Creating a Metric of Educational Opportunity for Assisted Households

Assisted Housing Research Cadre Report

Creating a Metric of Educational Opportunity for Assisted Households

Prepared for:
U.S. Department of Housing and Urban Development
Office of Policy Development and Research

Authors:
Ingrid Gould Ellen
Keren Mertens Horn
Furman Center for Real Estate and Urban Policy

February 2011

Assisted Housing
Research Cadre Report

ACKNOWLEDGMENTS

We thank HUD PD&R staff, especially Barbara Haley, David Chase, and Barry Steffens for their help and guidance. We also thank Alan Biller, Mary Farnsworth, Molly Greer, and Emily Osgood for excellent research assistance and Furman Center seminar participants for useful insights and feedback.

Preface

HUD's strategic plan identifies the use of "housing as a platform for improving quality of life" as one of its five strategic goals. It further establishes a sub-goal to improve educational outcomes and early learning and development for children in HUD-assisted housing. This research, *Creating a Metric of Educational Opportunity for Assisted Households,* is intended to advise HUD about how to use readily available data to create a metric for school quality. This metric is the measure of success in providing "access to schools scores at or above the local average" for children in assisted households. A suitable metric must be consistent across states and metropolitan areas and available nationwide.

The data that are readily available are reports to the U.S. Department of Education regarding individual schools' percentages achieving 'proficiency' in reading and math. In the near term, comparing the test scores of nearby schools to other schools in the area can serve as a measure to assess schools. The researchers recommend a ratio that compares the test scores of the elementary schools nearest subsidized households to the test scores of other schools in that same county or metropolitan area, with perhaps a comparison to the schools nearest other renters or low-income households. This ratio should be calculated using five-year American Community Survey (ACS) data at the block-group level, which will produce sufficiently localized population information. Using this local-comparison ratio can overcome differences in state methodologies for evaluating schools, differences in homeownership rates across metropolitan areas, and differences in income levels. This score will allow HUD to identify metropolitan areas to target for mobility efforts and to track progress over time.

The first problem addressed by the researchers was identification of schools attended by children living in assisted housing. The researchers suggest that the nearest elementary school is a reasonable proxy for the school to which a household is assigned. However, in the longer term, HUD should consider linking assisted households to zoned schools, using the School Attendance Boundary Information System (SABINS), which presently covers three complete states and numerous smaller areas.

The researchers surveyed the existing literature and conducted a case study of New York City in order to assess how to evaluate the quality of public schools and identify the school or schools that assisted households can attend. The research used data on assisted household locations from HUD administrative systems and data supplied by the U.S. Department of Education that include the school-level proficiency rate in math and English for 2009.

The researchers suggest that as additional data become available HUD should incorporate the following improvements to its school quality metric: shift to mean test scores instead of proficiency rates; focus only on fourth grade students; create multiple-year averages of test scores; incorporate measures of the performance of economically disadvantaged students; and rely on "value-added" measures of school quality. Additionally, a measure of school choice could be incorporated by using the number of charter/magnet schools within one to two miles of an assisted household as well as the average fourth grade test scores of the three closest charter/magnet schools.

TABLE OF CONTENTS

LIST OF TABLES

Executive Summary

In its recent strategic plan, the U.S. Department of Housing and Urban Development (HUD) listed "using housing as a platform to deliver a wide variety of services and improve the quality of life of its residents" as one of its five key goals. One subgoal focused on increasing access to high-quality public schools for children in HUD-assisted housing. This report aims to help HUD reach its strategic goals by determining how best to evaluate the quality of public schools and how best to identify the school or schools that assisted households can attend. Based on our survey of the existing literature and a case study of New York City, we have created a series of recommendations for HUD.

How Best To Evaluate Schools

Based on our analysis, we believe that HUD can create a useful measure of school quality using the existing data provided by the U.S. Department of Education (DoED). These data include the school-level proficiency rate for 2009 in math and English. We recommend that HUD focus on proficiency rates for elementary schools, as the location of one's home typically determines access to an elementary school but does not so clearly restrict the choices of middle and high schools.

As additional data become available, we suggest that HUD make the following improvements to its measure of school quality:

- Shift to reliance on mean test scores instead of proficiency rates.
- Focus only on fourth grade students rather than all students in a school.
- Create multiple-year averages of test scores.
- Incorporate measures of the performance of economically disadvantaged students.
- Rely on "value-added" measures of school quality.

In recent years there has been a movement toward increasing levels of school choice, largely in urban school districts. We therefore suggest that HUD also incorporate a measure of school choice in its metric of school quality. In particular, HUD should use the number of charter/magnet schools within one to two miles of an assisted household as well as the average fourth grade test scores of the three closest charter/magnet schools.

How Best To Identify Schools Assisted Households Can Attend

Our case study suggests that the nearest elementary school is a reasonable proxy for the school to which a household is assigned. In the longer term, however, we propose that HUD link assisted households to zoned schools, using the School Attendance Boundary Information System (SABINS). The SABINS data set already includes elementary school attendance zone boundaries for three complete states, nine regionally diverse metropolitan areas, and over 200 cities.

Creating a Metric

HUD requires a metric that is consistent across states and metropolitan areas. We recommend a ratio that compares the test scores of the schools available to subsidized households to the test scores of the schools available to other households (or perhaps other renter households) in that same county or metropolitan area. This ratio should be calculated using five-year American Community Survey (ACS) estimates of population—and population characteristics—at the block-group level. This metric overcomes both the differences between states' methods of evaluating schools and if limited to renters, it would eliminate differences in homeownership rates across metropolitan areas. HUD could also choose to limit the comparison group further to low-income households.

Using this ratio, HUD can create a simple score of relative school quality for each metropolitan area and each program type. A score above 1 means that assisted households are able to access higher quality schools than the average household in the Metropolitan Statistical Area (MSA) or county. A score below 1 indicates that assisted households are accessing lower quality schools than the average household in that geographic area. This score will allow HUD to identify metropolitan areas to target for mobility efforts and to track progress over time.

Introduction

In its recent strategic plan, the U.S. Department of Housing and Urban Development (HUD) listed "using housing as a platform to deliver a wide variety of services and improve the quality of life of its residents," as one of its five key goals. One subgoal focused on increasing access to high-quality public schools for children in HUD-assisted housing. This report aims to help HUD reach its strategic goals by determining how best to evaluate the quality of public schools and how best to identify the school or schools that assisted households can attend. Based on our survey of the existing literature and a case study of New York City, we have created a series of recommendations for HUD.

I. How To Evaluate Schools

Experts in the field of education continue to debate the best way to evaluate the performance of public schools. Ideally, we would like to evaluate schools on their ability to improve students' future employment outcomes, their earnings potential, or maybe even their future happiness or life satisfaction. We would like to identify which schools, if any, play a role in decreasing the likelihood that a child will turn to a life of crime or rely on public assistance. It is extremely rare, however, to have access to such long-term measures; moreover, it is not practical to wait ten years to learn how a school is performing. For the most part, researchers and policymakers have instead evaluated schools based on test scores, as these scores are easy to measure and give real-time feedback. Researchers justify this choice by arguing that test scores provide meaningful information about how well the school is educating its students, as well as how children attending these schools are likely to fare in the future (Currie and Thomas 2001).

Using Test Scores To Measure School Quality

Even among the experts who rely on test scores to assess school quality, there is still considerable debate about how to develop meaningful measures. School systems and researchers employ a wide range of approaches that use test scores to measure the quality of local schools. We begin this section by highlighting the differences in the choices made by states about the content or knowledge to be tested, which students to test, and how to test them. We next discuss the strengths and weaknesses of using levels or changes in test scores to evaluate school quality. We then discuss a number of additional measurement issues.

Content and Testing: Differences Across States

Although the implementation of No Child Left Behind (NCLB) in 2001 pushed the country toward national standards, states still have considerable autonomy in our current system of accountability.[i]

Subjects Tested

Perhaps the most fundamental choice a state must make is to define a set of curricular standards (Ladd 2001). The state must then decide which subjects will become the focus of its

accountability system. Although most states focus on math, English and reading, there is still variation in the subjects tested at the state level. Hanushek and Raymond (2002) report that in 2001, 45 states used subject based tests in English, 43 in math, 23 in history/social studies (mostly in middle and high schools) and 29 in science. NCLB requires all states to test students in English, math, and in some grades also science.[ii]

Grades Tested

NCLB requires states to test all students in grades three through eight in English and math each year, and in science at least twice in elementary school. Students must also be tested in English, math and science at least once in grades 10 through 12. Some states just meet these requirements, while others begin testing even younger students and continue to test students in every year of high school. According to Hanushek and Raymond (2002), eight states tested students in nine grades or more as of 2001.

Students Tested

States differ in their requirements of which students must be tested. For example, students with learning disabilities, limited English proficiency, or who are absent on the day of the test, are often exempted from taking the test. NCLB imposes limits on the proportion of students that can be exempted from testing (Kane and Staiger 2002). Rather than specifying which students must be included, currently, the rule is that 95 percent of the student body must be tested. A state can average its participation rate over two or three years to meet this requirement.[iii]

Type of Test

Once the subjects and grades are chosen, states must develop reliable and valid tools to measure how well students have mastered the curriculum. States must decide if they will rely on criterion-referenced tests, norm-referenced tests, or both. *Criterion-referenced assessment* refers to mastery of specific learning domains often assigned a cut point, above which is passing or proficient. Some states use "off-the-shelf" tests, such as the Stanford Achievement Test or the Iowa Test of Basic Skills, whereas many other states have developed their own tests. Alternatively, some tests are *norm-referenced* (such as the SAT and GRE), which compare students to a given population, usually scored with a percentile rank. Again using the state survey reported by Hanushek and Raymond (2002), we see that 34 states rely on norm-referenced tests, 35 rely on criterion-referenced tests, with 19 states relying on both.

There is also debate on the best type of testing technique, including multiple choice exams, short answer tests, essays or other types of written work, observational studies, or assessments of portfolios of work. Focusing on elementary school level testing, Hanushek and Raymond (2002) report that in 2001, 49 states used standardized tests with multiple choice formats, 36 included short-answer questions, and 44 used essay answers to evaluate English compositional skills. Only two states, Vermont and Kentucky, employ intensive methods of assessing portfolios of student work. Although this study is dated, it offers some sense of the variation across states.

Levels vs. Changes

Levels of Achievement

The primary method through which test scores are used to evaluate school quality is the average performance or proficiency rate of students in each school. The benefit of using this approach is that these data are readily available for all schools across the country. Additionally, this measure is easy to understand, as it represents the achievement level for all students in a school.

That said, children come to school with different degrees of readiness to learn, and therefore achievement status measures may capture more about the past achievement of students in the school and their family backgrounds, and less about the actual ability of the school to educate its students (Hanushek and Raymond 2002). In other words, high average test scores indicate that students at the school are high performers, but they do not demonstrate that the school is effectively teaching those students and leading them to be high performers. It may simply be that high performing students select into particular schools.

There is a large body of literature criticizing the use of achievement levels as a measure of school quality as these test scores are thought to reflect student socioeconomic status, as much as, or even more than, they reflect school quality. This literature builds on the 1966 *Equality of Education Opportunity Study* (EEOS), which collected data on test scores as well as "socioeconomic background" variables (including parental education, father's type of occupation, and other things that families may have such as cars or televisions) for over 600,000 students. In an analysis of this survey, Coleman (1966) reports that when controlling for socioeconomic background, differences between schools account for only a small fraction of the differences in student achievement. Rothstein (2004) confirms these findings, reporting that four decades of research consistently comes to the same conclusion. At least two-thirds of the variation in achievement among schools can be attributed to the family backgrounds of their students.

Regardless of whether levels of achievement reflect the underlying quality of the school, these levels are correlated with long-run outcomes and are valued by parents. Recent research by Chetty et al. (2010) finds that higher kindergarten class quality, as measured by classmates' end-of-class test scores (which capture both *ex-ante* variation in peer abilities and the effects of the teacher on students) increases earnings, college attendance rates, and other outcomes. Students randomly assigned to a class that is one standard deviation higher in quality score 6.27 percentile points higher on end-of-year tests and earn $483 (3%) more at age 27. Additionally, research shows that parents are willing to pay more for housing in school districts and attendance zones where school test scores are higher.[iv] In reviewing the existing literature, Machin and Black (2010) report that one standard deviation increase in school quality generates a 3-percent increase in house values on average.

Changes in Test Scores

To address some of the concerns outlined above, many researchers advocate using growth measures to assess school quality, rather than levels of achievement, as these indicators identify the school's role in advancing student learning.

Time Series Approach

The simplest model for assessing changes in performance is to track a series of status indicators (such as average test scores or proficiency rates) over time. This measure only requires data on the performance of a school or grade at two points in time, and these data are increasingly available.

But while this simple, *time series* approach helps to weed out underlying differences in students across schools, it too is confounded by factors other than school performance. As the model compares test scores for two different groups of students, it confounds the differences in school quality across two years with the differences in family background and preparation between these groups of students. If student populations are constant from year to year, then differences in school performance would likely dominate these measures rather than differences in student background. However, on average only 55 percent of students live in the same house over a three-year period, and the percentage is even smaller for low-income students (Hanushek, Kain, and Rivkin 2001). Looking at movement across schools, the average annual student mobility across schools in Texas is greater than 20 percent; in California the annual mobility is 15 percent (Fletcher and Raymond 2002). In New York City (NYC), 17 percent of public school children in grades one through four switched schools between 2007 and 2008 (Been et al. 2010). Thus it is unlikely that student populations are remain the same from year to year.

Additionally, one-year changes in test scores are not stable measures of school improvement. Kane and Staiger (2002) find that even among the 20 percent of schools with the largest numbers of students, over 60 percent of year-to-year changes in test scores were due to fairly random fluctuations (such as variation in the sample and individual student performance) and did not reflect meaningful improvement.

Furthermore, even if the student body in a school was constant over time and changes were stable, time series measures capture improvements in the school, not levels of school quality. If a school provides a large value-added to students each year this measure will be zero, as there will be no difference in student achievement from year to year. At the other extreme, if a school begins at a low quality level but improves each year, though still providing a lower value-added than the first school, it will have a positive change measure (Choi, Goldschmidt, and Yamashiro 2006).

Value-Added Approach

To address the problem of changing students, a number of states and school districts have begun to rely on panel data approaches that follow either individual students or a group of students over time, calculating the improvement or decline in performance for that individual or cohort. These

are called *value-added* measures. Value-added measures are preferable to measures that describe only the levels of student performance as they measure the contributions of the school to student learning from one year to the next (Ladd 2001).

Increasingly, states are beginning to rely on these growth measures to assess each school's contribution to a child's education. According to the Center for Greater Philadelphia, three states (Tennessee, Pennsylvania, and Ohio) used value-added measures in 2004, as did several hundred school districts in 21 other states.[v] Since 2004, more localities have begun creating measures of value-added, increasing the probability that these measures will be available at the national level in the next couple of years. Tennessee probably uses the most sophisticated model of value-added—the Tennessee Value Added Assessment System (TVAAS). TVAAS is a longitudinal database that tracks individual student achievement year by year, subject by subject, and teacher by teacher, based on the year-end Tennessee Comprehensive Assessment Program (TCAP). This model is now available for other states through the Education Practice at SAS. They offer a generic version, called Education Value Added Assessment System (EVAAS), which can be modified to meet each state's needs (Derringer 2010). This is currently the most popular and widely accepted value-added model (Amrein-Beardsley 2009).

Although the value-added model is widely viewed as the best way to measure school performance, experts still debate how to implement it. One issue is which students to include. As some students move in the middle of the year they are not a reliable source of school level value-added, therefore most measures focus on students that attend the school for the entire year. Additionally, experts debate whether special education students and those with limited English proficiency (LEP) should be included in measures of school level value-added. Hanushek et al. (2004) compare results using these different samples, and find very little difference in the measures of school level value-added, finding a correlation always above 0.98.

A second issue is how gains should be measured—as the absolute gain in test scores or a student's achievement gain relative to average gain in the state or city for students scoring in the same initial level. Hanushek et al. (2004) test the correlation of these different measures and again find a relatively high correlation between the absolute value of student gain and the relative measure of student gain (0.86), suggesting that both methods will lead to similar conclusions.

A third issue concerns how student-level performance should be aggregated to the level of the school. Hanushek et al. (2004) suggest that rather than creating a simple average of the improvement of all students, researchers might create a weighted gain measure in which different weights are placed on each subgroup according to their distribution throughout the state.[vi] This method allows researchers to summarize performance after controlling for the distribution of students.

Rather than weighting a school's performance by the distribution of students, some researchers use a regression-based approach, which includes controls for differences in family background and past performance of students. Choi, Goldschmidt, and Yamashiro (2006) create value-added measures controlling for a number of different background characteristics. They estimate uncontrolled models first, and then control for student initial score and socioeconomic status

(SES) as well as school level initial scores and SES. They find that the results from these different models are very highly correlated. After controlling for student initial performance, additional measures of student SES contribute little to this metric. This result indicates that their value added measure is not strongly correlated with the socioeconomic status of students at a school. Similar results hold for school level controls. Once initial student performance is included in the model, school level performance measures do not change the estimates.

Creating and implementing a clear and understandable value-added measure of school quality is challenging. For one thing, states must have high quality longitudinal test data (following students or cohorts over time), which not all states currently collect or maintain, although the number is growing (Amrein-Beardsley, 2009). For another, the construction of value-added measures requires so many choices that it is often seen as more of an art than a science, calling into question the validity and interpretation of these measures (Hanushek, Raymond, and Rivkin 2004). Finally, value-added measures are also criticized for their lack of transparency. As these models require sophisticated data sets and computations, they are often not so easy to interpret (Choi, Goldschmidt, and Yamashiro 2006).

Do These Measures Provide Different Information?

A few studies have explored whether these different measures of school performance provide similar information about the quality of a school. We summarize the evidence below.

Measuring Levels vs. Gains

Raudenbush (2004) calculates the correlation between math and reading performance measured through a value-added versus a levels approach, to determine whether school rankings vary when these two measures of school performance are used. He uses data from the Early Childhood Longitudinal Study (ECLS) to compare measures that rely on levels of student achievement with those that assess gains using a value-added approach. The ECLS is a nationally representative study of just under 4,000 children. Raudenbush uses end-of-year test scores as measures of student level achievement. To measure school level value-added, he uses a three-level hierarchical linear model to calculate each student's growth rate and school level averages (using the start-of-year test score as a baseline and the end-of-year test score as the final achievement level).[vii] He finds a correlation of 0.77 between average kindergarten test scores and kindergarten value-added measures for math, and 0.71 for English. Schools revealed as effective using mean achievement in kindergarten, in other words, also have a high probability of being identified as effective using value-added measures for kindergarten students. For first grade students, however, his results are quite different. He finds a correlation of only 0.06 between average performance levels and value-added measures in math, and a correlation of 0.55 in English. These low correlations show that first grade test scores provide little information about the average rate of learning or improvement among children in these schools.

Raudenbush (2004) also explores results for later grades, relying on two additional data sets. The first, the Sustaining Effects Study, is similar to the ECLS in the frequency with which students are tested, but it dates back to the early 1980s. Using the Sustaining Effects Study data, Raudenbush finds high levels of correlation between mean proficiency and value-added measures in the third grade: 0.78 in reading and 0.91 in math. These findings offer some

evidence that these two measures contain similar information about school quality. When using a second, and more recent data set from Washington DC (from 1998 to 2002), he finds correlations between 0.34 and 0.62, suggesting a far weaker relationship between levels of achievement and value-added measures.

Finally, Raudenbush (2004) also undertakes some analyses that focus on high-poverty schools (those in which more than 50 percent of students are eligible for free or reduced price lunch). He finds that though levels of achievement are much lower in these sets of schools, value-added measures are approximately the same in these high-poverty schools as compared to lower-poverty schools.

Comparing Measures of Gain

Data sets tracking individual students or cohorts of students are rare, and thus many researchers rely on changes in annual mean test scores in a school as a proxy measure for value-added. Hanushek, Raymond, and Rivkin (2004) compare these two approaches. They find low correlations between value-added measures[viii] and change in annual mean test score measures (0.64 and 0.49 respectively).[ix]

Choi, Goldschmidt, and Yamashiro (2006) explore whether value-added approaches provide different information about school quality than annual changes in mean school test scores, by comparing student level gains in schools that met Adequate Yearly Progress (AYP) to the gains of students in the schools that did not meet AYP. AYP is achieved when a school's proficiency rate rises by a pre-determined amount, thus AYP is a measure of the change in annual performance and not a value-added measure. Focusing on an urban school district in the Pacific Northwest, they find that 15 schools among 51 schools meeting AYP offer an estimated value-added gain that is smaller than the district average. They also find that among the remaining 36 schools that meet AYP, only 12 have statistically greater value-added gains than the average school in the district. In contrast, almost half of the schools that did not meet AYP offer an estimated value-added gain that is higher than the district average. These results highlight the fact that assessments based on changes in school level proficiency from year to year provide very different information about schools than do assessments based on student-level value-added measures. If the primary objective is to measure a school's contribution to student learning, then the changes in annual performance at the school level are a relatively weak proxy for a value-added approach.

Other Measurement Issues

Different Grade Configurations

The different grade configurations of elementary schools can make comparisons across schools challenging. As a solution to this problem, a number of researchers have focused on fourth grade test scores, as the fourth grade is represented in almost every elementary school (Black 1999; Bayer, Ferreira, and McMillan 2007). This solution, however, is also problematic as it does not describe the performance of all students in a school.

Mean vs. Proficiency

As NCLB requires that schools make adequate yearly progress each year, with an end-goal of 100 percent proficiency by 2014, proficiency rates have become the de facto indicator of school quality. The proficiency rate reports the percentage of students who have tested at or above a particular score that each state has deemed as the cutoff for "proficiency." Although this measure is simple and intuitive, it also has disadvantages. Most obviously, it is only sensitive to movements across the proficiency threshold, and may miss a great deal of change at the ends of the distribution (Choi, Goldschmidt, and Yamashiro 2006). (For example, if a student previously scoring 400 improved their score to 600, when the proficiency rate is 680, this student's gain would not count in the measure.) Reliance on a proficiency rate can create incentives for schools to teach toward the middle of the distribution and ignore people at the ends of the distribution. In a recent study, Neal and Schanzenbach (2010) show that in the Chicago Public Schools the introduction of NCLB in 2002 and the introduction of similar district-level reforms in 1996 generated a significant increase in reading and math scores for students in the middle of the achievement distribution, but not among the least academically advantaged students. Additionally, as each state determines their own proficiency threshold, these measures cannot be compared across states.

The average test score in a school is another commonly used measure of school performance. This measure is also readily available and simple, but it too is difficult to compare across jurisdictions as the scale and ranges of test scores vary widely across states. The mean does have an advantage over proficiency as it incorporates information about the entire distribution, rather than just the number of students above an arbitrary cutoff.

Volatility

Kane and Staiger (2002) highlight the tremendous amount of year-to-year volatility in test scores. As the average elementary school contains only 68 students per grade, the amount of variation stemming from differences in the student body across years is often large relative to the total amount of variation observed between schools. There is also a great deal of instability arising from one-time factors such as a dog barking during the test, a disruptive student, or a particularly successful rapport between a group of students and their teacher. Even small fluctuations in students' scores can have dramatic effects on school ranking, as between school test scores do not generally differ dramatically. Thus, the authors recommend that school systems report average test scores over a few years.[x]

Drawbacks of Relying on Test Scores and Potential Alternatives

As noted earlier, some researchers and policymakers raise more fundamental questions about the merits of using test scores at all to assess school quality. We elaborate on these concerns below.

Strategic Behavior

One of the most significant drawbacks with relying on test scores to evaluate school performance is that such a reliance can encourage "teaching to the test," or the shifting of resources in ways that improve performance on a test but do not actually increase mastery in a subject. For example, schools may coach students in the specific questions or formats that will be asked on the exam. In one striking example from New Jersey, when students were asked to add decimals in the familiar vertical format the passing rate was 86 percent; when the decimals were provided in horizontal format the passing rate was 46 percent; when students were asked to subtract rather than add, the passing rates in the vertical and horizontal formats were 78 and 30 percent respectively (Shepard 1988). When Koretz et al. (1996) asked teachers in Kentucky to report the importance of several different factors to account for the improvements in student test scores, more than half of the teachers said "increased familiarity" with Kentucky's accountability test and "work with practice tests and preparation materials" had been important, while only 16 percent reported that "broad improvements in knowledge and skills" accounted for the improvement.

Schools may also manipulate the students who are actually tested. Studying the introduction of test-based accountability for Chicago public schools, Jacob (2005) finds that the large increases in test scores after accountability went into effect were also accompanied by increases in special education placement and grade retentions. Deere and Strayer (2001a, 2001b) and Cullen and Reback (2006) also find increases in special education placement when accountability measures were introduced in Texas. Figlio and Getzler (2006) concentrate on special education placement after the introduction of a state accountability system in Florida. They find that placement rates increase relatively over time in grades that enter into the accountability system, whereas there is no similar increase in those grades that do not enter into the accountability system.

Alternatively, schools may place more resources in grades that "count." Boyd et al. (2002) consider whether teacher placement responds to the specific grades that are included in high stakes tests. They do find some attempt in urban schools to place the more experienced teachers in the grades tested when new teachers entered a school. Finally, there is concern that teachers and/or administrators may cheat to help improve test scores. Jacob and Levitt (2003) provide evidence that some teachers or principals in Chicago changed student answers after high stakes assessments in the 1990s. Their estimates suggest that 4 to 5 percent of classroom test scores in Chicago elementary schools were affected by cheating, as measured by common student responses combined with unexpected rises and subsequent declines in student performance. If such gaming is widespread, then test scores may not capture actual learning. However, as all schools face similar incentives to game the system, the threat of strategic behavior may not produce much bias in *relative* measures of school performance. Rather, gaming will simply inflate overall scores for a system as a whole.

Alternatives to Test Scores?

Even absent concerns about strategic behavior, standardized tests may not capture the learning that ultimately shapes the trajectory of a student's life. There is some evidence from school choice lotteries, for example, that lottery winners (students that are able to exercise choice and

move to better schools) do not experience improvements in individual scores, but they do experience improvements in non-traditional outcomes.[xi]

Some alternatives to using test scores to evaluate school quality include measures of school resources (such as teacher/pupil ratios or general funding levels per student) or the quality of teachers in a school. A number of states and cities use composite measures that include various school indicators to assess school quality.[xii] These measures provide information on different dimensions of the school that may contribute to student learning and offer a broader portrait of overall quality.

Project STAR (Student/Teacher Achievement Ratio), in which over 11,000 students in Tennessee and their teachers were randomly assigned to different classrooms within their schools from kindergarten to third grade, has allowed researchers to experimentally identify the relationship between school level inputs and student outcomes. Overall, the body of research on Project STAR has found that class size (Finn and Achilles 1990; Word et al. 1990; Krueger 1999), teacher quality (Nye, Konstantopoulos, and Hedges 2004; Dee 2004), and peer quality (Cascio and Schanzenbach 2007; Graham 2008; Sojourner 2009) are all predictors of increased test scores in the short run. In a recent study on the long term effects of Project STAR, Chetty et al. (2010) find that the effects of class size, teacher quality, and peer quality on test scores fade out by eighth grade. However, though test score effects are not permanent, they are able to link these school-level inputs to long term outcomes such as increased wages, college attendance, home ownership, neighborhood quality, and marital status.

The drawbacks of relying on a set of measures to assess school quality include the complexity of creating a measure based on multiple criteria, as well as the difficulty of assembling a large collection of school level indicators. There is also no evidence that these alternative measures provide a more complete picture of the student learning that occurs in a school than do test scores (Hanushek and Raymond 2002). Moreover, measuring teacher quality is challenging, with most of the current discussion focusing on measures of teacher level value-added (Goldhaber and Anthony 2007). As school level value-added measures are composite indicators of all teacher value-added within the school, inclusion of the school level value-added in a metric will by construction include a measure of overall teacher quality. Therefore inclusion of a value-added measure of school quality makes the creation of a separate measure of teacher quality obsolete.

Rather than create composite measures of school quality, researchers instead have relied primarily on test scores as these outcome measures provide information on the ability of a school to use their given inputs to produce the required outputs. In addition, researchers have found a connection between test scores and longer term outcomes, suggesting that test scores are appropriate indicators of a student's future trajectory. There is some research that has connected student test scores to longer term outcomes. For example, Neal and Johnson (1996) rely on the National Longitudinal Survey of Youth (NLSY) to examine the relationship between scores on a test administered when youth are between 14 and 21 and future wages. They find that when controlling for age, race, and ethnicity, test scores are highly significant predictors of wages at ages 26 to 29. Currie and Thomas (2001) use data from the British National Child Development Survey (NCDS) and find that test scores at age 16 are important determinants of wages and employment at age 33 for all individuals, including individuals of lower socioeconomic status.

A series of papers have tried to pinpoint the magnitude of the relationship between test scores and future earnings. Murnane et al. (2000) provide evidence on this relationship drawn from the "High School and Beyond" survey and the national longitudinal survey of the high school class of 1972. They find a one-standard deviation increase in test score performance predicts a 15-percent increase in earnings for male students, and a 10-percent increase in earnings for female students. Using a different data set and focusing on younger children, Lazear (2006) finds that a one-standard deviation gain in test scores correlates with a 12-percent boost to future earnings.[xiii]

Together, evidence that test scores are indicators of longer term outcomes but that they do not capture all dimensions in which schools can improve longer term outcomes suggests that there is value in relying on test scores as measures of school quality, but that additional indicators could be useful. Rather than focus on school-level inputs (for which there also is currently no consensus on which characteristics are most valuable in determining school quality), researchers have suggested that additional intermediate measures should be constructed that can measure a school's influence on a child's development of non-cognitive skills, such as motivation to learn, intellectual curiosity, or other emotional competencies (Cunha et al. 2006). If reliable measures of non-cognitive assessments were developed, they would be a valuable additional measure of school quality

.

II. Linking Assisted Households to Schools

The only geographic boundaries available for schools nationally are school district boundaries, so researchers have often focused on the quality of schools within a household's school district.[xiv] However, school districts are composed of heterogeneous schools, and therefore the average quality of schools within a district is not likely to capture the quality of the school a given student attends. Below, we describe five alternative methods for linking households to schools. We also describe a method for measuring the local level of school choice.

Identifying the Schools Students Attend

Ideally, one would want to know exactly which school a student in assisted housing attends, in order to evaluate the quality of their school. This could be accomplished by creating partnerships with state or local Departments of Education that track students over time. Jacob (2004), for example, matches children living in public housing and children in households with housing vouchers who were recently forced out of public housing due to demolitions to school records by combining data from the Chicago Housing Authority (CHA) to data from the Chicago Public Schools (CHiPS). He finds that students who remained in public housing, in comparison to those who were forced to exit, attend similar quality schools (as measured by the share of students proficient in math and reading as well as racial composition and size), even among voucher holders that move to low poverty neighborhoods.

Schwartz et al. (2010) use student-level data from the New York City Department of Education (DoED) to link students living in public housing to their schools. Their report finds that students living in public housing in New York City attend schools with similar resource levels to schools

attended by other comparable students, but they attend schools with peers who are poorer and perform substantially worse on standardized exams than the peer groups at other schools throughout the city.

Relying on administrative data from public schools, this approach omits private schools. Although most assisted households likely send their children to public school, this omission could be an issue in some areas.

Another way to link students to their actual school is to collect data directly from the assisted households during re-certification. Just as the U.S. Department of Housing and Urban Development (HUD) collects data on the income, household composition, etc. of assisted households, HUD could also collect data on the schools that children in assisted households attend. It would also be useful for HUD to learn more about why assisted households chose these schools. DeLuca and Rosenblatt (2010) interviewed Moving to Opportunity (MTO) participants in Baltimore to learn more about why these households were not sending their children to schools of higher "quality" even though they were living in neighborhoods with dramatically lower poverty rates. Some of the primary reasons were that they did not want to move their child from their current school, that they lacked information about local schools, or had low expectations of school in general.

Though both of these methods require a large amount of resources, they are worthwhile long-term plans. Still, as HUD's current objective is to determine whether assisted housing *can* provide access to high performing schools, it is valuable for HUD to get a sense of the quality of the local public schools in the neighborhoods where assisted households live, even if households do not send their children to these exact schools.

Attendance Zone Boundaries

In most of the country, the public elementary school a child can attend is determined by attendance zone boundaries. The zoned elementary school for each assisted household provides a baseline measure of the educational opportunity available to assisted households. As there is no national repository of elementary school attendance zones, we rely on a data set created by Sal Saporito at the College of William and Mary called the School Attendance Boundary Information System (SABINS), which is collecting elementary school attendance zone boundary files throughout the country.[xv] To date, the system has collected elementary school boundary files for three entire states (Minnesota, Delaware, and Oregon), seven regionally diverse metropolitan areas (Atlanta, Georgia; Hartford, Connecticut; Houston, Texas; Milwaukee, Wisconsn; Philadelphia, Pennsylvania; Tucson, Arizona; Washington, DC), and over 200 cities. The goal is to collect data from the 600 largest cities in the country.

Nearest School Match

Attendance zone boundaries are not available for the entire country, however, and some areas are not zoned. Thus relying on the nearest school can be used as an alternative. As elementary school students are less likely to travel far to attend an elementary school, and there is evidence that lower income families are particularly likely to base their school choice on distance, (Teske

Fitzpatrick, and Kaplan 2007), researchers can link households to the closest elementary school, within their school district, to create a proxy measure for educational opportunity.

This method is identical to that employed by Deng (2007) who uses "Thiessen" polygons[xvi] to create proximate zones around each elementary school. Deng compares the local public schools available to voucher holders to those available to Low-Income Housing Tax Credit (LIHTC) tenants, finding that both LIHTC tenants and voucher holders live in neighborhoods that contain public schools with lower levels of performance than the public schools attended by the typical renter household in the same metropolitan statistical area (MSA). When comparing the two programs, she finds that in Boston and New York, voucher holders have a clear advantage over LIHTC tenants and both programs in Boston and New York provide better opportunities than do these programs in the other four MSAs (Atlanta, Miami, Cleveland and San Jose).

A variant on this approach would be to create an average measure of school quality for the three schools closest to the assisted household, as the zoned school may in some cases be the second or third closest school, given how boundaries are drawn. Moreover, to the extent that students go to non-zoned schools, they are likely to attend those nearby.

Charter/Magnet Schools[xvii]

In recent years, there has been a movement toward increasing levels of school choice, particularly in urban school districts. As of 2008, 41 states had authorized charter schools, and 5,000 charter schools nationwide served over 1 million students—approximately 2.6 percent of the public school population (Hoffman 2009). The degree of choice available in the school district is also an important indicator of educational opportunity. One might create a measure of such choice by counting the number of charter/magnet elementary schools available within one or two miles of each assisted household (perhaps adjusting radius depending on population density). To measure the quality of the alternative schools available, one might track the average quality of the three closest charter/magnet schools.

III. Proposed Metric

We conclude from our review of the literature that if looking for a single, simple measure of school quality, some metric capturing test scores is the best choice. However, care must be taken when constructing and interpreting measures using test scores. We offer a number of recommendations for the U.S. Department of Housing and Urban Development (HUD).

Measuring Test Scores

First, we recommend focusing on elementary schools, as the location of one's home typically determines access to an elementary school but does not so clearly restrict the choices of middle and high schools.[xviii] Additionally, there is a growing body of empirical research highlighting the importance of early childhood education in predicting adult outcomes, suggesting that it is most useful for HUD to focus on the quality of elementary schools that children in assisted housing

may attend.[xix] Second, as test results for math and English are available for all states, we suggest that HUD limit its analysis to results for these subjects.

Based on our reading of the literature, we have several additional recommendations that HUD could adopt in the longer term as more data become available. First, we recommend focusing on mean test scores rather than proficiency rates, as the mean test score provides information about the performance of students for the entire distribution rather than just the number of students above a cutoff. Although this measure is not included on the data set DoED provided to HUD, it is collected by all state departments of education and therefore should be available nationally in the short run.

Second, we suggest focusing on performance for just one grade of students, specifically fourth grade, so that results for schools with different grade compositions can be more easily compared. According to No Child Left Behind (NCLB), states are required to report performance of each grade at the school level, so this data should soon be available nationally.

Third, we suggest averaging scores across two or three years, as annual school-level test scores exhibit a large amount of volatility. Again, as states have been required to collect test scores for a number of years, this data should also be available nationally in the near future. In this way HUD could then rely on two or three year moving averages to describe levels of school performance.

Fourth, since HUD is particularly interested in the lowest income populations, we also suggest focusing on achievement levels for these groups in a school. NCLB requires states to report on the performance of economically disadvantaged students, so these data too are currently collected in all states and should soon be accessible nationally.

Finally, in the longer run, we suggest including "value-added" measures at the school level, as these indicators quantify the contribution of the school to student learning more accurately than other methods such as time series approaches. Research suggests that value-added measures are superior to simple changes in mean school test scores; therefore, we do not recommend the inclusion of changes in average scores in the metric of school quality.

In our case study of New York City below, we assess the extent to which using these different metrics leads to different school rankings.

Comparison Group

As every state has a different set of evaluation measures, and every county or metropolitan area has a different population of students, we recommend that HUD evaluate its success in providing assisted households access to high performing schools relative to other households in the same county or metropolitan area. We suggest focusing on the metropolitan area for counties that are within metropolitan statistical areas (MSAs), and focusing on counties for areas outside of MSAs.

HUD could begin by comparing school quality available to assisted households to school quality available to all households in the same MSA or county. Given that HUD-assisted households are

all low income renters, if they had not participated in the program they would not have realistically had access to all housing units in the geographic area. HUD may want to give some serious thought as to whether it wants to narrow the comparison down further to renter households or low-income households. So, for instance, HUD might want to compare the quality of schools available to subsidized tenants to the quality of schools available to children living in households of similar income levels. Alternatively, it might want to keep the comparison broader, so as not to suggest that lower income households should expect to attend lower-quality schools.

Such comparison groups can be created by using American Community Survey (ACS) block group level data. The centroid of each block group can be joined to the elementary school zone in which it lies. Although this method does not provide a perfect match, it offers a reasonable approximation of the population living in each elementary school attendance zone.

Data are available at the block group level on the total number of households, on the number of households by tenure status, and on the number of households in a series of income buckets.[xx] As each of HUD's housing programs has a different eligibility cutoff, HUD might tailor the comparison group to most accurately reflect the income cutoff of the population served by the housing program. For example, to be eligible for a housing voucher, households must have incomes below 50 percent of Area Median Income (AMI). To be eligible to live in a tax credit development, households must have incomes below 60 percent of AMI. Also, as each metropolitan area has a different AMI, HUD should adjust the comparison group within each metropolitan area/county. Block group data also include tenure status, so researchers could calculate the average quality of the schools accessible to renters.

Once the appropriate comparison group is chosen, HUD then can compare the school quality available to subsidized households to the school quality available to all other households, other renter households, or households with similar incomes, who live in the same geographic area.[xxi] For example, if all subsidized households in a metropolitan area have access to schools with an average test score of 450, we can then compare this score to the average test score of schools that other renters can access, say 425. Then, even if the mean test score in the metropolitan area is 500, the ratio of these means shows us that subsidized households are accessing higher quality schools than other similar households. This ratio, the mean score of subsidized households divided by the mean score of all renters (within each metropolitan area or county), can be compared across metropolitan areas.

Interpretation

Using this ratio, HUD then can create a score for each metropolitan area and each program type. A score above 1 means that assisted households are able to access higher quality schools than the average low-income household in the MSA or county. A score below 1 means that assisted households are accessing lower quality schools than the average low-income household in that geographic area. This ratio will allow HUD to isolate areas that are doing particularly well and those that are doing particularly poorly, as well as programs that are succeeding more than others. This will assist HUD in identifying the causes for these discrepancies and improve

HUD's ability to help subsidized households access higher quality schools. We next create such a metric for New York City, providing a case study for HUD on our suggested approach.

IV. Case Study of New York City

Although New York City may not be typical among school systems, we focus on it for a number of reasons. First, New York City has the largest public school system in the country. Second, we have maps of the elementary school attendance zone boundaries for the city, allowing us to assess how well the nearest school works as a proxy for the actual zoned school. Third, we have access to a variety of different test score measures for the city's public schools, allowing us to construct and compare several alternative metrics of quality and to determine the extent to which school rankings are sensitive to the choice of measure. Fourth, New York City offers public school students a range of charter school and magnet school options, allowing us to experiment with measuring the degree of choice available to households living in subsidized housing, relative to similar households in New York City.

We begin this section by describing the New York City Public School System, including the accountability measures it uses and the degree of choice afforded to students. In part (b) we provide some background on subsidized housing in New York City. In part (c) we evaluate the different methods of linking students to schools, and in part (d) we compare a range of alternative metrics of quality.

New York City Public Schools

New York City has the largest system of public schools in the United States, serving over 1 million students. The city has 1,700 schools in total, including 795 elementary schools, which serve just under half a million students. The vast majority of elementary schools are "zoned" schools, with 8 percent magnets or charters. The students in the city's public schools are quite diverse: 32 percent black, 40 percent Hispanic, 14 percent Asian and 14 percent non-Hispanic white. Seventy-one percent of all students are eligible for free or reduced-price lunch.

New York City Accountability

New York City uses three separate and complementary accountability systems to evaluate schools. The first is a Progress Report, constructed by the New York City Department of Education. The second is a Quality Review, also conducted by the New York City Department of Education. The third is the New York State Annual School Report Card.

The New York City Progress Report, which began in the 2006-07 school year, provides an overall assessment of three main areas of a school: (I) School Environment, (II) Student Performance, and (III) Student Progress. School Environment includes metrics such as attendance and school safety. Student Performance includes proficiency rates in math and English. Student Progress is assessed with the city's value-added measure. Data for each of these areas are collected through surveys, administrative records, and standardized exams. Schools receive letter grades (A through F) in each of these three categories, as well as a single summary letter grade. Schools receive additional recognition for exemplary student outcomes by students

most in need of attention and improvement, such as economically disadvantaged students and students with limited English proficiency. To construct these grades, schools are compared to all schools citywide and to schools with student populations most like their own.

The Quality Review Score is a separate accountability score based on an on-site Quality Review of the school by an experienced educator. The score assesses the quality of efforts at the school to (1) track the capacities and needs of each student, (2) plan and set rigorous goals for each student's learning, and (3) evaluate the effectiveness of plans and practices, and revise them as needed. The Quality Review Score is measured on a four point scale: Well Developed, Proficient, Underdeveloped with Proficient Features, and Underdeveloped. The Quality Review Score is not incorporated into the Progress Report Grade, but is treated as a different, equally important indicator.

The New York State Annual School Report Card is yet a third accountability indicator, which reports the school's performance under the accountability system New York State has adopted as part of the federal No Child Left Behind (NCLB) Act. This Report Card includes the proficiency rate for students in each grade as well as a number of specific subgroups (such as economically disadvantaged students). A school's official NCLB status assesses whether a school has made Adequate Yearly Progress (AYP) or not, or whether students in a school have attained a pre-specified goal of proficiency in literacy and mathematics.

New York State Tests

New York State relies on standards-based tests developed to measure concepts, processes, and skills taught in schools throughout the state. Students in grades three through eight are tested in English Language Arts (ELA) and math. The tests are composed of multiple-choice and constructed response items. In ELA there are primarily three learning standards: S1–Information and Understanding, S2–Literary Response and Expression, and S3–Critical Analysis and Evaluation. In math there are five content standards in grades three to seven, and four content standards in grade eight. The reliability of the test is estimated using two types of statistics, Cronbach's alpha (Cronbach 1951) and Feldt-Raju coefficient (Qualls 1995). These two measures are used to estimate how well items that represent the same construct yield similar results. Overall, New York State tests are internally consistent with ELA scores in the 0.83 to 0.89 range and math scores in the 0.88 to 0.94 range. Additional tests were conducted to determine the standard error of measurement, classification consistency, and accuracy of the testing instrument.[xxii]

New York State Scores

New York State publishes several scores from these tests. The first, the scale score, is a quantification of the ability measured by the grade three to eight ELA and math tests at each grade level. The scale score is comparable within each grade level but not across grades, as they are not on a vertical scale. The second score is a proficiency level. Students are classified as either Level I (Not Meeting Learning Standards), Level II (Partially Meeting Learning Standards), Level III (Meeting Learning standards), or Level IV (Meeting Learning Standards with Distinction). New York State used a Bookmark Standard Setting Procedure (BSSP) to set

the standards for grades three through eight in ELA and math, which involves input from a variety of experts in a multiple-day discussion, and analysis of structured questions.[xxiii] The third score is a Standard Performance Index (SPI) score, which is an indicator of student ability, knowledge, and skills, in specific learning standards; it is used primarily to help teachers evaluate academic strengths and weaknesses of their students. Additionally, school level results are reported for a series of subgroups, including economically disadvantaged students, as required by NCLB.[xxiv]

New York City Value-Added Measure

In addition to reporting levels of achievement, New York City also computes a student level value-added measure which is aggregated to the school level. To be included in this measure, a student must be on the school's October 31, 2009 audited register, must be in at least fourth grade (as two years of test data are required for each student), and must take the test in one grade higher than in the previous year. New York City calculates this measure using a student's growth percentile, which indicates the percentage of students, starting at the same test score, whom the student's growth exceeded. For example, if a student scored 3.04 on the third grade ELA in 2009 and then scored 3.21 on the fourth grade ELA in 2010, their score is then compared to other students who scored a 3.04 in 2009. Let's say 62 percent of these students scored lower than 3.21, so this student's growth percentile would be 62.0.[xxv] This measure is intended to focus on the capacities students develop as a result of attending the school. New York City also computes this metric for each school's lowest one-third, which is based on the student's scores on the relevant test in the previous year.

School Choice

New York City has a well developed system of school choice, which includes charter schools and magnet schools. New York City has 82 charter schools, which serve 22,000 students. Of these, 53 serve 14,000 elementary school students. Charter schools are public schools that operate independently according to the terms of a five-year performance contract or "charter." All charter schools are free and open to all students. For schools where demand is greater than supply, a lottery must be used to determine admission to the school. Preference is given to students with siblings already attending the school and who live in the community school district in which the charter school is located. Some charter schools also give preference to students who are struggling to succeed, as evidenced by their test scores.[xxvi] Though all charter schools are different, as they are independently run, some features that distinguish charter schools in New York City include longer school days and a longer school year (some even provide school on Saturdays). Charter schools tend to have low student-teacher ratios, which offer more personalized attention for each child.[xxvii]

According to the Common Core of Data, New York City has 141 magnet schools, which serve 140,000 students. Most of these 141 schools generally use attendance zones to determine student access, but have a small magnet program that brings in some students from outside the catchment area. We thus classify most of these schools as neighborhood schools, and have identified eight elementary schools that are not 'zoned' schools. We include these eight schools with the 53 charter schools to create measures of school choice for children living in assisted housing. How

to correctly identify magnet schools is another issue that HUD must consider when replicating this methodology at the national scale. The Common Core defines magnet schools as those that have special programs designed to attract students of different backgrounds, or to provide an academic or social focus on a particular theme (for example, science/math, arts, etc.). Thus it may be true across the country that very often these schools are primarily neighborhood schools with magnet programs.

Subsidized Housing in New York City

The data set provided by HUD offers information on approximately 340,000 HUD-assisted households in New York City.[xxviii] In our case study we focus on the three largest HUD programs. Of these households, the majority are public housing residents (50%). In addition, 34 percent are voucher holders and a smaller share live in project based Section 8 developments (16%).[xxix] In the country as a whole, again focusing on the three largest HUD programs, public housing is a much smaller share of the HUD-assisted housing portfolio (26%), voucher holders are a much larger share of HUD-assisted households (49%), as are project based Section 8 developments (25%).[xxx]

Focusing first on public housing residents, the average household has an income of $22,000 and the 60th percentile public housing resident has an income of $18,000. The average public housing household includes 2.3 residents, 38 percent of these households have kids, and 27 percent of these households have elementary school aged children. As for racial composition, the majority of household heads are black (56%), and another large portion are Hispanic (35%).[xxxi] A small portion of these heads of households are white (5.4%), and another small portion is Asian (3.3%).

Moving on to voucher holders, the average voucher household in New York City has a household income of $15,000, and the 60th percentile housing voucher holder has an income of $13,000. The average voucher household has 2.5 residents. Of these households, 47 percent have kids and 35 percent have elementary school aged children. As for racial composition, a large share of voucher households are black (45%) and a similar portion are Hispanic (35%). Notably, a much greater share of voucher holders in New York are white (19%) than public housing residents. Again, a small share of voucher holders is Asian (1.3%).

Next looking at the households living in project based Section 8 developments, we see that the average household in New York City has an income of $14,000, and the 60th percentile tenant in a project based Section 8 development has an income of $13,000. The average household has 1.9 residents. Of these households, 30 percent have children, and 16 percent have elementary school aged children. As for racial composition, a large share of these households are black (40%) and a similar portion are Hispanic (38%), a smaller share are white (14%) and a larger share are Asain (5%).

In addition there are 1,399 affordable housing projects financed through the Low Income Housing Tax Credit Program (LIHTC). Currently, HUD does not collect data on tenants of LIHTC developments, so we know a very limited amount about these assisted households. We do know that subsidized LIHTC units must rent to households earning below 60 percent of Area

Median Income (AMI), which is currently approximately $45,000. Developers report that, absent vouchers, they need to rent to households that are earning very close to this level of income. Overall there are 53,000 low-income units in these projects, of which 42 percent have two or more bedrooms, signifying the possibility that children may live in these households.

Maps 1-4 show the distribution of each of these housing programs throughout the city. We see that there are voucher holders in virtually all neighborhoods in the city, but there are concentrations of voucher holders in areas near both public housing and tax credit developments, particularly in Brooklyn and the Bronx.

Comparing Methods of Linking Households to Schools

We next test the accuracy of our various methods of linking households to schools. As the majority of elementary school students in New York City attend their zoned elementary school, we first link each assisted household to its zoned school and compare our other methods of linking households to schools to this baseline measure. If there is more than one school in an attendance zone, we link the households to both schools, and use average proficiency rates. This comparison is useful as we do not have attendance zone boundaries nationally. If alternative methods link households to the same schools, then this suggests that HUD could create a similar metric for all HUD-assisted households in the nation.

First, we link each household to its closest school within the district, which is the same as Deng's method of using "Thiessen" polygons (Deng 2007). Using this method, we find that 74 percent of households are still linked to the same school. Therefore, the nearest school within the school district appears to be a good alternative matching method when elementary school attendance zones are not available, at least in New York City.

We also test the method of linking every household to the three nearest schools. As the average distance to the nearest school for each HUD household is 0.18 miles, and the average distance of the third closest school is 0.42 miles, we only include schools that are within 0.5 miles of the student's home. Using this method, 73 percent of households are merged to three schools, 16 percent to two schools, and 10 percent to a single school. We merge the 1 percent of assisted households who are not within 0.5 miles of an elementary school to their closest school. For 95 percent of households, one of the three closest schools is their zoned school. In the longer run, we encourage HUD to obtain actual school attendance zone boundaries, but in the short run, it appears that relying on the nearest school is a reasonable (though imperfect) alternative.

As for school choice, we identify the number of charter/magnet schools available within one mile of each assisted household. We then average the quality of the three closest charter/magnet schools as a measure of the quality of the choices available to households in subsidized housing. In New York City, there are 43 charter schools that serve elementary school students and include a third or fourth grade (which are the grade in which testing begins). We have also identified eight magnet elementary schools that have no geographic requirements when determining school eligibility. Map 5 displays the location of all these magnet and charter schools in New York City, relative to the concentration of HUD-assisted households in each block group. We see that 63 percent of HUD-assisted households have at least one charter school within one mile of their

home. We also see that the average distance of the closest charter school is one mile, and the average distance of the third-closest charter school is two miles. Therefore, in this measure we only include charter schools within two miles of the assisted household's home. Using this bound, 66.6 percent of assisted households have three charter schools within two miles of their home, 8.3 percent have two, and 11.4 percent have one school within two miles of their home.

Comparing Metrics of School Quality

Selection of Data for Metric

As a test of the short-run options available to HUD, we compare the rankings of schools, generated using data currently available to HUD, to the wider range of indicators we have available in New York City.

We rank all New York City public elementary schools according to alternative metrics. Table 1 shows the correlation between the rankings, using proficiency rates, and those using average school test scores, and shows that these measures are very highly correlated (0.92) for both math and ELA. Based on these results for New York City, it appears that relying on a proficiency rate rather than an average test score may not significantly alter the school rankings.

Table 1: Using proficiency rate vs. mean test score

	Rankings by Proficiency Rate, ELA, 4th Grade	Rankings by Proficiency Rate, Math, 4th Grade
Rankings by mean, ELA, 4th Grade	0.92	--
Rankings by mean, Math, 4th Grade	--	0.92

We next compare the rankings based on proficiency rates for the entire school to those calculated just for fourth grade proficiency rates, and find very high levels of correlation (0.93) for both math and ELA. Again based on the New York City results, rankings using the entire student body are quite similar to rankings that rely only on fourth grade students. In other school systems, however, where there may be more variation in the grade composition of elementary schools, the rankings generated by the two measures may differ more substantially.

Table 2: Using all students vs. fourth grade students

	Rankings by Proficiency Rate, ELA, All students	Rankings by Proficiency Rate, Math, All students
Rankings by proficiency rate, ELA, 4th grade	0.93	--
Rankings by proficiency rate, math, 4th grade	--	0.93

We then compare the school rankings based on overall proficiency rates in math and ELA to the school rankings based on the performance of economically disadvantaged students (students eligible for free or reduced-price lunch). In math, we see the correlation is relatively high, 0.77, and in ELA it is slightly lower, at 0.63. This suggests that overall proficiency in a school and proficiency of economically disadvantaged students provide somewhat different information about a school. Comparing rankings using the mean rather than the proficiency rate, we find higher levels of correlation. Here, we find that mean scores in math and ELA is correlated to the mean scores of economically disadvantaged students at a rate of 0.80.

Table 3: All students vs. economically disadvantaged students

	Rankings by Proficiency Rate, ELA, 4th Grade	Rankings by Proficiency Rate, Math, 4th Grade
Rankings by proficiency rate, ELA, 4th Grade Econ Disadvantaged	0.63	--
Rankings by proficiency rate, Math, 4th Grade Econ Disadvantaged	--	0.77
	Rankings by Mean, ELA, 4th Grade	Rankings by Mean, Math, 4th Grade
Rankings by mean, ELA, 4th Grade Econ Disadvantaged	0.80	--
Rankings by mean, Math, 4th Grade Econ Disadvantaged	--	0.80

As for levels versus changes, Table 4 compares achievement levels and time series measures of changes in average school proficiency rates, and suggests very low levels of correlation.

Table 4: Achievement levels vs. time series changes

	Rankings by Proficiency Rate, ELA, 4th Grade	Rankings by Proficiency Rate, Math, 4th Grade
Rankings by time series changes, ELA, 4th Grade	0.15	--
Rankings by time series changes, Math, 4th Grade	--	0.46

We find higher correlations between achievement levels and value-added measures (see Table 5), but the levels of correlation are still quite low. For math, the proficiency rate and value-added measure have a correlation of 0.64, and for ELA, a correlation of 0.48. Comparing the proficiency rate to the value-added measure for the bottom third of the school population we obtain even lower levels of correlation—for math, 0.31 and ELA, 0.13.[xxxii] This shows us that proficiency rates at a school provide very different information from school-level value-added measures. This highlights the need to incorporate school-level value-added measures, particularly for the lowest achieving students, as these data provide very different information about school performance than do proficiency levels.

Table 5: Achievement levels vs. value-added

All Students	Rankings by Proficiency Rate, ELA, 4th Grade	Rankings by Proficiency Rate, Math, 4th Grade
Rankings by value-added, ELA, 4th Grade	0.48	--
Rankings by value-added, Math, 4th Grade	--	0.64
Bottom Third of Students	Rankings by Proficiency Rate, ELA, 4th Grade	Rankings by Proficiency Rate, Math, 4th Grade
Rankings by value-added, ELA, 4th Grade	0.13	--
Rankings by value-added, Math, 4th Grade	--	0.31

We continue by comparing the rankings using time series measures to those using value-added measures (Table 6), finding that these measures are correlated but clearly provide different information about school performance.

Table 6: Time series changes vs. value-added

	Rankings by Time Series Changes, ELA, 4th Grade	Rankings by Time Series Changes, Math, 4th Grade
Rankings by value-added, ELA, 4th Grade	0.49	--
Rankings by value-added, Math, 4th Grade	--	0.61

Finally, we compare single-year metrics to those based on two years of performance (Table 7). We see these two-year average measures lead to similar rankings, as the rankings are highly correlated (0.99). Given the New York City example, relying on one year of proficiency data at the school level does not appear to be problematic.

Table 7: One-year vs. two-year averages

	One Year	
Two Years	Rankings by Proficiency Rate, ELA, 4th Grade	Rankings by Proficiency Rate, Math, 4th Grade
Rankings by proficiency rate, ELA, 4th Grade	0.99	--
Rankings by proficiency rate, Math, 4th Grade	--	0.99

Overall, based on this analysis in combination with the literature survey, we believe that HUD can use the data it currently has available from the U.S. Department of Education (DoED) to create a useful metric of school quality. One potential issue with the existing data is that high and low scores are truncated, which does not allow for a complete ranking of schools. Although we

do not believe that this is a very large problem in the creation of the overall ratio (as there are few schools at the extremes), we rely on the non-censored version of this data for the construction of our metric in New York City (provided through the New York City Department of Education). Additionally, between the 2009 and 2010 school year, New York changed their proficiency requirements, dramatically lowering the average performance of schools and increasing the variation observed in proficiency rates. We therefore rely on the most recent data for this analysis, as these data provide a more accurate picture of the current performance of New York City school.[xxxiii]

Map 6 and Map 7 show the distribution of high quality schools throughout New York City, using school-level proficiency rates. From these maps, we see that the highest performing schools (with proficiency rates above 80 percent) are in Manhattan, Queens and Brooklyn, and the lowest performing schools are concentrated in the Bronx and Brooklyn. There is a great deal of variation in proficiency rates at the school level in New York City, with proficiency in math ranging from 17 percent to 99 percent (with a mean of 57 percent) and proficiency in ELA ranging from 9 percent to 100 percent (with a mean of 45 percent).

Although we believe that HUD can create a reasonable metric of school quality using its existing data, we still recommend that, in the longer run, the agency add measures of school level value-added, particularly for the lowest income students, as they provide very different information about schools and they will increasingly become available at the state and school district level.

Linking Households to Schools

We link each assisted household to their zoned elementary school using elementary school attendance zone boundaries. For the small number of zones with multiple schools, we create an average measure of school quality based on all schools in a given zone. We assigned the few zones with no schools to their closest elementary school so that each assisted household will be linked to at least one public elementary school.

Construction of Comparison Group

As noted, given that each state relies on different tests and has a different meaning for proficiency, simple mean proficiency rates are not comparable across states. Instead, we recommend that HUD create a ratio measure, which compares the quality of schools that HUD-assisted households can access to those available to other households who have not received housing assistance.

To construct such a comparison group we rely on census block-group data, as previously described. We have linked every census block group with an elementary school attendance zone and its associated school, following the same method we used for assisted households. We rely on census tabulations at the block-group level that report the total number of households, households by income and the number of households by tenure status. In this way we can calculate the quality of the school attended by the typical household in a county or metropolitan area. As for income-based comparisons, since each housing assistance program serves a different population, we create three different income-based comparison groups, based on the 60 percent

percentile income of households in that housing program. We compare public housing residents to other households with incomes below $20,000, and compare voucher households and project based section 8 households to other households with incomes below $15,000. Finally, we compare LIHTC households to other households with incomes below $45,000, as we do not have any information on the income of households living in these developments. Another alternative could be to compare the locations of assisted households to other poor households or near-poor households across the country. Though we do not include this counterfactual in our analysis the methodology for including this comparison group is exactly the same.

Results

Table 8 summarizes our analysis. We find that public housing tenants have access to the lowest quality schools among assisted households. The schools that public housing tenants[xxxiv] can attend have an average proficiency rate of 44.6 percent in math and 33.5 percent in ELA. The tenants in other place-based housing have access to somewhat stronger schools. LIHTC tenants can access schools with proficiency rates of 48.5 percent in math and 37.9 percent in ELA, while Project based Section 8 households have access to schools with proficiency rates of 48.9 percent in math and 37.7 percent in ELA. Housing choice voucher holders have access to slightly stronger schools, with proficiency rates of 50.1 percent in math and 39.3 percent in ELA.[xxxv]

The table also shows that we find that assisted households are zoned for lower performing schools than the average household in NYC. When making the comparison to only renter households, the ratios are closer to one but only slightly. When using an even more restrictive comparison group (only households with the same income) ratios are higher, but again still below one.

Looking at the ratios based on households of each income group, we obtain a ratio of 0.80 in math and 0.76 in ELA for public housing residents, suggesting that other households earning below $20,000 in New York City can access higher quality schools than those attended by public housing residents.[xxxvi] For LIHTC households, we obtain a ratio of 0.84 in math and 0.83 in ELA. For voucher households (choice and project based), we compare them to other households earning below $15,000 and obtain a ratio of 0.88 in math and 0.86 in ELA for project based Section 8 and 0.90 in math and in ELA for housing choice vouchers. In sum, assisted households have access to schools that are lower performing than other households with similar incomes. Of the three programs, voucher holders live in neighborhoods with the highest performing schools in comparison to households of similar incomes, but they still live in neighborhoods with lower performing schools than other households with similar income levels.

Table 8: School quality of zoned elementary schools for assisted households

	Mean Proficiency in Math	Mean Proficiency in ELA
Subsidized Households		
Public housing	44.6%	33.5%
LIHTC	48.5%	37.9%
Project Section 8	48.9%	37.7%
Vouchers	50.1%	39.3%
Comparison Groups		
All households	61.6%	50.1%
Renter households	59.0%	47.6%
Households earning under $45,000	57.8%	45.9%
Households earning under $20,000	56.1%	44.3%
Households earning under $15,000	55.6%	43.8%
Ratios		
Comparison to All households		
Public housing/All households	0.72	0.67
LIHTC/All households	0.79	0.76
Project Section 8/All households	0.79	0.75
Vouchers/All households	0.81	0.78
Comparison to Renter households		
Public housing/Renter households	0.76	0.70
LIHTC/Renter households	0.82	0.80
Project Section 8/Renter households	0.83	0.79
Vouchers/Renter households	0.85	0.83
Comparison to Households by Income Group		
Public housing/under $20,000	0.80	0.76
LIHTC/under $45,000	0.84	0.83
Project Section 8/under $15,000	0.88	0.86
Vouchers/under $15,000	0.90	0.90

Metric of School Choice

We also create a measure of choice available to students at the elementary school level. Results are presented in Table 9. First, we include the number of charter/magnet schools within one mile of the assisted household. We see that the average public housing resident has 3.0 charter/magnets within one mile, the average LIHTC resident has 3.8 charter/magnets in this range, the average voucher holder has 2.6 and the average project based Section 8 household has 3.3 charter or magnet schools in this range. When comparing assisted households to other

households in the same MSA we see that overall assisted households live in neighborhoods with more charter and magnet school options than do other households with similar income levels.

As for the quality of the three closest alternative schools (which is limited to schools within two miles of residents), the average charter/magnet school near public housing residents has a proficiency rate of 65.8 percent in math and 50.2 percent in ELA. For voucher holders, these proficiency rates are 67.1 percent for math and 50.2 percent for ELA. The proficiency rates are similar for households living in project based Section 8 as well, with an average proficiency of 66.4 percent in math and 49.6 percent in ELA. For LIHTC, the average charter/magnet school within 2 miles is somewhat lower performing, with proficiency rates of 48.0 percent in math and 37.6 percent in ELA. We see that the charter options near both public housing residents and voucher holders are higher performing than their zoned schools. For LIHTC households, the performance of the three nearest charters is equivalent to the performance of the zoned school. When comparing the quality of charter/magnet options near assisted households to those near other households we see that overall the quality of charter/magnets is lower for assisted households than for other households in the city. This analysis highlights the importance of considering the quality of the local choice options, as well as the quantity, when creating a measure of local school quality.

Table 9: Charter/Magnet School Options for Assisted Households

	Number of Charter/Magnets within 1 Mile	Mean Proficiency of Chaters within 2 miles in Math	Mean Proficiency of Chaters within 2 miles in Math
Subsidized Households			
Public housing	3.0	65.8%	50.2%
LIHTC	3.8	48.0%	37.6%
Project Section 8	3.3	66.4%	49.6%
Vouchers	2.6	67.1%	50.3%
Comparison Groups			
All households	2.1	70.3%	55.0%
Renter households	2.2	70.1%	54.8%
Households earning under $45,000	2.3	68.7%	52.5%
Households earning under $20,000	2.5	68.3%	52.1%
Households earning under $15,000	2.5	68.2%	52.0%
Ratios			
Comparison to All households			
Public housing/All households	1.42	0.94	0.91
LIHTC/All households	1.80	0.68	0.68
Project Section 8/All households	1.57	0.94	0.90
Vouchers/All households	1.23	0.95	0.91
Comparison to Renter households			
Public housing/Renter households	1.34	0.94	0.92
LIHTC/Renter households	1.70	0.68	0.69
Project Section 8/Renter households	1.48	0.95	0.91
Vouchers/Renter households	1.16	0.96	0.92
Comparison to Households by Income group			
Public housing/under $20,000	1.22	0.96	0.96
LIHTC/under $45,000	1.62	0.70	0.72
Project Section 8/under $15,000	1.32	0.97	0.95
Vouchers/under $15,000	1.04	0.98	0.97

V. Conclusion/Recommendations

We believe that the U.S. Department of Housing and Urban Development (HUD) can create a valid metric for school quality given the data provided through the U.S. Department of Education (DoED). These data include the proficiency rate in math and English Language Arts (ELA) for all students in a school for the 2008-09 school year. We suggest that HUD focus on elementary school quality, as the location of one's home typically determines access to an elementary school (which is not the case for household choices of middle and high schools). We also offer several additional recommendations as more data become available, such as reporting on multiple-year averages, using fourth grade test scores, focusing on the performance of disadvantaged students, and relying on mean scores rather than proficiency levels. However, our study of New York City suggests that these refinements will not fundamentally change the relative ranking of a school. In the longer-run, we also suggest that HUD incorporate measures of value-added at the school level when these data become available, which will provide different—and we believe more reliable—measures of school quality.

As there has recently been a movement toward increasing levels of school choice, we suggest that HUD incorporate a measure of school choice in evaluating school access. Specifically, we recommend that HUD incorporate both the number of charter/magnet schools within one mile of an assisted household as well as the average proficiency rate of the three closest charter/magnet schools.

As for linking households to schools, we suggest that, in the longer run, HUD link assisted households to schools using the School Area Boundary Information System (SABINS) data set, which includes elementary school attendance zones for a large sample of cities. However, results from our New York City case study show that linking households to their nearest school within a school district will link 75 percent of all households to the same schools. Therefore, linking households to the closest school within a district will be reasonably close to replicating the zoned schools approach.

Finally, we recommend that HUD evaluate the quality of schools that assisted households can access in comparison to other households living in the same county or metropolitan area. As states vary in their methods of evaluating schools, and distinct geographic areas have a different composition of households, the creation of this counterfactual will allow HUD to create a meaningful metric of local school quality. This ratio enables HUD to isolate areas that are doing particularly well and those that are doing particularly poorly, as well as programs that are succeeding more than others. This will assist HUD in their efforts to help subsidized households access higher quality schools. In rural areas, we suggest that HUD use counties as the geographic unit of analysis.

We have two additional suggestions for how HUD can best use housing as a platform to improve educational achievement. We believe that HUD can play a crucial role in providing information to assisted households about the quality of local schools and the availability of alternative school choices in their district. We also suggest that HUD could work toward linking assisted housing with opportunities for early childhood education, which could lead to dramatic improvements in

educational attainment for children living in assisted housing. We discuss each of these recommendations below.

HUD as Distributor of Information

We believe HUD can play an important role in using housing as a platform to enhance educational achievement through increasing the access that HUD-assisted households have to information on school quality. DeLuca and Rosenblatt (2010), in their interviews with voucher holders from the Baltimore site of the Moving to Opportunity (MTO) study, find that one of the primary reasons parents did not send their children to a higher performing public school was a lack of information on the quality of nearby schools. Additional qualitative studies find that lower-income families are quite often information poor when it comes to selecting schools for their children (Teske et al. 2007). DeLuca and Rosenblatt cite the *Thompson Housing Mobility Program* as one model of the way in which additional information about educational opportunities can be distributed to assisted households. Also, as the public education system—particularly in urban areas—continues to increase the amount of choice available to households in the public school system, HUD-assisted households will be more likely to take advantage of these increased opportunities if they are well informed of their options.

There is also experimental evidence showing that increased access to information on public schools can dramatically improve the quality of the public schools that families select. Hastings and Weinstein (2008) examine a natural experiment as well as a field experiment where lower income families in a public school choice plan in Charlotte-Mecklensburg were provided with direct information on school test scores. They found that receipt of this information significantly increased the fraction of parents that chose higher-performing schools. They also found that parents with high-scoring nearby alternatives were more likely to choose nonguaranteed schools with higher test scores. They then rely on the random variation in each of these experiments, and identify a causal relationship between attending higher-scoring schools and increased student test scores. These results highlight the importance of information on school quality as well as the importance of good local options in terms of ultimately improving educational outcomes for lower income students.

In creating these information packets, we suggest that HUD include a wide range of indicators about local public schools. As children and families have heterogeneous needs and preferences, providing a broad range of information could improve the match quality of students to schools. As states and school districts collect different information about local schools, we suggest that school reports be created in each locality to provide households with the richest information available about local public schools.

Linking Housing to Early Childhood Education

Finally, there is a growing body of literature which shows that a large portion of the achievement gap between low-income households and higher income households, and between minority households and non-minority households, begins before children are school aged (Reardon 2011). This suggests that even if households with housing assistance are able to gain access to higher quality schools, it may not be enough to close the achievement gap. There is growing

evidence that early childhood interventions can reduce the size of this initial achievement gap as well as lead to sustained improvements in longer term life outcomes. Almond and Currie (2010) review the literature that explores these early childhood interventions, particularly focusing on studies that identify the relationship between early childhood education and longer term life outcomes.

Almond and Currie (2010) find three interventions (which have been studied using sophisticated research designs) that have been shown to be effective in improving early childhood achievement as well as longer term outcomes. The first program they describe is *Nurse Home Visiting*. This program is designed to give young, poor, uneducated, and/or unmarried women assistance with their children at very young ages. In a study by Olds et al. (1999, 2007) which used randomized control trails, they find that by age two children in the nurse home visiting programs are much less likely to be seen in a hospital emergency room for unintentional injuries, and by age 15, the children of visited mothers are less likely to have been arrested or to have run away from home, had fewer sexual partners, and smoked and drank less. They also find these children are less likely to have been involved in verified incidents of child maltreatment. Almond and Currie suggest that this type of parenting education program may be more effective than other home visiting programs such as visits by social workers, as nurses may be more acceptable to parents, and therefore they may be more likely to take advice from a health professional.

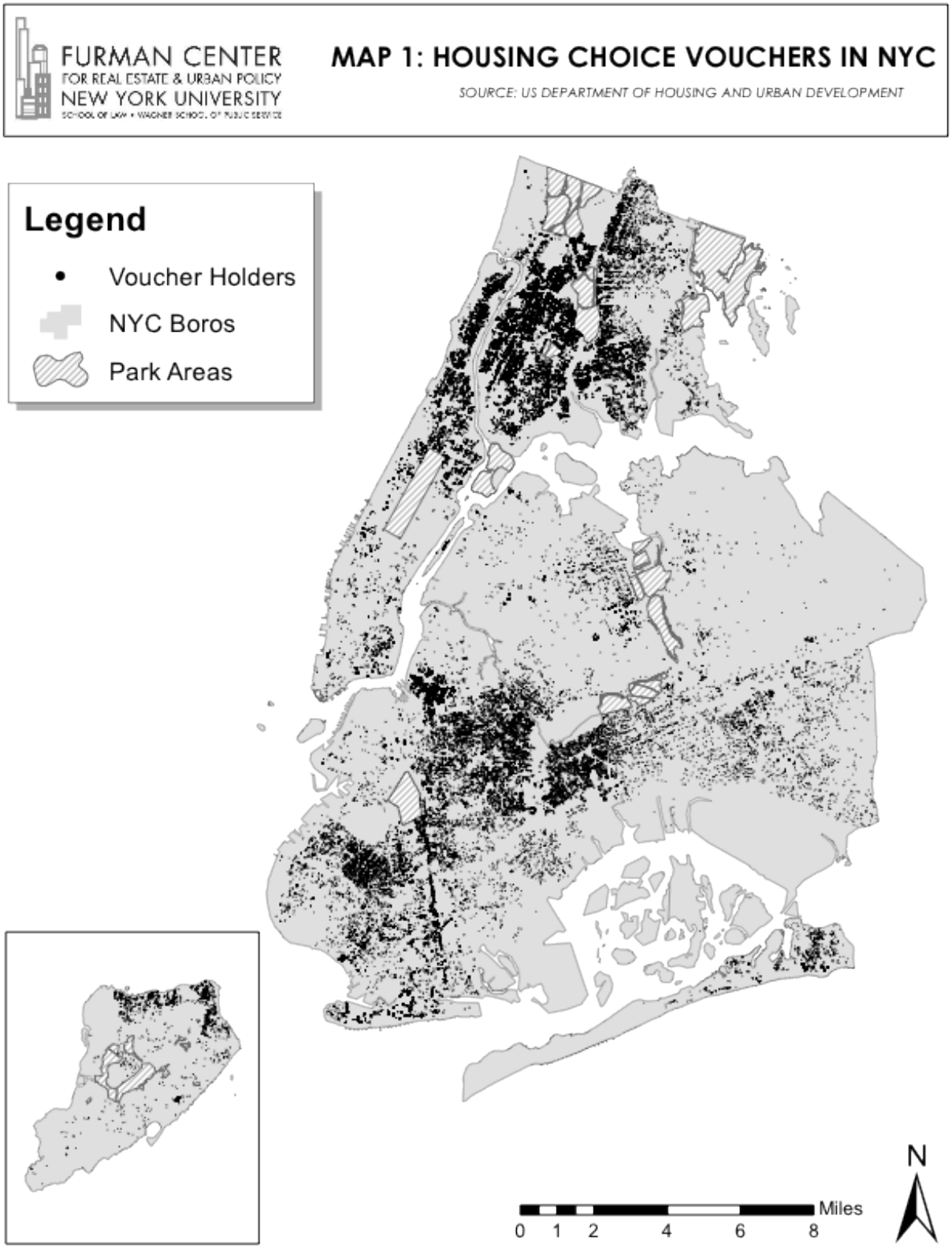
The second program is the U.S. Supplemental Feeding Program for Women, Infants and Children (WIC). WIC is a program targeted at pregnant and lactating women, infants, and children up to age five. Women who participate in this program receive vouchers which they can then use to buy particular types of food at participating retailers. Participants generally have to go to the WIC office to receive these vouchers, and they then receive nutrition education. Many WIC offices are run within clinics and can also facilitate access to medical care. There have been numerous studies of this program[xxxvii] which have shown that women who participate in WIC during pregnancy have longer gestations, have children with higher birth weights, and have generally healthier infants; the effects of these programs tend to be the largest for children of the most disadvantaged mothers. Though many economists criticize these studies for not appropriately adjusting for selection bias, a series of more sophisticated studies have found evidence to support some of these positive effects (though not across the board).

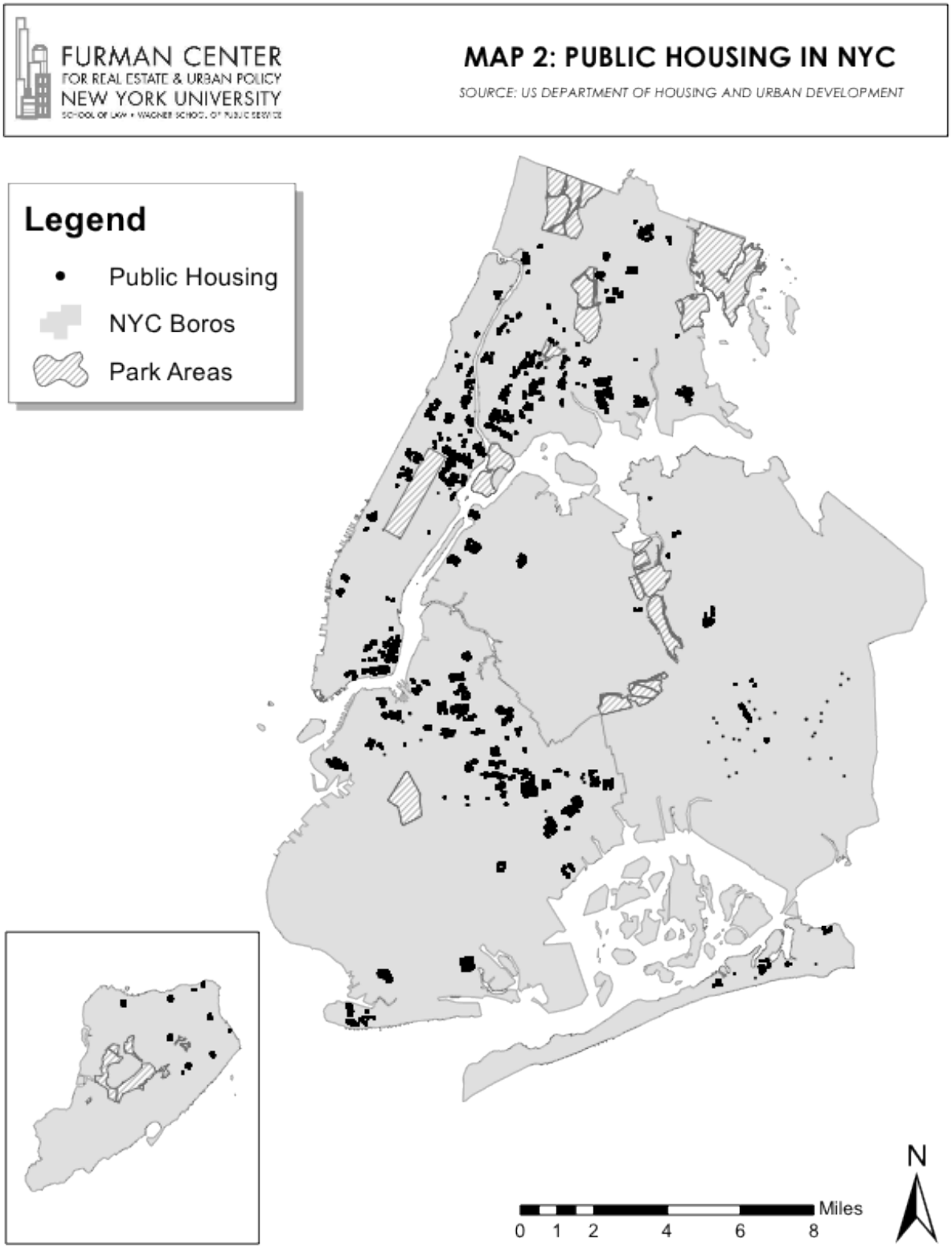
The third program is child care. Almond and Currie report that, overall, many of these studies involve experimental evaluations of model programs that serve relatively small numbers of children, and involve intensive services delivered by well-trained and well-supervised staff. These studies generally find that child care has long-lasting effects on schooling attainment and other outcomes such as teen pregnancy and crime, even when these programs do not result in any lasting increase in cognitive skills as measured by test scores. There have also been a series of evaluations of Head Start, which is a national preschool program for disadvantaged three-, four-, and five-year-olds, which currently serves about 800,000 children. Head Start is not of the same quality as the model interventions, and the quality of these programs varies from center to center, but overall these programs are of higher average quality than other preschool options available to low-income households. Experimental evaluations of Head Start show that participating children make gains, particularly in terms of language ability (Puma et al. 2010). A non-experimental study by Garces et al. (2002) found that participation in Head Start is associated with reductions

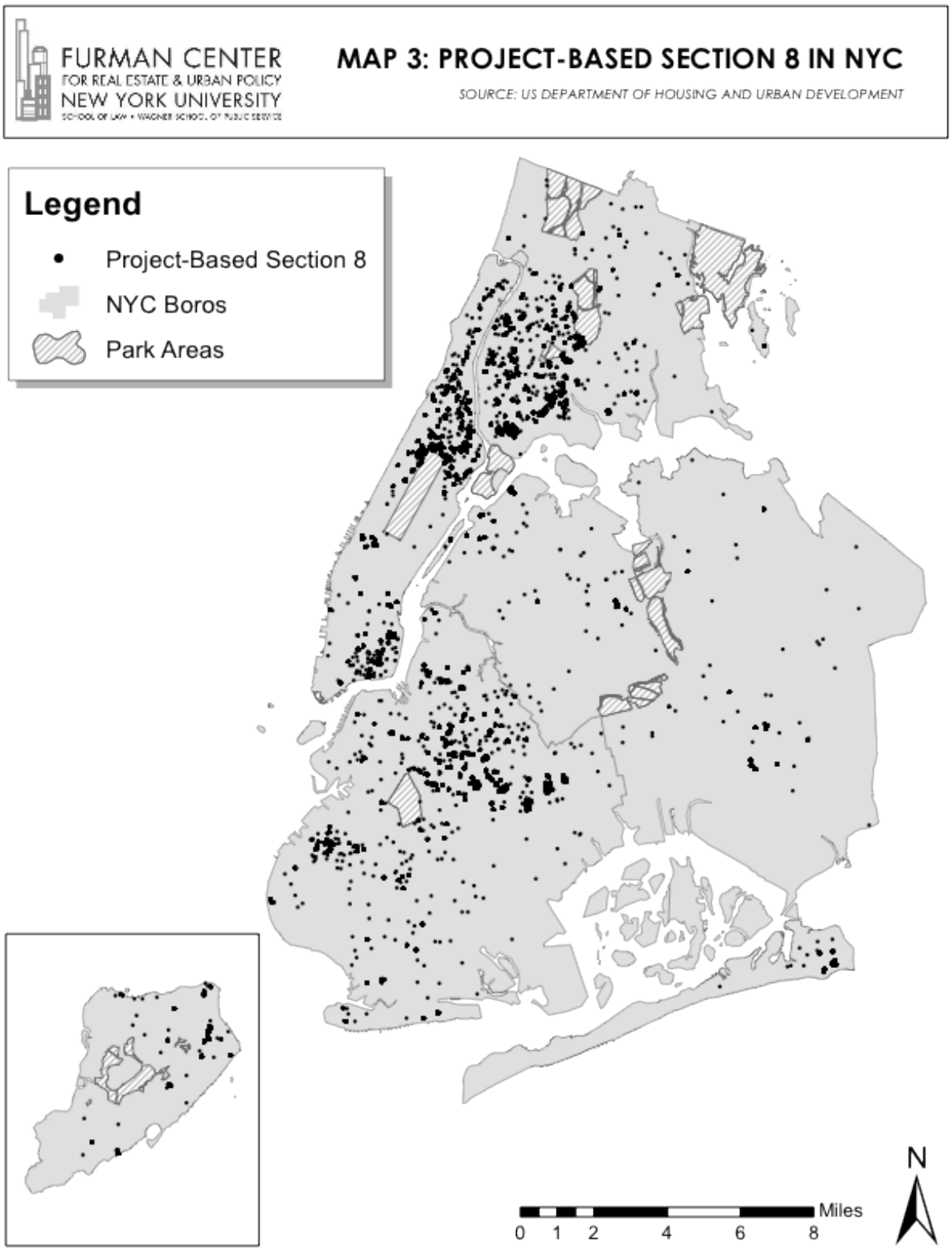
in probabilities of being booked or charged with a crime, for black youths. A more recent study by Carneiro and Ginja (2008), which uses a regression discontinuity approach, finds that participation in Head Start is associated with reductions in behavior problems, grade repetition, depression, and obesity.

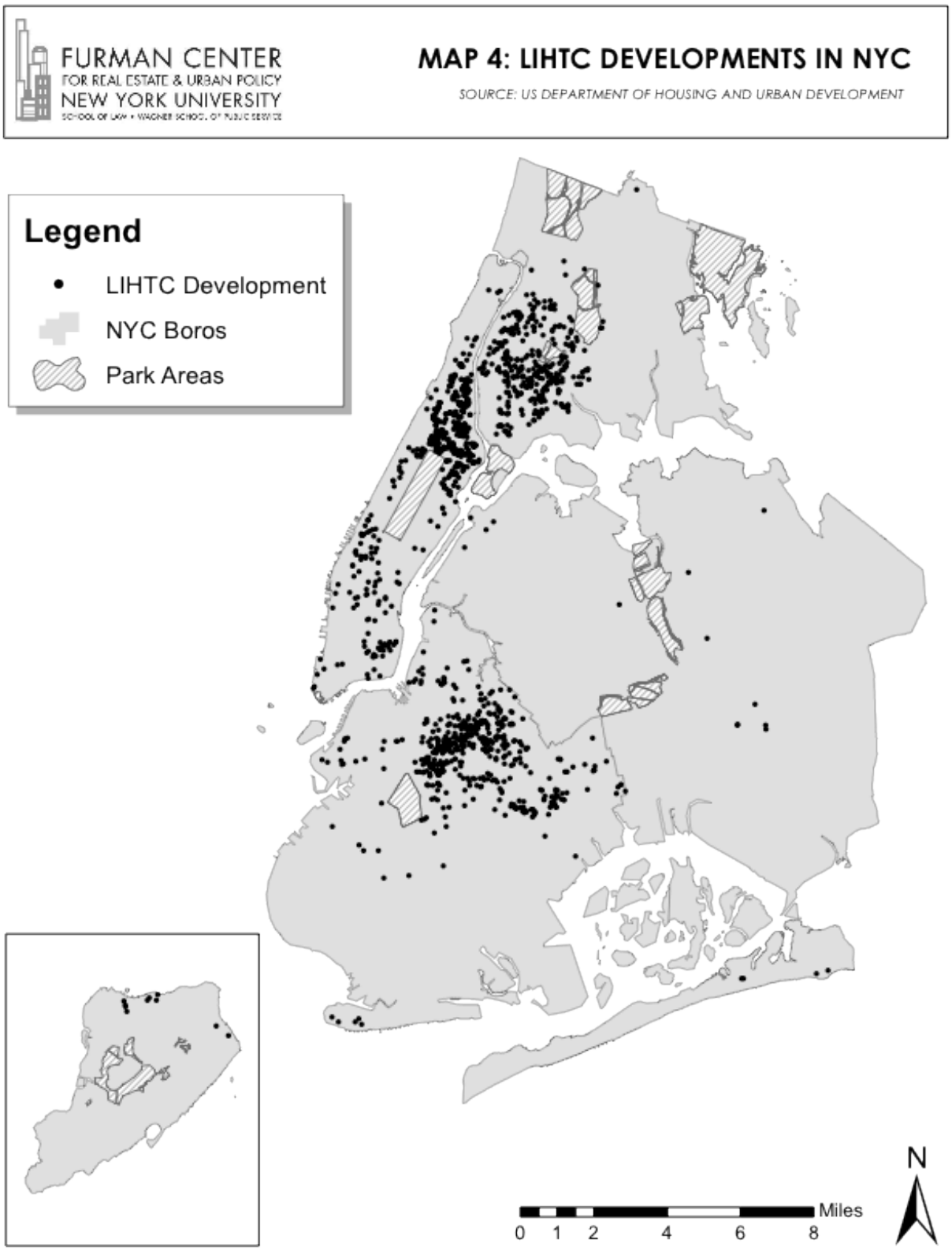
Given this growing body of high quality research, we suggest that one way HUD could help use housing as a platform to increase educational opportunity is through linking assisted housing with early childhood education. Perhaps HUD could provide households with additional resources that enable assisted households to send their children to higher quality preschool or nursery school. Alternatively, HUD could provide information to assisted households about local public assistance programs that provide early childhood education. Helping families with young children navigate existing programs could help dramatically improve educational outcomes for the children of HUD-assisted households.

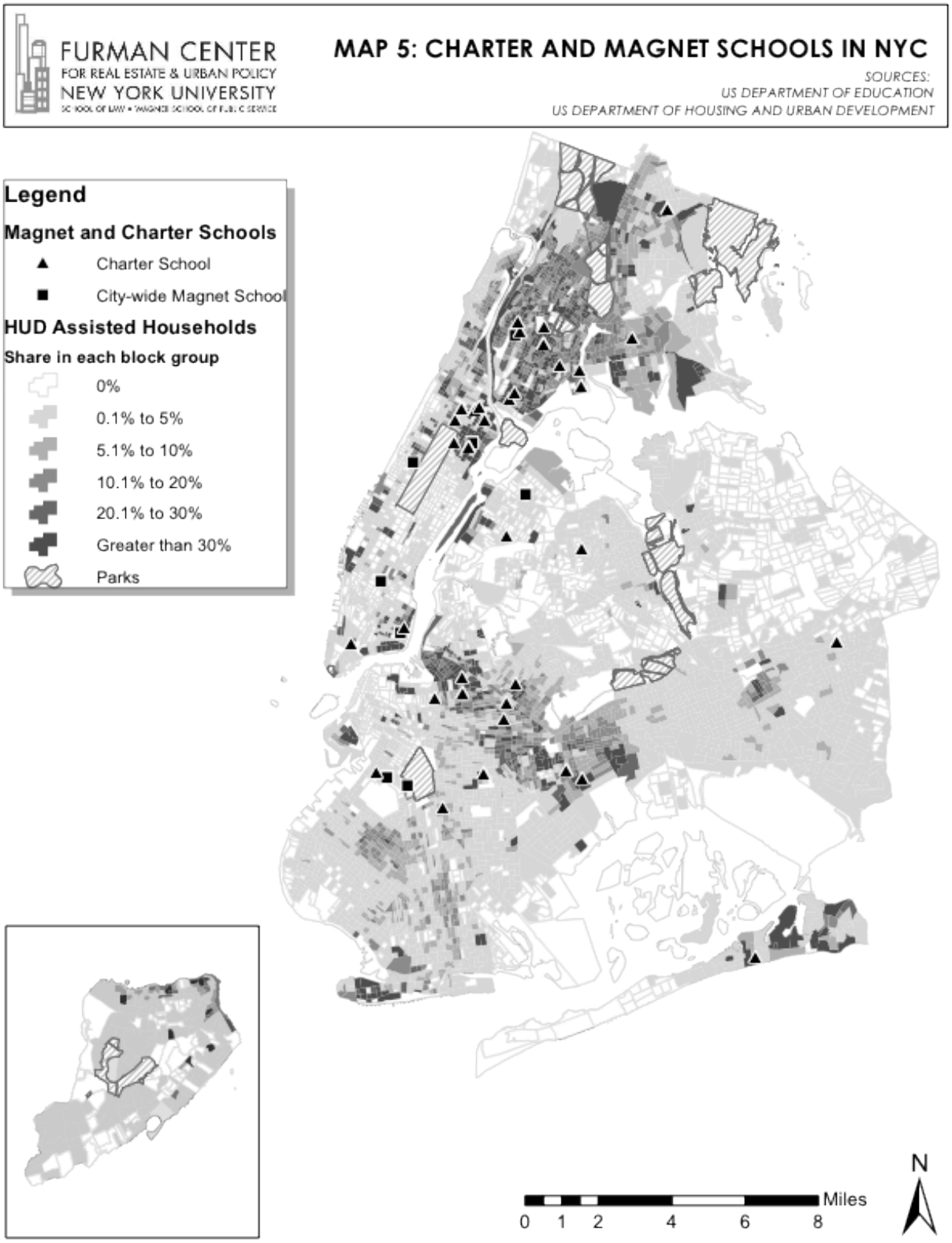
Appendix: Maps

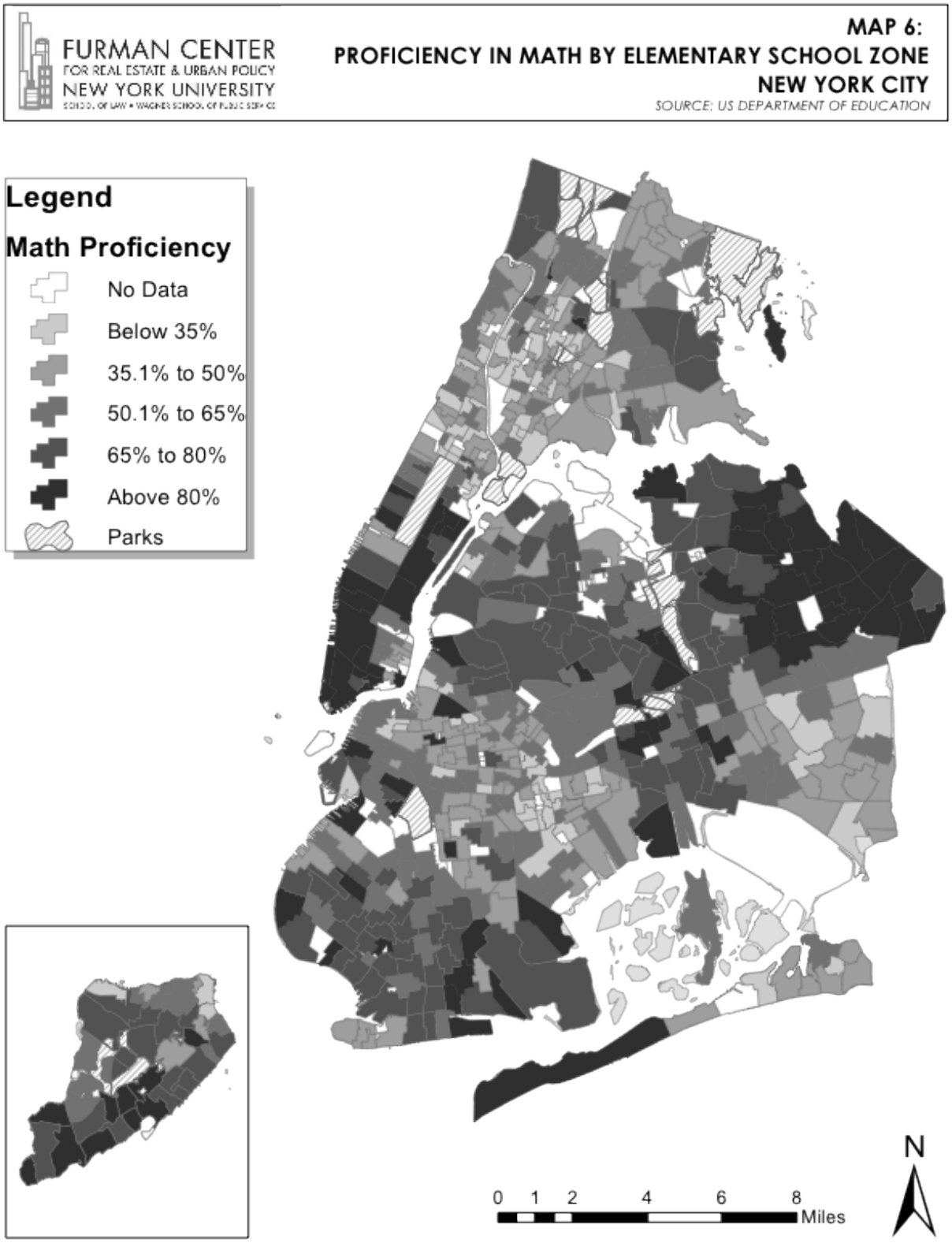


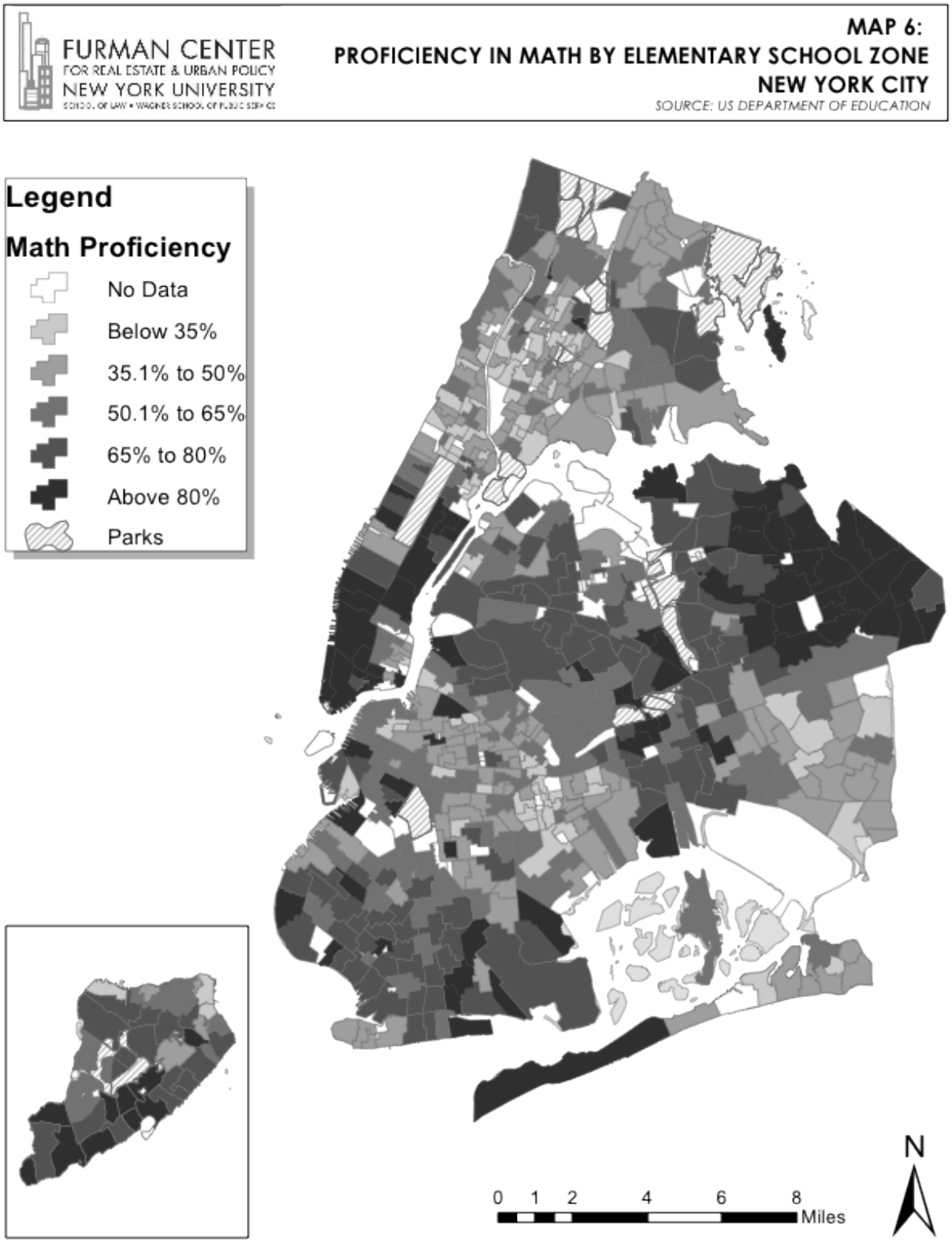

FURMAN CENTER
FOR REAL ESTATE & URBAN POLICY
NEW YORK UNIVERSITY
MAP 2: PUBLIC HOUSING IN NYC
SOURCE: US DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
Legend
Public Housing
NYC Boros
Park Areas
Miles
0 1 2 4 6 8
N

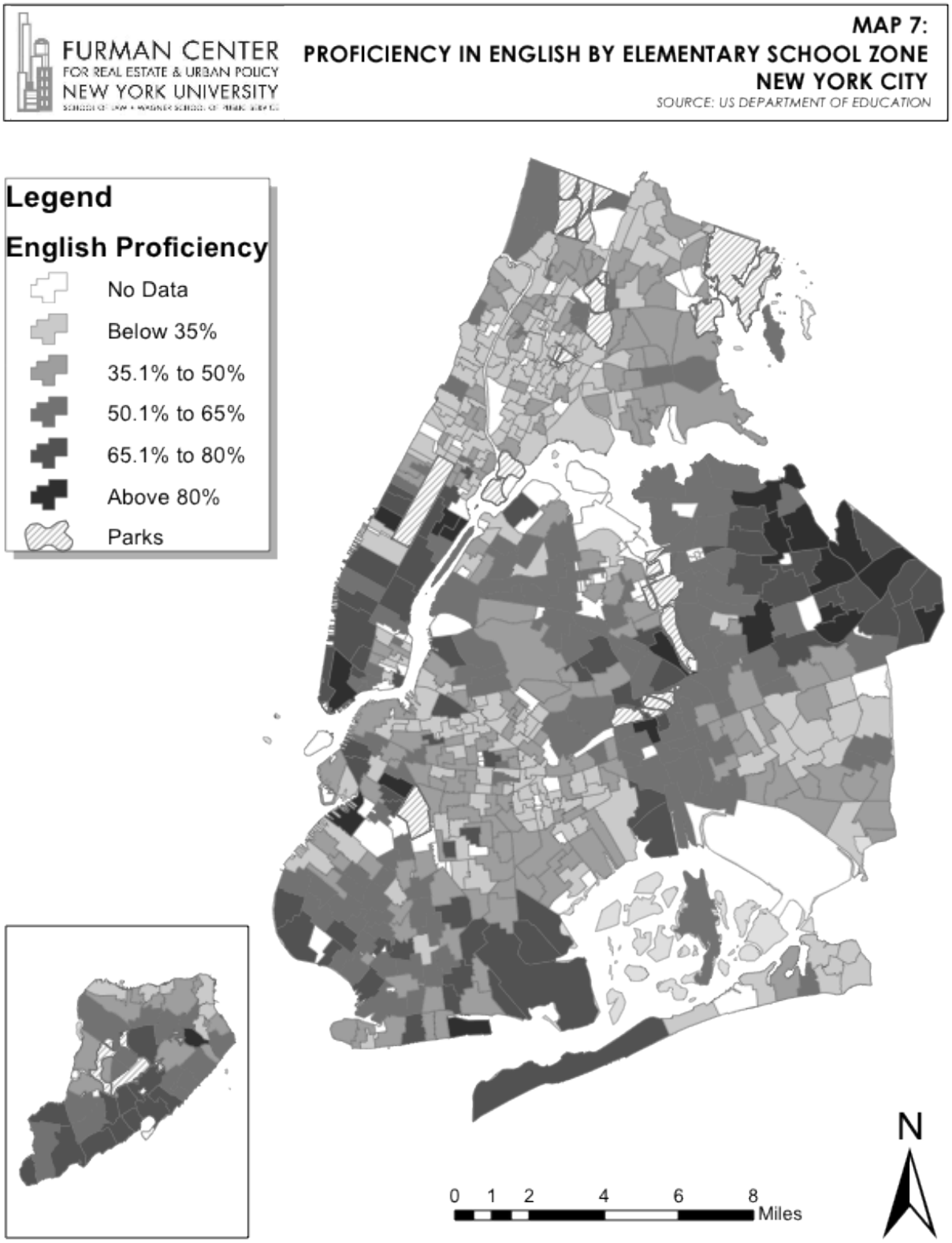

FURMAN CENTER
FOR REAL ESTATE & URBAN POLICY
NEW YORK UNIVERSITY
MAP 3: PROJECT-BASED SECTION 8 IN NYC
SOURCE: US DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
Legend
Project-Based Section 8
NYC Boros
Park Areas
Miles
0 1 2 4 6 8
N

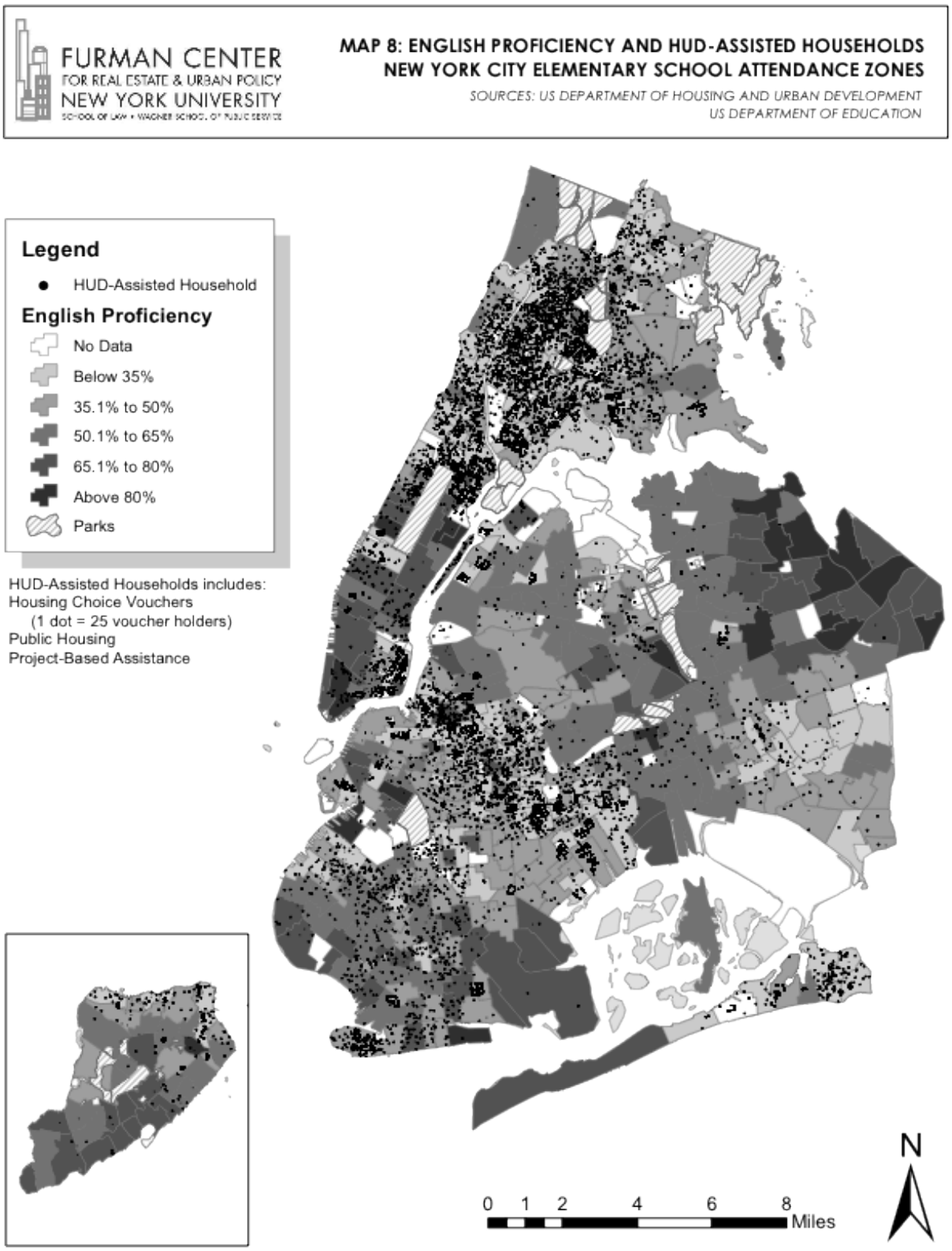

FURMAN CENTER
FOR REAL ESTATE & URBAN POLICY
NEW YORK UNIVERSITY
MAP 4: LIHTC DEVELOPMENTS IN NYC
SOURCE: US DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
Legend
LIHTC Development
NYC Boros
Park Areas
Miles
0 1 2 4 6 8
N


FURMAN CENTER
FOR REAL ESTATE & URBAN POLICY
NEW YORK UNIVERSITY
MAP 5: CHARTER AND MAGNET SCHOOLS IN NYC
SOURCES:
US DEPARTMENT OF EDUCATION
US DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
Legend
Magnet and Charter Schools
Charter School
City-wide Magnet School
HUD Assisted Households
Share in each block group
0%
0.1% to 5%
5.1% to 10%
10.1% to 20%
20.1% to 30%
Greater than 30%
Parks
Miles
0 1 2 4 6 8
N


FURMAN CENTER
FOR REAL ESTATE & URBAN POLICY
NEW YORK UNIVERSITY
MAP 6:
PROFICIENCY IN MATH BY ELEMENTARY SCHOOL ZONE
NEW YORK CITY
SOURCE: US DEPARTMENT OF EDUCATION
Legend
Math Proficiency
No Data
Below 35%
35.1% to 50%
50.1% to 65%
65% to 80%
Above 80%
Parks
0 1 2 4 6 8
Miles
N


FURMAN CENTER
FOR REAL ESTATE & URBAN POLICY
NEW YORK UNIVERSITY
MAP 6:
PROFICIENCY IN MATH BY ELEMENTARY SCHOOL ZONE
NEW YORK CITY
SOURCE: US DEPARTMENT OF EDUCATION
Legend
Math Proficiency
No Data
Below 35%
35.1% to 50%
50.1% to 65%
65% to 80%
Above 80%
Parks
0 1 2 4 6 8
Miles
N


FURMAN CENTER
FOR REAL ESTATE & URBAN POLICY
NEW YORK UNIVERSITY
MAP 7:
PROFICIENCY IN ENGLISH BY ELEMENTARY SCHOOL ZONE
NEW YORK CITY
SOURCE: US DEPARTMENT OF EDUCATION
Legend
English Proficiency
No Data
Below 35%
35.1% to 50%
50.1% to 65%
65.1% to 80%
Above 80%
Parks
0 1 2 4 6 8
Miles
N


FURMAN CENTER
FOR REAL ESTATE & URBAN POLICY
NEW YORK UNIVERSITY
MAP 8: ENGLISH PROFICIENCY AND HUD-ASSISTED HOUSEHOLDS
NEW YORK CITY ELEMENTARY SCHOOL ATTENDANCE ZONES
SOURCES: US DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
US DEPARTMENT OF EDUCATION
Legend
HUD-Assisted Household
English Proficiency
No Data
Below 35%
35.1% to 50%
50.1% to 65%
65.1% to 80%
Above 80%
Parks
HUD-Assisted Households includes:
Housing Choice Vouchers
(1 dot = 25 voucher holders)
Public Housing
Project-Based Assistance
0 1 2 4 6 8
Miles
N

Endnotes

[i] As of the beginning of the 2010 school year, 40 states have adopted the Common Core state standards, so within the next couple of years national indicators should be available at the school level. http://www.corestandards.org/.

[ii] For details on the law see http://www.nochildleftbehind.com/nclb-law-contents.html.

[iii] http://www.edtrust.org/dc/publication/the-abcs-of-ayp

[iv] See Machin and Black (2010) for a review of this literature.

[v] http://www.cgp.upenn.edu/ope_value.html#9.

[vi] To construct this measure Hanushek et al divide test scores into 20 intervals. They then compare the gain that each student made to the gains of other students that began the year in the same interval. To aggregate this measure to the school level they weight the gains of students in each interval by their representation throughout the state.

[vii] See Raudenbush and Bryk (2002) for technical details on the calculation of this value-added measure.

[viii] They calculate two versions of a value-added measure. The first version relies on the absolute value of student gain in the construction of a value-added measure. For the second value-added measure they build a relative measure of student gain, comparing the gain each student made to gains of other students with similar baseline test scores.

[ix] They create their measure of change in annual performance based on the most restricted sample of students, those present for the entire year, but even with this restricted sample the two measures do not provide the same information about a school.

[x] Kane and Staiger (2002) also discuss the "optimal" weights for averaging performance over time, which maximize the reliability of the resulting performance measure. The optimal weights incorporate information on the signal variance, sampling variation, and the degree of persistence in the signal over time. They find that optimally weighted averages of past test scores were much more successful in picking schools that were likely to perform well one or two years in the future, more than doubling the R-squared of the forecast compared to using a single year of data. In most cases, however, simple averages of past test scores (in this case three-year averages) achieved more than half of the gain in forecast performance, suggesting that investment in optimal weights may not be necessary.

[xi] Cullen, Jacob, and Levitt (2006) find no connection between winning a school choice lottery in Chicago and individual test scores, but they do find a connection between attending one's first-choice school and declines in self-reported criminal behavior. Similarly, exploring the school choice lottery in Charlotte-Mecklenburg, Deming (2010) and Deming et al. (2010) find that high school lottery winners from low-performing neighborhood schools are more likely to graduate from high school, attend college, and are less likely to commit a crime, though they find little or no impact of winning a lottery on short-run test scores.

[xii] See Fletcher and Raymond (2002) for a more complete list of alternative indicators that states use to measure school quality.

[xiii] Additional studies exploring the relationship between test scores and long term outcomes include Zax and Rees (1998); Murnane, Willett and Levy (1995); Bishop (1989); Blackburn and Neumark (1993); Bound, Griliches, and Hall (1986); Cameron and Heckman (1993); Cohn and Kiker (1986); and Kiker and Condon (1981).

[xiv] For example, Hayes and Taylor (1996) and Dills (2004) use school district boundaries. For additional studies that rely on school district boundaries, see Machin and Black (2010).

[xv] http://www.wm.edu/as/sabins/.

[xvi] Thiessen polygons are polygons created by ARCInfo through triangulation and have the in the unique feature that each polygon contains only one input point, and any location within a polygon is closer to its associated point than to the point of any other polygon.

[xvii] A charter school is a publicly funded school that has been granted a charter from the state, exempting it from selected state or local rules and regulations. Charter schools are typically governed by a group or organization which is under a contract or charter with the state. In exchange for funding and autonomy, the charter school is required to meet accountability standards. A magnet school is a public elementary or secondary school which provides unique or specialized curriculum in order to attract a diversified student body. Traditionally, magnet schools are distinct from other public schools because they offer specialized academic focuses or themes, known as a magnet program. They are open to students outside the normal school district boundaries.

[xviii] There is a large literature that has explored how much families are willing to pay for schools, and for the most part these studies highlight the strong connection between where a child lives and the elementary school they are able to attend. See Machin and Black (2010) for a review of this literature.

[xix] See Almond and Currie (2010) for a review of the literature that explores the importance of early childhood interventions in the context of longer term human capital development.

[xx] There are 16 different buckets (1) under $10,000, (2) $10,000-$14,999, (3) $15,000-$19,999, (4) $20,000-$24,999, (5) $25,000-$29,999, (6) $30,000-$34,999, (7) $35,000-$39,999 , (8) $40,000-$44,999, (9) $45,000-$49,999, (10) $50,000-$59,999, (11) $60,000-$74,999, (12) $75,000-$99,999, (13) $100,000-$124,999, (14) $125,000-$149,999, (15) $150,000-$199,999, (16) $200,000 or more.

[xxi] In creating this counterfactual we did not subtract assisted households from the block group data, but HUD could consider doing so in the creation of this metric to create a cleaner measure of the locations where unsubsidized households live.

[xxii] See http://www.p12.nysed.gov/osa/reports/2009/ela-techrep-09.pdf for details on ELA and http://www.p12.nysed.gov/osa/reports/2009/math-techrep-09.pdf for details on math.

[xxiii] For details of the Bookmark Standards Setting Procedure in ELA, see http://www.p12.nysed.gov/osa/pub/2006/els-sstr-06.pdf and in math, see http://www.p12.nysed.gov/osa/pub/2006/math-sstr-06.pdf.

[xxiv] http://www.whive.com/KingBee/blog/&p=5.

[xxv] Before a student's growth percentile is calculated their current score may be adjusted if the student is in a special education program, based on the school's percentage of Title 1 Free Lunch students. This means that the typical median will be above 50.0 percentile as the demographic adjustments can only raise a student's growth percentile.

[xxvi] http://schools.nyc.gov/NR/rdonlyres/50117A5B-DB95-4231-82CC-16C853A8EE2B/0/CharterSchoolDirectory_English.pdf.

[xxvii] See http://www.nber.org/~schools/charterschoolseval/nyc_charter_schools_report_july2007.pdf for more details about NYC charter schools.

[xxviii] The HUD data set does not include data on households living in LIHTC or state or locally-subsidized developments.

[xxix] There are a few smaller programs reported in the HUD dataset which we did not include in the case study as they comprise a very small share of assisted households.

[xxx] This is based on the publicly available data on assisted households, Picture of Subsidized Households 2008. The remaining programs that are not included in the analysis comprise a small share of the HUD portfolio and include 236 Projects and other multifamily assisted properties with Federal Housing Authority (FHA) insurance or HUD subsidy.

[xxxi] We define households that report being both black and Hispanic as black. Otherwise all racial categories other than Hispanic include non-Hispanic households.

[xxxii] As New York City does not calculate the growth rate for economically disadvantaged students, we rely on the growth rate for the bottom third of the school population as our key subgroup indicator.

[xxxiii] We construct our ratios using the censored and uncensored 2008-09 test scores as well and find very small differences between these ratios (the largest gap is 0.03).

[xxxiv] We restrict our analysis to non-elderly public housing units. The non-restricted results are quite similar, with differences in proficiency rates of about 1 percent.

[xxxv] HUD could also consider focusing entirely on units with two or more bedrooms as these are where households with children will most likely live. Also, HUD could focus only on housing choice voucher holders with children.

[xxxvi] This is consistent with findings reported in Schwartz et al. (2010).

[xxxvii] See Currie (2003) for a review of these studies.

References

Almond, Douglas and Janet Currie. 2010. "Human Capital Development Before Age Five." NBER Working Paper No. w15827. http://www.nber.org/papers/w15827.

Amrein-Beardsley, Audrey. 2009. "Value-Added Tests: Buyers, Be Aware." *Educational Leadership* 67(3): 38-42.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115(4): 588-638. doi: 10.1086/522381.

Been, Vicki, Ingrid G. Ellen, Amy E. Schwartz, Leanna Stiefel, and Meryl Weinstein. 2010. "Does Losing Your Home Mean Losing Your School?: Effects of Foreclosure on the School Mobility of Children." Paper presented at the Association for Public Policy Analysis and Management Conference, Boston, MA, November 2010.

Bishop, John. 1989. "Is the Test Score Decline Responsible for the Productivity Growth Decline?" *American Economic Review* 79(1): 178-97. http://www.jstor.org /stable/1804780.

Black, Sandra. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* 114(2): 577-99. http://www.spa .ucla.edu/ps/pdf/s99/PS294assign/Black.pdf.

Blackburn, McKinley L. and David Neumark. 1993. "Omitted-Ability Bias and the Increase in Returns to Schooling." *Journal of Labor Economics* 11(3): 521-44. http://www.jstor.org/stable/2535084.

Bound, John, Zvi Griliches, and Bronwyn H. Hall. 1986. "Wages, Schooling and IQ of Brothers and Sisters: Do the Family Factors Differ?" *International Economic Review* 27(1): 77-105. http://www.jstor.org/stable/2526608.

Boyd, Don, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2002. "Do High-Stakes Tests Affect Teachers' Exit and Transfer Decisions? The Case of the 4th Grade Test in New York State." Stanford Graduate School of Education, Stanford, CA. Mimeo.

Cameron Steven V. and James J. Heckman. 1993. "The Nonequivalence of High School Equivalents." *Journal of Labor Economics* 11(1): 1-47. http://www.jstor.org /stable/2535183.

Carneiro, Pedro and Rita Ginja. 2008. "Preventing behavior problems in childhood and adolescence: Evidence from Head Start." Prepared for the Society of Labor Economists Annual Meetings, New York, NY, March 2008. http://client.norc.org/jole/soleweb /8201.pdf.

Cascio, Elizabeth and Diane W. Schanzenbach. 2007. "First in the Class? Age and the Education Production Function." NBER Working Paper No. 13663. http://www .nber.org/papers/w13663.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2010. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." NBER Working Paper No. w16381. http://www.nber.org/papers/w16381.

Choi, Kilchan, Pete Goldschmidt and Kyo Yamashiro. 2006. "Exploring Models of School Performance: From Theory to Practice." CSE Report 673, National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation, University of California. http://www.cse.ucla.edu/products/reports/r673.pdf.

Cohn, Elchanan and Kiker, B. F. 1986. "Socioeconomic Background, Schooling, Experience and Monetary Rewards in the United States." *Econometrica* 53(212): 497-503. http://www.jstor.org /stable/2554099.

Coleman, James S. 1966. "Equality of Educational Opportunity (Coleman) Study, 1966. Codebook and Study Report." Ann Arbor: Inter-university Consortium for Political and Social Research.

Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika*. 16(3): 297-334.

Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt. 2006. "The Effect of School Choice on Participants Evidence from Randomized Lotteries." *Econometrica* 74(5): 1191-1230. http://www.jstor.org/stable/3805923.

Cullen, Julie B. and Randall Reback. 2006. "Tinkering Toward Accolades: School Gaming under a Performance Accountability System." In ed. Michael Baye and John Maxwell, *Improving School Accountability*, *Advances in Applied Microeconomics* 14: 1-34. doi: 10.1016/S0278-0984(06)14001-8.

Cunha, Flavio, James J. Heckman, Lance J. Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, edited by Eric A. Hanushek and Frank Welch, 697-812. Amsterdam: North-Holland: Elsevier.

Currie, Janet. 2003. "Means-Tested Transfer Programs in the United States." In *U.S. Food and Nutrition Programs*, edited by Robert Moffitt, 199–290. Chicago: University of Chicago Press for NBER.

Currie, Janet and Duncan Thomas. 2001. "Early Test Scores, School Quality and SES: Longrun Effects on Wage and Employment Outcomes." In *Worker Wellbeing in a Changing Labor Market*, *Research in Labor Economics,* edited by Solomon W. Polachek, volume 20. Oxford: JAI.

Currie, Janet and Aaron Yellowitz. 2000. "Are Public Housing Projects Good for Kids?" *Journal of Public Economics* 75(1): 99-124. doi: 10.1016/S0047-2727(99)00065-1.

Dee, Thomas S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *Review of Economics and Statistics* 86(1): 195-201. http://www.jstor .org/stable/3211667.

Deere, Donald and Wayne Strayer. 2001a. "Closing the Gap: School Incentives and Minority Test Scores in Texas." Department of Economics, Texas A&M University, College Station, TX. Mimeo.

——. 2001b. "Putting Schools to the Test: School Accountability, Incentives, and Behavior." Working Paper No. 113, Private Enterprise Research Center, Texas A&M University, College Station, TX. www.tamu.edu/perc/perc/Publication/0411.pdf.

DeLuca, Stephanie and Peter Rosenblatt. 2010. "Does Moving To Better Neighborhoods Lead to Better Schooling Opportunities? Parental School Choice in an Experimental Housing Voucher Program." *Teachers College Record* 112(5): 1443-91.

Deming, David, Justine Hastings, Thomas Kane, and Douglas Staiger. 2010. "School Choice and College Attendance: Evidence from Randomized Lotteries." NBER Working Paper No. w10113, National Bureau of Economic Research. http://www.nber.org /papers/w10113.

Deming, David. 2010. "Better Schools, Less Crime?" Harvard Kennedy School, Cambridge, MA. http://www.sites.google.com/site/demingdavid/Deming _BetterSchoolsLessCrime.pdf.

Deng, Lan. 2007. "Comparing the Effects of Housing Vouchers on Low-Income Housing Tax Credits on Neighborhood integration and School Quality." *Journal of Planning Education and Research* 27(1): 20-35.

Derringer, Pam. 2010. "RTT in Tennessee: Assessment Done Right." *Technology and Learning* 31(1): 40-43.

Dills, Angela. 2004. "Do Parents Value Changes in Test Scores? High Stakes Testing in Texas." *Contributions to Economic Analysis and Policy* 3(1). http://proquest.umi.com /pqdlink?did=686499121&sid=1&Fmt=2&clientId=9269&RQT=309&VName=PQD.

Figlio, David N. and Lawrence S. Getzler. 2006. "Accountability, Ability and Disability: Gaming the System?" In *Improving School Accountability*, *Advances in Applied Microeconomics,* edited by Michael Baye and John Maxwell, volume 14: 35-49.

Finn, Jeremy D. and Charles M. Achilles. 1990. "Answers and Questions About Class Size: A Statewide Experiment." *American Educational Research Journal* 27(3): 557-77. http://www .jstor.org/stable/i248864.

Fletcher, Stephen and Margaret Raymond. 2002. "The Future of California's Academic Performance Index." Center for Research on Education Outcomes (CREDO), Stanford University, Stanford, CA. http://credo.stanford.edu/downloads/api.pdf.

Garces, Eliana, Duncan Thomas and Janet Currie. 2002. "Longer-Term Effects of Head Start." *The American Economic Review* 92(4): 999-1012. http://www.jstor.org/stable /3083291.

Graham, Bryan. 2008. "Identifying Social Interactions Through Conditional Variance Restrictions." *Econometrica* 76(3): 643-660. http://www.jstor.org/stable/40056460.

Goldhaber, Dan, and Emily Anthony. 2007. "Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching." *The Review of Economics and Statistics* 89(1): 134-150. doi: 10.1162/rest.89.1.134.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2001. "Disruption versus Tiebout improvement: the costs and benefits of switching schools." *Journal of Public Economics* 88(9-10): 1721-1746. doi:10.1016/S0047-2727(03)00063-X.

Hanushek, Eric A. and Margaret E. Raymond. 2002. "Lessons about the Design of State Accountability Systems." Paper prepared for *Taking Account of Accountability: Assessing Policy and Politics*, Harvard University, Cambridge, MA, June 2002.

Hanushek, Eric A., Margaret E. Raymond, and Steven G. Rivkin. 2004. "Does it Matter How We Judge School Quality?" Paper prepared for annual meetings of the American Education Finance Association, Salt Lake City, UT, March 2004.

Hastings, Justine S. and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics* 123(4): 1373-1414.

Hayes, Kathy J. and Lori L. Taylor. 1996. "Neighborhood School Characteristics: What Signals Quality to Homebuyers?" *Federal Reserve Bank of Dallas Economic Review* 4: 2-9.

Hoffman, Lee. 2009. "Numbers and types of public elementary and secondary schools from the Common Core of Data: School year 2007-08." NCES 2010-305, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC. http://www.nces.ed.gov/pubsearch/pubsinfo.asp ?pubid=2010305.

Jacob, Brian A. 2004. "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago." *The American Economic Review* 94(1): 233-58. http://www.jstor.org/stable/3592777.

——. 2005. "Accountability, Incentives and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89(5–6): 761–96. doi:10.1016 /j.jpubeco.2004.08.004.

Jacob, Brian A. and Steven Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118(3): 843–77. http://www.jstor.org/stable/25053925.

Kane, Thomas J. and Douglas O. Staiger. 2002. "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." *Brookings Papers on Education Policy* 5: 235-83. doi: 10.1353/pep.2002.0010.

Kiker, B.F. and C.M. Condon. 1981. "The Influence of Socioeconomic Background on the Earnings of Young Men." *The Journal of Human Resources* 16(1): 99-104. http://www.jstor.org /stable/145221.

Koretz, Daniel, Sheila Barron, Karen J. Mitchell, and Brian M. Stecher. 1996. "The Perceived Effects of the Kentucky Instructional Results Information System." Rand Monograph Report MR-792-PCT/FF, Rand Corporation, Santa Monica, CA. http://www. rqpi.org/pubs/monograph _reports/2007/MR792.pdf.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114(2): 497-532. http://www.jstor.org/stable/i324116.

Ladd, Helen. 2001. "School-Based Educational Accountability Systems: The Promise and the Pitfalls." *National Tax Journal* 54(2): 385-400. http://proquest.umi.com/pqdlink ?did=75352023&sid=1&Fmt=4&clientId=9269&RQT=309&VName=PQD.

Lazear, Edward P. 2006. "Speeding, Terrorism, and Teaching to the Test." *Quarterly Journal of Economics* 121(3): 1029–61.

Machin, Stephen and Sandra E. Black. 2010. "Housing Valuations of School Performance." Chapter prepared for ed. Eric Hanushek and Finis Welch, *Handbook of the Economics of Education.* Amsterdam: North-Holland: Elsevier.

Murnane, Richard J., John B. Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *The Review of Economics and Statistics* 77: 251-66. http://www.jstor.org/stable/2109863.

Murnane, Richard J., John B. Willet, Yves Duhaldeborde, and John H. Tyler. 2000. "How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?" *Journal of Policy Analysis and Management* 19(4): 547-68. http://www.jstor.org/stable/3325574.

Neal, Derek A. and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104: 869-95. http://www.jstor .org/stable/2138945.

Neal, Derek A. and Diane W. Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *The Review of Economics and Statistics* 92(2): 263-83.

Newman, Sandra J. and Joseph M. Harkness. 2002. "The Long-Term Effects of Public Housing on Self-Sufficiency." *Journal of Policy Analysis and Management* 21(1): 21-43. http://www.jstor .org/stable/3325946.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. "How Large are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26(3): 237-57. http://www.jstor.org /stable/i370881.

Olds, David L., Charles R. Henderson, Harriet J. Kitzman, John J. Eckenrode, Robert E. Cole, and Robert C. Tatelbaum. 1999. "Prenatal and infancy home visitation by nurses: recent findings." *The Future of Children* 9(1): 44-65. http://www.jstor.org/stable /1602721.

Olds, David L., Harriet Kitzman, Carole Hanks, Robert Cole, Elizabeth Anson, Kimberly Sidora-Arcoleo, Dennis W. Luckey, Charles R. Henderson, Jr., John Holmberg, Robin A. Tutt, Amanda J. Stevenson, and Jessica Bondy. 2007. "Effects of nurse home visiting on maternal and child functioning: Age-9 follow-up of a randomized trial." *Pediatrics* 120(4): 832-45.

Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, Gary Shapiro, Pam Broene, Frank Jenkins, Philip Fletcher, Liz Quinn, Janet Friedman, Janet Ciarico, Monica Rohacek, Gina Adams, and Elizabeth Spier. 2010. "Head Start Impact Study: Final Report; Executive Summary." Technical report prepared for the Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC.

Qualls, Audrey L. 1995. "Estimating the Reliability of a Test Containing Multiple Item Formats." *Applied Measurement in Education* 8(2): 111-120. doi: 10.1207/s15324818 ame0802_1.

Raudenbush, Stephen W. 2004. "Schooling, Statistics, and Poverty: Can We Measure School Improvement?" Presented at Ninth Annual William H. Angoff Memorial Lecture, Educational Testing Service, Princeton, NJ, April 2004.

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, *(2nd edition).* Newbury Park, CA: Sage.

Reardon, Sean F. 2011. "The Widening Academic-Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations." Forthcoming in *Social Inequalities and Educational Disadvantage,* ed. Richard Murnane and Greg Duncan. Washington, DC: Brookings Institution.

Rothstein, Richard. 2004. "Class and Schools; Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap." Washington DC: Economic Policy Institute.

Schwartz, Amy E., Brian J. McCabe, Ingrid G. Ellen and Colin C. Chellman. 2010. "Public Schools, Public Housing: The Education of Children Living in Public Housing." *Urban Affairs Review* 46(1): 68-89. doi:10.1177/1078087410367780.

Shepard, Lorrie. 1988. "Should instruction be measurement driven? A Debate." Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, January 1988.

Sojourner, Aaron. 2009. "Inference on peer effects with missing peer data: Evidence from Project STAR." Carlson School of Management, University of Minnesota. Mimeo.

Teske, Paul, Jody Fitzpatrick and Gabriel Kaplan. 2007. "Opening Doors: How Low-Income Parents Search for the Right School." Center on Reinventing Public Education. Daniel J. Evans School of Public Affairs, University of Washington. http://www.crpe.org /cs/crpe/view/csr_pubs/106.

Word, Elizabeth, John Johnston, Helen P. Bain, B. Dewayne Fulton, Charles M. Achilles, Marth N. Lintz, John Folger, and Carolyn Breda. 1990. "The State of Tennesee's Student/Teacher

Achievement Ratio (STAR) Project: Technical Report 1985-1990." Tennessee State Department of Education.

Zax, J., and D. Rees. 1998. "Environment, Ability, Effort and Earnings." Working Paper 9801, Center for Research on Economic and Social Policy, Denver, CO. www.colorado .edu/econ/CEA/papers98/wp98-35.pdf.