
3.8 <= Peer Index < 4.0	8	1.88	8	2.41
TOTAL	425	100.00	332	100.00

Notes: due to missing data and school closures, not all schools assigned a peer index received a Progress Report score in 2010-11. The subset of schools with reported scores are summarized in the two rightmost columns.

Appendix Table E.2: Average Percentage of Peers in the Same Borough and Average Percentage of Peers Using the Same Admissions Method

	Average peers	Citywide
% of peers in the same borough (all)	31.4	24.6
<u>for high schools in:</u>		
Brooklyn	32.7	28.0
Manhattan	28.9	25.0
Queens	25.4	16.6
Staten Island	4.4	2.7
Bronx	38.4	27.7
% of peers with the same selection method (all)	38.0	26.4
<u>for high schools with selection method:</u>		
Audition	9.6	5.9
Educational option	23.5	21.0
Limited unscreened	54.0	38.6
Screened	42.0	25.0

Notes: based on the 332 schools in 2010-11 with an overall Progress Report score. The citywide (all) values in the rightmost column represents, for the average school, the percent of peers that would be expected to be in the same borough if they were chosen at random.

Appendix Table E.3: Actual Progress Report Grade vs. Grade under a Pure Citywide System (Applying the Original Grade Distribution Rather than Cutpoints)

		Actual letter grade on Progress Report (row percentages)					# of schools	% of total
		A	B	C	D	F		
Letter grade based on pure citywide comparison	A	75.5%	23.6%	0.9%	0.0%	0.0%	110	33.1%
	B	23.8	54.3	19.1	1.9	1.0	105	31.6
	C	2.6	28.6	54.6	14.3	0.0	77	23.2
	D	0.0	0.0	42.3	42.3	15.4	26	7.8
	F	0.0	0.0	0.0	28.6	71.4	14	4.2
# of schools		110	105	74	28	15	332	100.0
% of total		33.1%	31.6%	22.3%	8.4%	4.5%	100.0	

Notes: Includes only schools with an overall score reported in 2010-11. "Additional credit" points are added equally to the "pure citywide" score and original formula.

Appendix Table E.4: Correlations between Peer/City Category Scores and Student/School Characteristics

	Peer index	8th Gr Prof	Enrollment	Pct Free Lunch	Pct ELL	Pct IESP	Pct Self-Cont	Pct Overage	Pct Asian	Pct White	Pct Black	Pct Hisp	Pct Missing Score
Overall score	0.380	0.348	-0.147	-0.112	0.081	-0.315	-0.346	-0.075	0.233	0.258	-0.221	-0.055	0.151
Environment	0.146	0.105	-0.188	0.041	0.261	-0.198	-0.238	0.102	0.080	0.146	-0.271	0.147	0.263
Academic exp	-0.059	-0.079	-0.183	0.151	0.188	0.007	-0.125	0.168	-0.115	0.021	-0.142	0.200	0.200
Communications	-0.040	-0.056	-0.161	0.087	0.140	0.027	-0.141	0.115	-0.091	0.068	-0.121	0.135	0.162
Engagement	-0.076	-0.092	-0.085	0.105	0.201	0.022	-0.106	0.180	-0.093	0.085	-0.165	0.172	0.213
Safety	0.060	0.025	-0.252	0.139	0.301	-0.121	-0.219	0.162	0.038	0.100	-0.382	0.319	0.250
Attendance rate	0.371	0.325	-0.088	-0.139	0.184	-0.401	-0.239	-0.068	0.298	0.197	-0.212	-0.069	0.195
Performance	0.425	0.441	-0.102	-0.185	-0.166	-0.231	-0.228	-0.243	0.196	0.241	-0.072	-0.180	-0.102
Gradrate 4yr	0.548	0.552	-0.082	-0.258	-0.271	-0.310	-0.295	-0.388	0.215	0.285	-0.049	-0.241	-0.189
Gradrate 6yr	0.444	0.408	-0.251	-0.208	-0.079	-0.367	-0.329	-0.224	0.133	0.189	-0.023	-0.180	0.005
Weighted dip 4yr	0.204	0.247	-0.009	-0.103	-0.129	-0.006	-0.127	-0.121	0.162	0.176	-0.111	-0.079	-0.069
Weighted dip 6yr	0.210	0.232	-0.045	-0.101	0.030	-0.138	-0.067	-0.038	0.166	0.135	-0.085	-0.093	0.050
Progress	0.319	0.275	-0.117	-0.087	0.133	-0.307	-0.343	-0.019	0.233	0.228	-0.223	-0.034	0.208
Credit acc 1yr	0.390	0.365	-0.069	-0.230	0.039	-0.305	-0.343	-0.084	0.305	0.360	-0.227	-0.150	0.127
Credit acc 1yr low	0.316	0.288	-0.112	-0.151	0.094	-0.291	-0.305	0.011	0.254	0.272	-0.172	-0.125	0.152
Credit acc 2yr	0.376	0.346	-0.103	-0.197	0.148	-0.333	-0.336	-0.024	0.327	0.295	-0.271	-0.080	0.175
Credit acc 2yr low	0.406	0.376	-0.124	-0.194	0.111	-0.336	-0.349	-0.126	0.310	0.275	-0.264	-0.062	0.109
Credit acc 3yr	0.448	0.415	-0.039	-0.255	0.057	-0.388	-0.349	-0.089	0.338	0.319	-0.190	-0.189	0.157
Credit acc 3yr low	0.428	0.393	-0.059	-0.222	0.103	-0.405	-0.338	-0.058	0.308	0.294	-0.215	-0.127	0.201
Avg Regents	0.504	0.526	-0.013	-0.267	-0.224	-0.262	-0.224	-0.376	0.296	0.342	-0.184	-0.170	-0.146
Wgt Regents Eng	-0.261	-0.316	-0.010	0.274	0.379	0.000	0.044	0.320	-0.060	-0.165	-0.124	0.266	0.358
Wgt Regents math	-0.034	-0.046	-0.072	0.099	-0.071	0.062	-0.080	-0.042	-0.143	-0.032	0.085	0.015	0.003
Wgt Regents sci	-0.142	-0.164	-0.132	0.252	0.024	0.075	0.004	0.061	-0.140	-0.218	0.024	0.183	0.038
Wgt Regents global	-0.099	-0.134	-0.092	0.215	0.046	0.045	-0.085	0.038	-0.156	-0.158	0.043	0.142	0.085
Wgt Regents US	-0.075	-0.093	-0.022	0.134	0.071	0.055	-0.086	0.062	-0.049	-0.045	-0.037	0.097	0.094

Appendix F: Idiosyncratic Variation in Peer Ranges

As described in Section 2, the *peer percent of range* is a school's outcome expressed as a percent of the distance from the peer minimum to the peer maximum. The peer minimum and maximum are *unique to each school*, and are based exclusively on the outcomes observed in its 40-school peer group.

Peer ranges tend to be broadly similar for schools with similar peer indices, as seen in Figures 4.2 and 6.2. However, the peer range will be sensitive to the specific schools included in one's peer group, especially when included schools have atypical values. As a result, it is possible for the peer range (and in turn, the peer scores) of neighboring schools to differ in unintended and idiosyncratic ways.

Figures F.1 and F.2 provide an illustration for a narrow band of schools on the peer index—the 92 schools with a peer index between 2.0 and 2.25. Figure F.1 shows the peer range width for 4-year graduation rates for these schools (the peer maximum minus the peer minimum), while F.2 provides the same information for Weighted Regents Passing Rates in math.

In Figure F.1, schools in circle 1 and circle 2 have approximately the same peer index (2.16 and 2.17, respectively). Yet the range of peer graduation rates is nearly 8 percentage points wider for the topmost school in circle 1 versus the bottommost school in circle 2. Even schools with the *same* peer index experience variability in peer range width, as the seven schools with a peer index of 2.17 illustrate. In this example, despite having nearly the same peer group/peer index, neighboring schools are benchmarked against very different peer ranges. In such cases, it is possible for similar schools with the same outcome to receive quite different peer scores, depending on idiosyncratic variation in the composition of their peer group.

Whether or not these variations in peer groups (and peer ranges) have a significant effect on the scores of affected schools in practice is not something we explore here. However, we point out that comparisons based on 40-school groups are inherently sensitive to minor differences in the peer group, especially when the included schools are dissimilar.

Figure F.1: Width of peer range for schools with peer index between 2.0 and 2.25, 4-year graduation rate

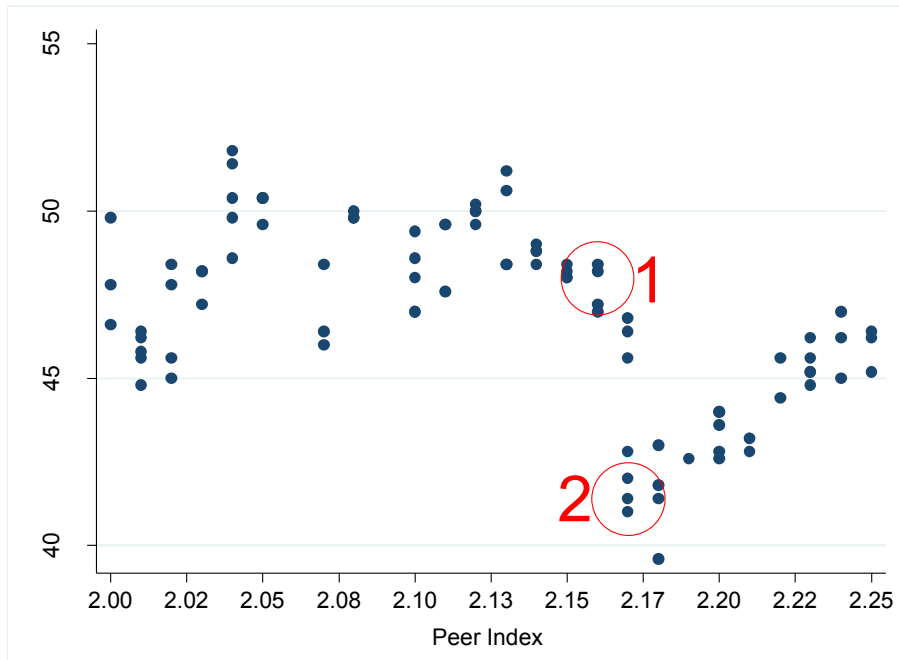
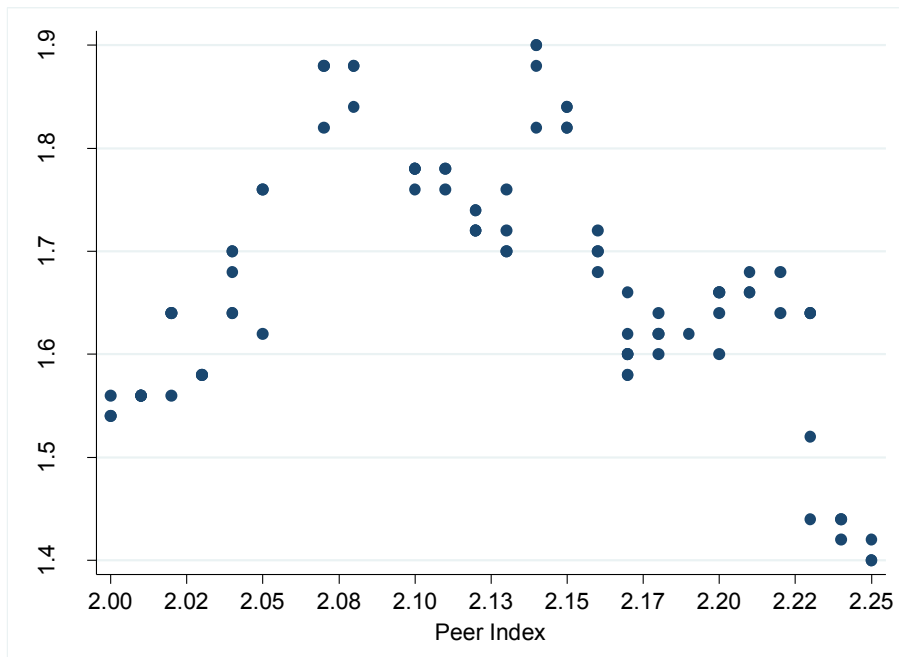


Figure F.2: Width of peer range for schools with peer index between 2.0 and 2.25, Weighted Regents Passing Rate – Math



¹ These figures show the percent of *9th graders* with no 8th grade test score reported in the past three years. (This allows for 9th grade repeaters). These percentages would likely be higher if 10th-12th graders were included.

² “Recent immigrant” does not always imply academic disadvantage. In fact in many schools these students progress at a rate faster than native-born students. Schools with these students fall into the “School B” archetype.

³ For the overall Progress Report score, we find that schools 1 standard deviation higher in missing 8th grade reading scores (18.4 percentage points) had overall scores that were 5 points higher on average. (This controls for the peer index and a set of other school characteristics).