

# UNSATISFACTORY:

The Distribution of Teacher Quality  
in New York City



Students**First**NY





## StudentsFirstNY

is New York's leading voice for students who depend on public education for the skills they need to succeed, but who are too often failed by a system that puts special interests, rather than the interests of children, first. We believe that the students of New York State and New York City cannot wait one minute more for an end to broken policies that stand in the way of ensuring a quality teacher in every classroom and an end to tired rhetoric that serves only to delay meaningful fixes to the problems in our schools.

Learn more at:

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

# UNSATISFACTORY:

## The Distribution of Teacher Quality in New York City

January 2013

---

# EXECUTIVE SUMMARY

---

The importance of highly effective teachers to the life outcomes of their students was cast in stark relief by a December 2011 study, *The Long-Term Impacts of Teachers*, by Raj Chetty and John N. Friedman of Harvard University and Jonah E. Rockoff of Columbia University. Their analysis found that students assigned consistently to higher quality teachers are more likely to go to college, go to better colleges, earn higher salaries and live in more affluent neighborhoods – and are less likely to have children as teenagers.

While the challenges that students face outside of school can create powerful obstacles to academic success, these findings demonstrate the importance of highly effective teaching in any child's quest to overcome those challenges. This should serve as a clarion call for schools to provide a quality teacher to every student, and gives particular urgency to the need to do so in the classrooms of those students for whom a great education is their best chance at the American dream.

How is teacher quality distributed across New York City public schools? That's the question that StudentsFirstNY sought to answer by analyzing teacher evaluation data provided (without identifying information) by the New York City Department of Education for the 2011-2012 school year.<sup>1</sup> We set out to determine whether the distribution of ineffective teachers, as represented by the relative few who were rated "Unsatisfactory," was reasonably even across all schools, or whether it disproportionately affected certain schools or student populations. (Below, we variously use the terms "U-rating" and "U-rated teacher" for shorthand.)

U-rated teachers, just 2.78% of all teachers in our data set of 1,509 schools, almost certainly represent a subset of the universe that would be deemed ineffective under a more rigorous and robust evaluation system, such as the new system required by state law. But for our purposes, U-rated teachers serve as a reasonable proxy for ineffective teachers, generally. Indeed, 62% of U-rated math teachers had value-added scores (as calculated by the City's Department of Education) in the bottom quartile in the 2009-2010 school year – compared to 22% of teachers with "Satisfactory" ratings. In the State Department of Education's calculation of value-added scores (for elementary and middle school math and English teachers in the 2011-2012 school year), teachers rated "ineffective" were more than twice as likely as their peers to be U-rated. Given the social disincentives within schools that discourage principals from going so far as to label teachers Unsatisfactory, the pool of U-rated teachers likely includes many of the worst performers in the system.<sup>2</sup>

While there is large school-to-school variation (principals in 748 schools, for example, did not give out a single U-rating in the 2011-2012 school year), discernible patterns in the distribution of U-rated teachers across groups of schools would validate the notion that the ratings are not, overall, randomly or arbitrarily assigned. On the contrary, such patterns may indicate underlying inequities in the distribution of teacher talent that would be even more clearly – and alarmingly – illuminated by the results of a better evaluation system.

Our analysis was straightforward. We selected independent variables to represent poverty rates, racial composition and achievement levels, respectively, of student populations in schools. We then examined the relationship between schools' U-rating rates and their student populations – grouping similar schools with respect to the variable under examination, and assessing the prevalence of U-rated teachers within each group.

In every case, our findings allowed us to reject a null hypothesis that there was no difference in the prevalence of U-rated teachers between schools grouped according to poverty, race or student achievement levels. Instead, we found support for the notion that teacher talent is distributed in a way that we would consider regressive:

**POVERTY:** Schools with the highest levels of student poverty had a disproportionately high number of U-rated teachers;

**RACE:** Schools with the highest percentages of students of color had a disproportionately high number of U-rated teachers;

**ELEMENTARY AND MIDDLE SCHOOL STUDENT ACHIEVEMENT:** Elementary schools and middle schools with the lowest rates of student proficiency had a disproportionately high number of U-rated teachers; and

**HIGH SCHOOL STUDENT ACHIEVEMENT:** High schools with the lowest rates of college readiness had a disproportionately high number of U-rated teachers.

**In other words: Overall, students with the greatest need for our best educators appear to be least likely to have access to them.**

An ineffective teacher in any classroom is a failure of the system at the expense of the children in that classroom. Ineffective teachers in many classrooms is a big failure. And concentrations of ineffective teachers serving specific student populations is an injustice. We have an obligation to students to seek remedies to these problems, but New York City schools are virtually powerless to do so in the current policy environment.

To change this paradigm, we recommend several policy shifts:

- ☑ Implement the State's more robust and comprehensive teacher evaluation system;
- ☑ Make it easier for top college graduates to enter teaching, and provide financial incentives for them to do so;
- ☑ Provide significant salary increases to highly effective teachers who stay in the classrooms of high-needs schools;
- ☑ Empower principals as managers by reforming the outmoded State law that grants tenure virtually automatically to all teachers after three years, while they are still learning their craft;
- ☑ Impose a cap on how many ineffective teachers may be allowed to remain at any one school year after year, require parental consent for a student to be taught by an ineffective teacher, and prohibit schools from assigning to the class of an ineffective teacher any student taught by an ineffective teacher in the previous year;
- ☑ Require annual reporting by the New York City Department of Education on the distribution of teacher quality across schools and student populations;
- ☑ Hold principals and schools accountable for rigorous implementation of the new evaluation system; and
- ☑ Continue to expand high-quality school options for families, and remove address-driven restrictions that segregate and consign students to inferior schools.

# KEY FINDINGS

# POVERTY ANALYSIS

Our analysis of schools based on the percentage of students receiving Free and Reduced Price Lunch (FRPL) found that Low Poverty schools (Mean FRPL Percentage of 29.45%) had a Mean U-Rating Rate of 1.14%; Medium Poverty schools (Mean FRPL Percentage of 64.78) had a Mean U-Rating Rate of 3.1%; and High Poverty schools (mean FRPL

Percentage of 85.34%) had a Mean U-Rating Rate of 3.91%.

**In other words, students in High Poverty schools were more than three times as likely to be taught by a U-rated teacher as students in Low Poverty schools.**

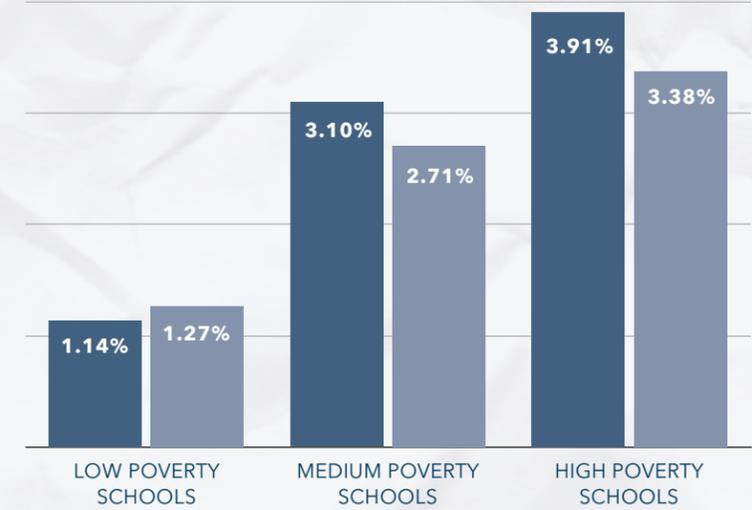
Group	No. of Schools	Mean FRPL Pct.	Mean U-Rating Rate	Group U-Rating Rate	Total Teachers	Proportional Share of U-Rated Teachers	Actual U-Rated Teachers	Pct. of All Teachers	Pct. of U-Rated Teachers	Pct. of Proportional Share
Low Poverty	216	29.45%	1.14%	1.27%	10,307	287	131	15.73%	7.18%	45.66%
Medium Poverty	584	64.78%	3.10%	2.71%	25,834	719	701	39.42%	38.43%	97.48%
High Poverty	709	85.34%	3.91%	3.38%	29,386	818	992	44.85%	54.39%	121.27%
Total	1,509	69.38%	3.20%	2.78%	65,527	1,824	1,824	100.00%	100.00%	100.00%

## Mean and Group Percentages of U-Rated Teachers in Schools

by Poverty Group

**MEAN U-RATING RATE**  
(Average of each school's percentage)

**GROUP U-RATING RATE**  
(Aggregate percentage of the group)



The results of the ANOVA test, the Welch test and the Brown-Forsythe all showed statistically significant results at the p=.05 level. Thus, the null hypothesis can be rejected. Moreover, the Games-Howell post hoc test found a statistically significant difference between each pair of groups at the p=.05 level.

In real terms, Low Poverty schools, comprising 15.73% of all teachers, included just 7.18% of U-rated teachers - less than half (45.66%) of those schools' proportional share.

By contrast, High Poverty schools, comprising 44.85% of all teachers, included 54.39% of U-rated teachers - 121.27% of those schools' proportional share.

## POVERTY ANALYSIS: Groupings of Schools by Percentage of Students Receiving Free and Reduced Price Lunch



PERCENT OF ALL TEACHERS BY POVERTY GROUP



PERCENT OF U-RATED TEACHERS BY POVERTY GROUP



# DEMOGRAPHIC ANALYSIS

Schools with the lowest percentages of black and Hispanic students (Mean Percentage Students of Color of 23.84%) had a Mean U-Rating Rate of 1.06%; schools with medium percentages of black and Hispanic students (Mean Percentage Students of Color of 59.29%) had a Mean U-Rating Rate of 1.99%; and schools with the highest percentage of black and Hispanic students (Mean Percentage

Students of Color of 94.03%) had a Mean U-Rating Rate of 4.13%.

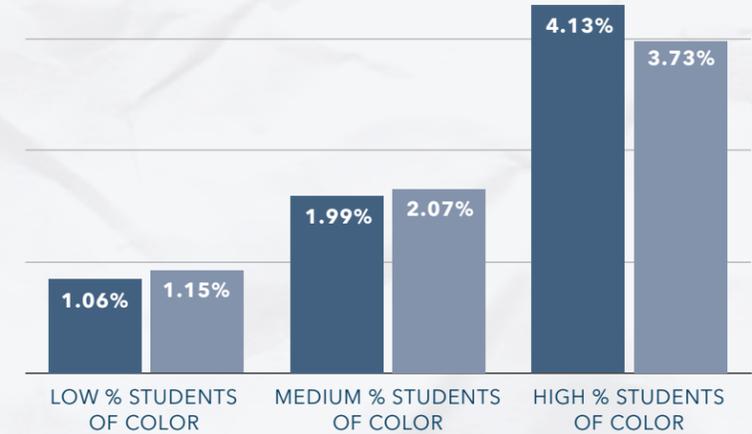
**Students in schools with high percentages of black and Hispanic students were almost four times as likely to be taught by a U-rated teacher as students in schools with far fewer students of color.**

Group	No. of Schools	Mean Pct. Black/Hispanic	Mean U-Rating Rate	Group U-Rating Rate	Total Teachers	Proportional Share of U-Rated Teachers	Actual U-Rated Teachers	Pct. of All Teachers	Pct. of U-Rated Teachers	Pct. of Proportional Share
Low Percent Students of Color	278	23.84%	1.06%	1.15%	14,967	417	172	22.84%	9.43%	41.28%
Medium Percent Students of Color	259	59.29%	1.99%	2.07%	14,184	395	294	21.65%	16.12%	74.46%
High Percent Students of Color	972	94.03%	4.13%	3.73%	36,376	1,013	1,358	55.51%	74.45%	134.12%
Total	1,509	75.13%	3.20%	2.78%	65,527	1,824	1,824	100.00%	100.00%	100.00%

## Mean and Group Percentages of U-Rated Teachers in Schools by Demographic Group

**MEAN U-RATING RATE**  
(Average of each school's percentage)

**GROUP U-RATING RATE**  
(Aggregate percentage of the group)

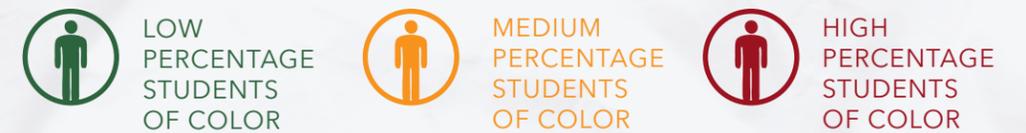


The results of the ANOVA test, the Welch test and the Brown-Forsythe all showed statistically significant results at the  $p=.05$  level. Thus, the null hypothesis can be rejected. Moreover, the Games-Howell post hoc test found a statistically significant difference between each pair of groups at the  $p=.05$  level.

In real terms, the group of schools with the lowest percentage of black and Hispanic students, comprising 22.84% of all teachers, included just 9.43% of U-rated teachers - less than half (41.28%) of those schools' proportional share.

By contrast, the group of schools with large concentrations of students of color comprised 55.51% of all teachers and fully 74.45% of U-rated teachers, or 134.12% of those schools' share.

## DEMOGRAPHIC ANALYSIS: Groupings of Schools by Percentage of Students who are Black or Hispanic



PERCENT OF ALL TEACHERS BY DEMOGRAPHIC GROUP



PERCENT OF U-RATED TEACHERS BY DEMOGRAPHIC GROUP



# ACHIEVEMENT ANALYSIS

## Elementary Schools

High Proficiency elementary schools (Mean Proficiency Percentage of 82.85%) had a Mean U-Rating Rate of 1.32%; Medium Proficiency elementary schools (Mean Proficiency Percentage of 54.15%) had a Mean U-Rating Rate of 3.03%; and Low Proficiency elementary

schools (Mean Proficiency Percentage of 20.07%) had a Mean U-Rating Rate of 4.67%.

Students in Low Proficiency elementary schools were more than three times as likely to be taught by a U-rated teacher as students in High Proficiency schools.

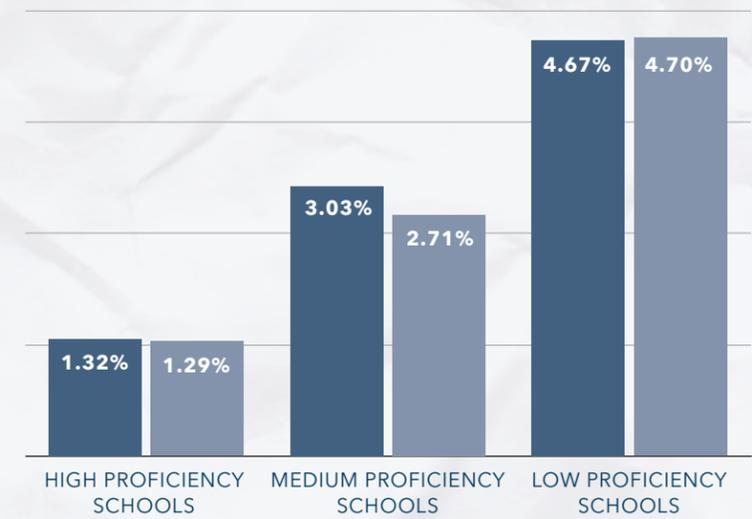
Group	No. of Schools	Mean Proficiency Pct.	Mean U-Rating Rate	Group U-Rating Rate	Total Teachers	Proportional Share of U-Rated Teachers	Actual U-Rated Teachers	Pct. of All Teachers	Pct. of U-Rated Teachers	Pct. of Proportional Share
High Proficiency	333	82.85%	1.32%	1.29%	16,312	365	211	47.55%	27.47%	57.78%
Medium Proficiency	308	54.15%	3.03%	2.71%	14,504	325	393	42.28%	51.17%	121.04%
Low Proficiency	96	20.07%	4.67%	4.70%	3,490	78	164	10.17%	21.35%	209.91%
Total	737	75.13%	2.47%	2.24%	34,306	78	768	100.00%	100.00%	100.00%

## Mean and Group Percentages of U-Rated Teachers in Schools

### by Proficiency Group, 4th Grade Math Exams

**MEAN U-RATING RATE**  
(Average of each school's percentage)

**GROUP U-RATING RATE**  
(Aggregate percentage of the group)



The results of the ANOVA test, the Welch test and the Brown-Forsythe all showed statistically significant results at the p=.05 level. Thus, the null hypothesis can be rejected. Moreover, the Games-Howell post hoc test found a statistically significant difference between each pair of groups at the p=.05 level.

In real terms, High Proficiency elementary schools, comprising 47.55% of all teachers, included just 27.47% of U-rated teachers - 57.78% of those schools' proportional share.

By contrast, Low Proficiency elementary schools, comprising 10.17% of all teachers, included 21.35% of U-rated teachers - more than twice (209.91%) those schools' proportional share.

## ACHIEVEMENT ANALYSIS: Groupings of Elementary Schools By Proficiency on 4th Grade Math Exam



HIGH PROFICIENCY SCHOOLS

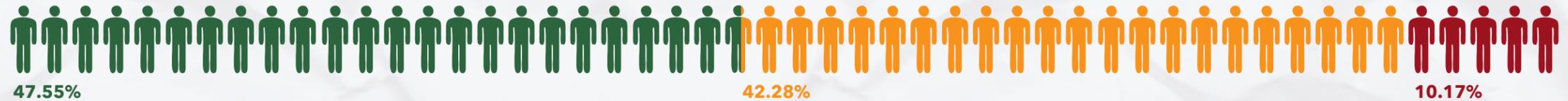


MEDIUM PROFICIENCY SCHOOLS



LOW PROFICIENCY SCHOOLS

### PERCENT OF ALL TEACHERS BY PROFICIENCY GROUP



### PERCENT OF U-RATED TEACHERS BY PROFICIENCY GROUP



### Middle Schools

Middle schools in our analysis had the highest overall Mean U-Rating Rate: 4.46%, compared to 3.7% for high schools in the achievement analysis and 2.47% for elementary schools in the achievement analysis. 3.20% was the Mean U-Rating Rate for all schools in the data set.

High Proficiency middle schools (with a Mean Proficiency Percentage of 75.82%) had a Mean U-Rating Rate of 1.63%; Medium Proficiency middle schools (Mean Proficiency Percentage of 46.05%) had a Mean U-Rating Rate of 5.13%; and

Low Proficiency middle schools (Mean Proficiency Percentage of 20.15%) had a strikingly high Mean U-Rating Rate of 7.06% (and a nearly identical Group U-Rating Rate).

**Students in Low Proficiency middle schools, as defined by this analysis, were more than four times as likely to be taught by a U-rated teacher as students in High Proficiency schools.**

The results of the ANOVA test, the Welch test and the Brown-Forsythe all showed statistically significant results at the p=.05 level. Thus, the null hypothesis can be rejected. While Games-Howell post hoc test

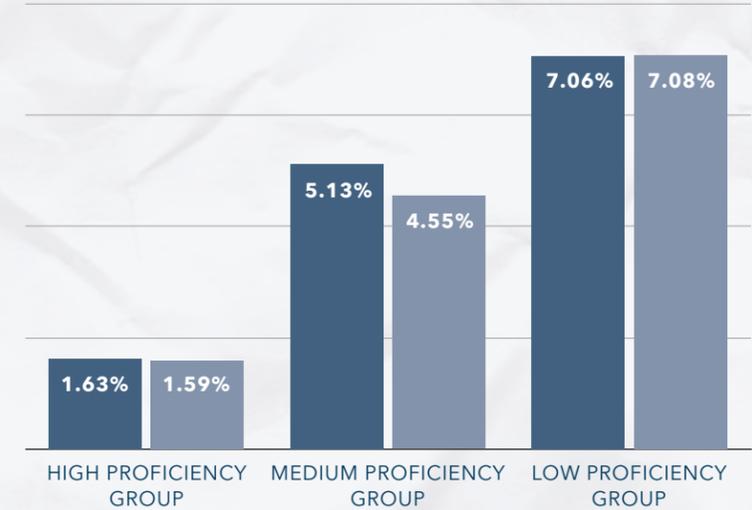
Group	No. of Schools	Mean Proficiency Pct.	Mean U-Rating Rate	Group U-Rating Rate	Total Teachers	Proportional Share of U-Rated Teachers	Actual U-Rated Teachers	Pct. of All Teachers	Pct. of U-Rated Teachers	Pct. of Proportional Share
High Proficiency	94	75.82%	1.63%	1.59%	4,963	187	79	42.44%	17.95%	42.31%
Medium Proficiency	112	46.05%	5.13%	4.55%	4,571	172	208	39.09%	47.27%	120.95%
Low Proficiency	73	20.15%	7.06%	7.08%	2,161	81	153	18.48%	34.77%	188.18%
Total	279	49.30%	4.46%	3.76%	11,695	81	440	100.00%	100.00%	100.00%

## Mean and Group Percentages of U-Rated Teachers in Schools

### by Proficiency Group, 8th Grade Math Exams

**MEAN U-RATING RATE**  
(Average of each school's percentage)

**GROUP U-RATING RATE**  
(Aggregate percentage of the group)



found a statistically significant difference between Low Proficiency and Medium Proficiency schools, and between Low Proficiency and High Proficiency schools, Medium Proficiency and High Proficiency schools were found not to have statistically different means of Unsatisfactory teacher ratings.

In real terms, High Proficiency middle schools, comprising 42.44% of all teachers, included just 17.95% of U-rated teachers - less than half (42.31%) of those schools' proportional share.

By contrast, Low Proficiency middle schools, comprising 18.48% of all teachers, included 34.77% of U-rated teachers - 188.18% of those schools' proportional share.

## ACHIEVEMENT ANALYSIS: Groupings of Middle Schools By Proficiency on 8th Grade Math Exam



### PERCENT OF ALL TEACHERS BY PROFICIENCY GROUP



### PERCENT OF U-RATED TEACHERS BY PROFICIENCY GROUP



### High Schools

High College Readiness schools (Mean College Readiness Rate of 86.66%) had a Mean U-Rating Rate of 1.94%; Medium College Readiness schools (Mean College Readiness Rate of 41.7%) had a Mean U-Rating Rate of 2.73%; and Low College Readiness schools (Mean College Readiness Rate of 8.35%) had a Mean U-Rating Rate of 4.11%.

Students in high schools with Low College Readiness rates were more than twice as likely to be taught by a U-rated teacher as students in schools with High College Readiness rates.

The results of the ANOVA test, the Welch test and the Brown-Forsythe all showed statistically significant results at the p=.05 level. Thus, the null hypothesis can be rejected. While

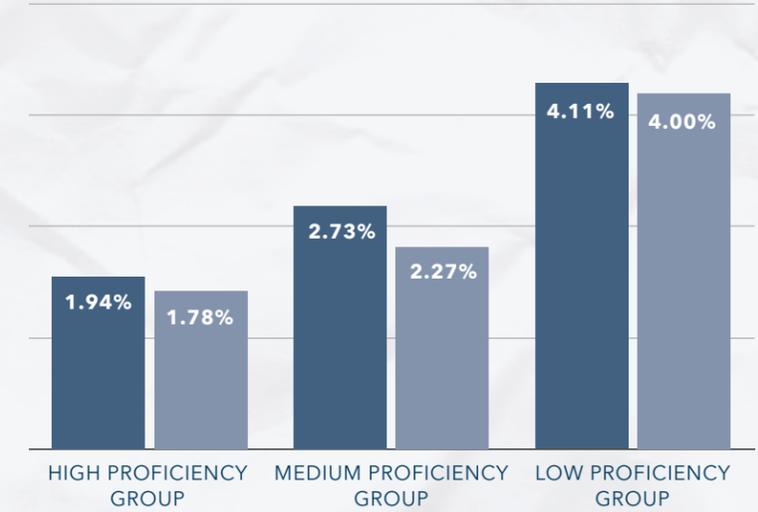
Group	No. of Schools	Mean Proficiency Pct.	Mean U-Rating Rate	Group U-Rating Rate	Total Teachers	Proportional Share of U-Rated Teachers	Actual U-Rated Teachers	Pct. of All Teachers	Pct. of U-Rated Teachers	Pct. of Proportional Share
High College Readiness	28	86.66%	1.94%	1.78%	1,685	55	30	9.10%	4.99%	54.87%
Medium College Readiness	91	41.70%	2.73%	2.27%	5,935	193	135	32.04%	22.46%	70.11%
Low College Readiness	334	8.35%	4.11%	4.00%	10,903	354	436	58.86%	72.55%	123.25%
Total	453	19.89%	3.70%	3.24%	18,523	354	601	100.00%	100.00%	100.00%

## Mean and Group Percentages of U-Rated Teachers in Schools

### by College Readiness Rate, High School

**MEAN U-RATING RATE**  
(Average of each school's percentage)

**GROUP U-RATING RATE**  
(Aggregate percentage of the group)



Games-Howell post hoc test found a statistically significant difference between Low College Readiness and Medium College Readiness schools, and between Low College Readiness and High College Readiness schools, Medium College Readiness and High College Readiness schools were found not to have statistically different means of Unsatisfactory teacher ratings.

In real terms, High College Readiness schools, comprising 9.10% of all teachers, included just 4.99% of U-rated teachers - 54.87% of those schools' proportional share.

By contrast, Low College Readiness schools, comprising 58.86% of all teachers, included 72.55% of U-rated teachers - 123.25% of those schools' proportional share.

## ACHIEVEMENT ANALYSIS: Groupings of High Schools By College Readiness



HIGH COLLEGE READINESS SCHOOLS



MEDIUM COLLEGE READINESS SCHOOLS



LOW COLLEGE READINESS SCHOOLS

### PERCENT OF ALL TEACHERS BY COLLEGE READINESS GROUP



### PERCENT OF U-RATED TEACHERS BY COLLEGE READINESS GROUP

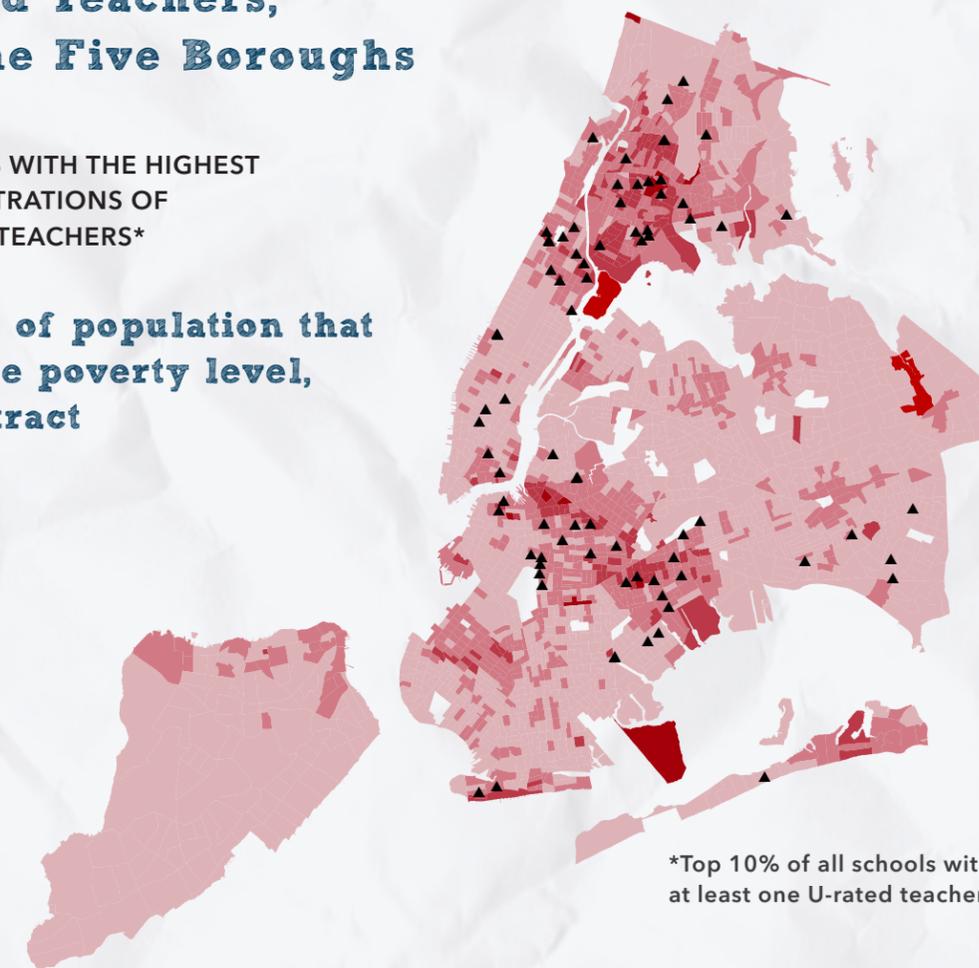
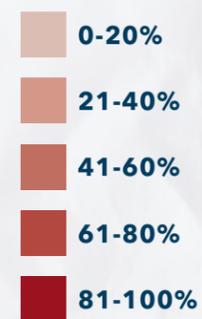


# CONCLUSION AND RECOMMENDATIONS

## Schools with the Highest Percentages of U-Rated Teachers, Across the Five Boroughs

▲ SCHOOLS WITH THE HIGHEST CONCENTRATIONS OF U-RATED TEACHERS\*

Percentage of population that is below the poverty level, by census tract



\*Top 10% of all schools with at least one U-rated teacher

Based on the findings of this study, the distribution of teachers with Unsatisfactory ratings is, in our view, fundamentally regressive. Schools with the highest rates of poverty and the lowest rates of student achievement have substantially more than their fair share of teachers with Unsatisfactory ratings, as do schools with high concentrations of students of color. Conversely, wealthier, higher-achieving schools have fewer U-rated teachers.

With just 3% of New York City teachers receiving U-ratings, the disparities identified in this report translate into relatively small differences in raw numbers. But if U-ratings are the tip of the teacher quality iceberg, then the incidence of U-rated teachers at high-poverty schools, for example, at a rate more than three times the rate at low-poverty schools may indicate a broader inequity that harms many New York City schoolchildren each year.

The scope of the problem is impossible to assess with the data we have, but that should not be necessary to agree upon its unfairness to children and implement solutions. And to the extent that variation in the

data amplifies less extreme underlying disparities, it is likely because of the under-assignment, not the over-assignment, of U-ratings in certain schools - either because principals have decided that it isn't worth the trouble to U-rate ineffective teachers (absent real authority to remove them), or because principals have been lulled into a sense of complacency by decent achievement scores at their schools, which might simply be the result of serving more affluent students. In other words, if the disparities narrow as the data fill out, it will almost certainly be because fewer schools are exempt from teacher quality challenges than the U-rating data would indicate.

Regardless, the data undeniably show that many principals faced with high numbers of ineffective teachers are right now trying to effect change by identifying low-performers, using the only rating system they have, despite its limitations. Yet few of these U-rated teachers will be removed from the classroom, protected by law and contract. To highlight what these principals are dealing with, we ranked all schools with at least one U-rating, and looked at the 10 percent - 77 schools - at the top of the list. Overall, these schools had a U-rating rate of 18.64% - compared to a citywide rate of 3.20%. (They are mapped above.)

New York City schools do not target resources to recruit highly effective teachers, to retain them, or to distribute them in a progressive way. At the same time, there are natural incentives driving a regressive distribution of teacher talent: Affluent schools are far more attractive places to work for most teachers, and we suspect that, as a result, these schools wind up with a higher quality pool of applicants for each teaching vacancy. Strong incentives, focused on our best educators, are necessary to offset this effect.

Alarming, the characteristics of these schools underscore the findings of our study. While black and Hispanic students make up 68% of all New York City public school students in our data set, they make up 95% of the students in these schools. While 67% of all students are eligible for Free and Reduced Price Lunch, 78.4% of students in these schools are eligible. And, needless to say, these are, as a group, places where students are achieving at depressingly low levels.

Among these were schools where fully one-third of the staff was rated Unsatisfactory. Imagine the inadequacy of the education being provided in a school whose principal is willing to absorb the blowback that comes from rating one-third of his or her staff Unsatisfactory. The sad truth is that there are almost certainly schools in the same sinking educational boat that aren't on this list.

Allowing this to go on, without action, is unconscionable. Policy relief must be provided.

First, the current evaluation system - while providing revealing data about systemwide trends - creates nothing close to a robust picture of the strengths and weaknesses of individual teachers. Implementation of the State's more comprehensive evaluation and feedback system still awaits the resolution of precarious negotiations between the City and its teachers union. Success in those negotiations is a necessary and critical first step to addressing teacher quality deficiencies, generally, and will create an important tool to address inadequacy and inequity in teacher talent and the way it is distributed across schools.

**RECOMMENDATION #1: Implement the State's robust and comprehensive teacher evaluation system, to provide a clearer picture of teacher quality in City schools and important tools to improve the status quo.**

Second, the State law on evaluations makes clear that measuring teacher quality is not enough; the information must be acted upon:

Such annual professional performance reviews shall be a significant factor for employment decisions including but not limited to, promotion, retention, tenure determination, termination, and supplemental compensation, which decisions are to be made in accordance with locally developed procedures negotiated pursuant to the requirements of article fourteen of the civil service law (New York Education Law 3012-c).

But none of this has come to pass; of the list of "employment decisions," only termination will be affected by implementation of the new evaluation system. New York City schools do not target resources to recruit highly effective teachers, to retain them, or to distribute them in a progressive way. While the City proposed substantial salary increases for highly effective teachers, the idea was declared dead on arrival by the teachers union, whose agreement is necessary under the State law. The school system of Newark, by contrast, just approved a new collective bargaining agreement that focuses on these imperatives by tying all salary increases to teacher quality and providing substantial bonuses to highly effective teachers in high-needs schools and hard-to-staff subjects.

At the same time, there are natural incentives driving a regressive distribution of teacher talent: Affluent schools - where students enter the classroom with fewer external challenges - are far more attractive places to work for most teachers, and we suspect that, as a result, these schools wind up with a higher quality pool of applicants for each teaching vacancy.<sup>3</sup> Strong incentives, focused on our best educators, are necessary to offset this effect.

**RECOMMENDATION #2: Provide financial incentives to encourage top college graduates to enter the teaching profession, and ease the regulatory obstacles that stand in their way.**

**RECOMMENDATION #3: Consistent with State law, provide significant salary increases to highly effective teachers who stay in the classrooms of high-needs schools.**

Also, New York City principals have few tools and little management authority to address teacher quality deficiencies once they have taken root in schools. While the City's Department of Education has gone to great lengths to work around a State law that awards tenure virtually automatically (to all "persons who have been found competent, efficient and satisfactory"), that law remains an enormous handicap to the work of principals as managers of talent, substantially eroding authority that is an intrinsic part of almost all relationships between managers and their staffs in professional settings.

And, of course, the difficulty of removing ineffective teachers, once they are in the classroom and tenured, has been well established. There is some hope that the new evaluation system will provide relief through a lower standard for termination based on incompetence, but that standard applies only after teachers have spent two years ineffectively educating children, and substantial bureaucratic bottlenecks loom.

**RECOMMENDATION #4: Replace the outmoded State law that grants tenure virtually automatically after just three years with one that emphasizes the retention of highly effective teachers and empowers principals to manage and improve the quality of instruction in their schools.**

**RECOMMENDATION #5: Enact additional reforms in State law and collective bargaining agreements to ensure that ineffective teachers do not remain in the classroom.**

These are tools that we believe can help the system and individual principals meaningfully address teacher talent deficiencies and inequities. But measures should also be put in place as backstops to protect students and ensure that the tools actually produce results.

**RECOMMENDATION #6: Require parental consent for any child to be assigned to the class of an ineffective teacher, and prohibit schools from assigning to the class of an ineffective teacher any student taught by an ineffective teacher the previous year.**

**RECOMMENDATION #7: Require annual reporting by the New York City Department of Education on the distribution of teacher quality across schools and student populations.**

**RECOMMENDATION #8: Impose a cap on how many ineffective teachers may be allowed to remain in any individual school year after year.**

**RECOMMENDATION #9: Make the diligent, rigorous and student-focused implementation of the new evaluation system a component of principal and school-based accountability measures, including principal evaluations, school progress reports and quality reviews.**

Finally, aiming to put a highly effective teacher in every classroom and ensure comprehensive educational equity in every school should be bedrock principles on which education policy is made. While we continue to strive toward these noble and important goals, until they are realized it is indefensible to force certain students to attend schools that are demonstrably unequal and subpar – where, say, one-third of the teaching staff has been deemed Unsatisfactory – when their peers have access to something better simply because they live in a different neighborhood. We must continue to expand policies that give families real options.

**RECOMMENDATION #10: Continue to expand high-quality school options, and remove address-driven restrictions that segregate and consign students to inferior schools.**

It is beyond argument that students benefit from consistent access to highly effective teachers – and that they are educationally harmed by ongoing exposure to ineffective pedagogues. Critics may wish to distract from this self-evident truth by emphasizing other causes of student success or failure, or they may dispute the magnitude of the impact of teacher quality – but they cannot dispute that it matters, meaningfully, in the lives of our children.

The further notion that we are concentrating our most ineffective teachers – whatever the numbers – in the classrooms of poor children, children of color, and children who face the greatest educational challenge should provoke anger.

We hope it will also provoke action.

# APPENDIX

# METHODOLOGY

Any analysis of Unsatisfactory vs. Satisfactory ratings to assess patterns and relationships must confront and solve for an enormous amount of school-to-school variation in the assignment of U-rated teachers. Indeed, 748 schools in our data set - with a wide array of school characteristics - had no U-ratings at all. Regression analyses found that while race (which is highly correlated with poverty) and student achievement were statistically significant predictors of the number of U-ratings in a school, they accounted for just 15% of all variation amongst schools with at least one U-rating, and approximately 11% of the variation amongst all schools.

At the same time, the statistical significance of school characteristics as predictors of U-ratings tells us that the ratings are not, in the aggregate, randomly assigned. So, too, do the relationships between City and State value-added data and U-ratings. This is consistent with other research that shows strong relationships between the qualitative observations of principals and other measures of teacher quality.

To get a better sense of any relationship between school characteristics - specifically students' poverty levels, race and achievement levels - and the prevalence of U-rated teachers, one must look across schools to create groups that, in effect, smooth out the school-to-school variations that exist assignment of U-ratings. (Unlike for race and poverty, no single variable could be used to

represent student achievement across elementary, middle and high schools. We therefore conducted three different analyses, using a different variable for each school level, to explore student achievement.)

Specifically, the independent variables we considered in grouping schools were<sup>4</sup>:

**Poverty:** the percentage of students receiving Free and Reduced Price Lunch (FRPL) at each school;

**Race:** the combined percentage of black and Hispanic students at each school;

**Student achievement (elementary school level):** the combined percentage of students at each elementary school (includes K-8 and K-12 programs) scoring at a Level 3 or Level 4 on the 2011-2012 New York State 4th Grade Math Examination;

**Student achievement (middle school level):** the combined percentage of students at each middle school scoring a Level 3 or Level 4 on the 2011-2012 New York State 8th Grade Math Examination; and

**Student achievement (high school level):** the college readiness rate for each high school (secondary schools were included in the high school grouping).

For each of these independent variables, we initially ranked schools and created ten groupings. We saw stark differences moving across deciles in the prevalence of U-rated teachers. But the uneven distribution of the independent variables meant that many of the groups were not, in fact, statistically different.

With no definitive theoretical basis from the literature for dividing schools into a specific number of groups, three groups were decided upon and created for each independent variable

to ensure simplicity and a parsimonious analysis. A K-means cluster analysis was conducted to ensure that, for each independent variable, the groups were as different as possible based solely upon the variable under examination. "Analysis of variance" ("ANOVA") tests were run to ensure that the group means were in fact statistically different, and the tests confirmed these assumptions for each variable. The groupings for income, race and student achievement, and their basic descriptive statistics, are below.

## POVERTY ANALYSIS:

### Groupings of schools by percentage of students receiving Free and Reduced Price Lunch (FRPL)

Group	No. of Schools	Mean FRPL Pct.	Std. Deviation	Median FRPL Pct.	Minimum FRPL Pct.	Maximum FRPL Pct.	Total Teachers	Pct. of All Teachers
Low Poverty	216	29.45%	0.11291	30.60%	4.00%	47.00%	10,307	15.73%
Medium Poverty	584	64.78%	0.07149	65.58%	47.00%	75.00%	25,834	39.42%
High Poverty	709	85.34%	0.06405	85.11%	75.00%	100.00%	29,386	44.85%
<b>Total</b>	<b>1,509</b>	<b>69.38%</b>	<b>0.20334</b>	<b>73.82%</b>	<b>4.00%</b>	<b>100.00%</b>	<b>65,527</b>	<b>100.00%</b>

## DEMOGRAPHIC ANALYSIS:

### Groupings of schools by percentage of students who are black or Hispanic

Group	No. of Schools	Mean Pct. Black/Hispanic	Std. Deviation	Median Pct. Black/Hispanic	Minimum Pct. Black/Hispanic	Maximum Pct. Black/Hispanic	Total Teachers	Pct. of All Teachers
Low Pct. Students of Color	278	23.84%	0.09619	22.79%	4.00%	41.00%	14,967	22.84%
Med. Pct. Students of Color	259	59.29%	0.10813	59.05%	42.00%	77.00%	14,184	21.65%
High Pct. Students of Color	972	94.03%	0.05259	95.90%	77.00%	100.00%	36,376	55.51%
<b>Total</b>	<b>1,509</b>	<b>75.13%</b>	<b>0.28515</b>	<b>91.41%</b>	<b>4.00%</b>	<b>100.00%</b>	<b>65,527</b>	<b>100.00%</b>

**ACHIEVEMENT ANALYSIS:**

**Groupings of elementary schools by proficiency on 4th grade math exam**

Group	No. of Schools	Mean Proficiency Pct.	Std. Deviation	Median Proficiency Pct.	Minimum Proficiency Pct.	Maximum Proficiency Pct.	Total Teachers	Pct. of All Teachers
High Proficiency	333	82.85%	0.08710	82.89%	69.00%	100.00%	16,312	47.55%
Medium Proficiency	308	54.15%	0.08301	54.42%	37.00%	68.00%	14,504	42.28%
Low Proficiency	96	20.07%	0.14577	24.81%	0.00%	37.00%	3,490	10.17%
<b>Total</b>	<b>737</b>	<b>75.13%</b>	<b>0.23274</b>	<b>65.00%</b>	<b>0.00%</b>	<b>100.00%</b>	<b>34,306</b>	<b>100.00%</b>

**ACHIEVEMENT ANALYSIS:**

**Groupings of middle schools by proficiency on 8th grade math exam**

Group	No. of Schools	Mean Proficiency Pct.	Std. Deviation	Median Proficiency Pct.	Minimum Proficiency Pct.	Maximum Proficiency Pct.	Total Teachers	Pct. of All Teachers
High Proficiency	94	75.82%	0.10532	73.99%	61.00%	100.00%	4,963	42.44%
Medium Proficiency	112	46.05%	0.07381	45.41%	34.00%	60.00%	4,571	39.09%
Low Proficiency	73	20.15%	0.10292	22.57%	0.00%	33.00%	2,161	18.48%
<b>Total</b>	<b>279</b>	<b>49.30%</b>	<b>0.23486</b>	<b>48.18%</b>	<b>0.00%</b>	<b>100.00%</b>	<b>11,695</b>	<b>100.00%</b>

**ACHIEVEMENT ANALYSIS:**

**Groupings of high schools by college readiness**

Group	No. of Schools	Mean Proficiency Pct.	Std. Deviation	Median Proficiency Pct.	Minimum Proficiency Pct.	Maximum Proficiency Pct.	Total Teachers	Pct. of All Teachers
High College Readiness	28	86.66%	0.01940	91.35%	64.80%	100.00%	1,685	9.10%
Medium College Readiness	91	41.70%	0.02730	38.30%	25.30%	63.70%	5,935	32.04%
Low College Readiness	334	8.35%	0.04110	7.60%	0.00%	24.30%	10,903	58.86%
<b>Total</b>	<b>453</b>	<b>19.89%</b>	<b>0.03700</b>	<b>11.30%</b>	<b>0.00%</b>	<b>100.00%</b>	<b>18,523</b>	<b>100.00%</b>

We then calculated a U-rating rate for each school: the number of U-rated teachers divided by the total number of teachers. These rates were averaged for the schools in each group, for each variable, to create a “Mean U-Rating Rate.” (Because all schools in a group, regardless of size, are weighted equally calculating the Mean U-Rating Rate, that rate is different from the “Group U-Rating Rate,” which was calculated by dividing the group’s total number of U-rated teachers by its total number of teachers.)

For each variable’s set of groups, the null hypothesis that was tested was that there was no difference in the Mean U-Ratings Rate among each of the three groups. An ANOVA test was run to test this null hypothesis and determine if, indeed, the Mean U-Rating rates were statistically different. For each set, it was noted that the homogeneity of

variances assumption was violated and thus the Welch and Brown-Forsythe statistics were also examined.

Post hoc tests were then conducted to look specifically at the differences between each group. The Games-Howell post hoc test was specifically chosen because of its robustness against violations of the homogeneity of variances assumption.

In addition, we compared each group’s proportional share of all teachers to its proportional share of U-ratings to get a sense for the relative burden each group bears with regard to ineffective teachers.

# STUDYING TEACHER TALENT DISTRIBUTION

The near-universal validation indicated by the assignment of Satisfactory ratings to 97% of teachers in a school system with such glaring gaps in student achievement has led to the deserved maligning of New York City's existing, binary method of evaluating teachers. It is true that not much can be gleaned from the sweeping assignment of Satisfactory ratings and that the binary system is a failure as a management tool. Yet it seems reasonable to assume that the miniscule number of teachers who principals deem sufficiently incompetent to receive an Unsatisfactory rating are representative of a larger group of ineffective teachers, an assumption supported by the strong relationship between U-ratings and low value-added scores, by research from other jurisdictions<sup>5</sup>, and by common sense. This is why we believe that this system still has something to tell us.

Some may attempt to dismiss our findings, pointing to the long overdue movement away from the U. vs. S. system to one that is more robust, and to the large school-to-school variation in U-rating rates. But the suggestion that U-ratings are, generally speaking, randomly or arbitrarily assigned – and therefore that the ratings indicate nothing about teacher talent distribution – is belied by the very patterns that emerge in this study.

It has also been suggested that there are principals who assign high numbers of U-ratings, without merit, to their faculty members, driven by malice or retaliation. While this claim is difficult to assess, with just 3% of teachers receiving Unsatisfactory

ratings, we think it almost certain that we have the opposite problem: that, in fact, there are far more undeserved Satisfactory ratings assigned each year. Given the difficulty of terminating an ineffective teacher, there is little incentive for a principal to give one of his or her faculty members an Unsatisfactory rating. Doing so often involves the loss of substantial time and effort to bureaucratic procedure, as well as the loss of social capital within a school – not inconsequential disincentives, given the low probability of any resultant action. Indeed, just 11% of principals assigned Unsatisfactory ratings to more than 10% of their teachers.

With the data currently available, and without a better mode of evaluating teacher practice in place, it is hard to empirically test just how representative the U-rating data is of ineffective teaching, generally (beyond examining the relationship between U-ratings and value-added measures). If New York City implements the State's new teacher evaluation system, and begins employing the Danielson Rubric for qualitative review of teacher practice on a citywide basis, researchers will have the means to examine relationships between the old system and the new – and also to explore all of the questions raised by this study with far richer data.

Interestingly, the “U. vs. S” system, despite its limitations, is likely of greater use in exploring the question of teacher talent distribution than value-added measures that aim to quantify each teacher's contribution to student learning.

Because value-added measures are driven by student performance on test scores, and such performance is highly correlated to student characteristics, architects of value-added measures must wrestle with the question of endogeneity: How much of measured outcomes are driven by student characteristics rather than teacher impact? As a result, most value-added measures – such as those used in the past by New York City and presently by New York State – control heavily for student, classroom and school characteristics, and compare teachers against their peers in similar schools. Almost by definition, differences that track with school characteristics are wiped away as these measures compare teachers to their peers in similar schools.

Indeed, critics of value-added methodology have charged that the State hasn't gone far enough in this regard, and that under the State's recently released student growth measures, “teachers in classrooms with high numbers of poor or disabled students tended to get slightly lower ratings” (“Report finds potential problems with new state rating system for teachers,” *New York Post*, December 11, 2012).

State officials dispute the charge, and point out that the bottom-line finding of the study referenced by critics was one of fairness in the State's methodology. They also point out that “the fact that teachers in high-performing classes got marginally better growth scores than their peers supports a recent study that found that the best teachers often work in top-notch classrooms or schools.” (This may refer to a 2006 Education Trust study, *Teaching Inequality: How Poor and Minority Students Are Shortchanged on Teacher Quality*, by Heather G. Peske and Kati Haycock, whose findings from other states and jurisdictions were consistent with our analysis of New York City schools.)

This back-and-forth, however, underscores just why qualitative measures are necessary to compare teacher performance in, for example, high-poverty and low-poverty schools. While different kind of schools may attract or promote different styles of practice, there's no reason that student characteristics, across schools, should be related to practice that is better or worse, generally – unless teacher talent is being distributed inequitably.

---

# ENDNOTES

---

<sup>1</sup> To create fair comparisons, we removed from the analysis schools for which the Department of Education did not provide demographic information. This includes most District 75 schools, which serve “students who are on the autism spectrum, have significant cognitive delays, are severely emotionally challenged, sensory impaired and/or multiply disabled.” The remaining data set comprised 1,509 schools.

<sup>2</sup> A spotlight was thrown on this subject, and the difficulties facing principals in rating and removing ineffective teachers, by journalist Steven Brill in “The Rubber Room: The battle over New York City’s worst teachers,” *The New Yorker*, August 31, 2009.

<sup>3</sup> *The New Teacher Project*, in its report on teacher retention, *The Irreplaceables*, highlighted the negative impact of poor working conditions on teacher retention. Needless to say, this is a cause that should be directly addressed, in addition to the creation of incentives to offset it, and the challenges that are inherent in teaching at even the best schools serving high-needs students.

<sup>4</sup> Where data was not available for a given school, that school was excluded from the analysis. For example, a handful of elementary schools, which taught grades K-2 or K-3, were excluded. 43 high schools, primarily because of their newness, do not yet have college readiness rates.

<sup>5</sup> *Gathering Feedback for Teaching, Combining High-Quality Observations with Student Surveys and Achievement Gains*, Bill and Melinda Gates Foundation, January 2012.

Learn more at:  
[www.StudentsFirstNY.org](http://www.StudentsFirstNY.org)

