NBER WORKING PAPER SERIES

NEIGHBORHOODS AND ACADEMIC ACHIEVEMENT:
RESULTS FROM THE MOVING TO OPPORTUNITY EXPERIMENT

Lisa Sanbonmatsu
Jeffrey R. Kling
Greg J. Duncan
Jeanne Brooks-Gunn

ABSTRACT

Families originally living in public housing were assigned housing vouchers by lottery, encouraging

moves to neighborhoods with lower poverty rates. Although we had hypothesized that reading and

math test scores would be higher among children in families offered vouchers (with larger effects

among younger children), the results show no significant effects on test scores for any age group

among over 5000 children ages 6 to 20 in 2002 who were assessed four to seven years after

randomization. Program impacts on school environments were considerably smaller than impacts

on neighborhoods, suggesting that achievement-related benefits from improved neighborhood

environments are alone small.

Lisa Sanbonmatsu
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138
lsanbonm@nber.org

Jeffrey R. Kling
The Brookings Institution
1775 Massachusetts Avenue NW
Washington, DC 20036
and NBER
jkling@brookings.edu

Greg J. Duncan
Institute for Policy Research
Northwestern University
2040 Sheridan Road
Evanston, IL 60208
greg-duncan@northwestern.edu

Jeanne Brooks-Gunn
Teachers College
Columbia University
New York, NY 10027
brooks-gunn@columbia.edu

I. Introduction

Children educated in large urban school districts in the U.S. have substantially lower academic performance than children in the nation as a whole.[1] Children attending schools with high concentrations of poor students fare particularly poorly, facing numerous disadvantages including less-educated parents, low performing schools, and distressed communities outside of school (Lippman et al., 1996). In an attempt to identify the effects of social context that are distinct from individual and family factors, this paper examines the extent to which changes in residential neighborhood affect children's academic achievement.

Our analysis utilizes a randomized housing mobility experiment, the Moving to Opportunity (MTO) for Fair Housing demonstration program of the U.S. Department of Housing and Urban Development (HUD), to estimate the causal effects on children's educational outcomes of moving out of high-poverty neighborhoods. Through a lottery for housing vouchers among families initially living in public housing, MTO randomly assigned families into three groups. Families in an "experimental" group received housing vouchers eligible for use in low-poverty neighborhoods. Families in a "Section 8" group received traditional housing vouchers without neighborhood restrictions. Families in a control group did not receive either voucher, but were still eligible for public housing.

While family and individual attributes may strongly influence children's educational outcomes, the experimental design of MTO enables us to isolate the impact of residential neighborhood characteristics on educational outcomes. If neighborhoods influence the quality and learning environment of schools attended, then residential relocation programs such as MTO should improve educational outcomes among children who experience moves through the

---

[1] See the National Assessment of Educational Progress for urban schools (U. S. Department of Education, 2003a, 2003b).

1

program. Neighborhoods may also affect the educational norms, values, and resources in the community outside of school. These community influences may be particularly important for young children who have spent the largest fraction of their lives in new locations, may be more adaptable to a new social environment, and are learning language at a rapid rate (Shonkoff and Phillips, 2000). Children not old enough to attend school prior to their families' MTO enrollment had the opportunity to begin their schooling in a more affluent neighborhood. Both school readiness and school success in the early grades may be important for later school success and human capital formation (Rouse, Brooks-Gunn, and Sara McLanahan, 2005; Slavin, Karweit, and Wasik, 1993; Heckman, 2000).

In the analysis that follows, we focus on estimating the magnitude of impacts on educational outcomes and evaluating the mechanisms through which neighborhoods may produce them. In addition to assessing the test scores and behavioral gains overall, we test the hypothesis that younger children would experience greater gains than older children. We also investigate the possibility of differential effects based on demographic characteristics including gender, race and ethnicity, and educational risk factors. Furthermore, using school address histories and data on school characteristics (including self-reports on school climate), we analyze the extent to which moves out of high-poverty neighborhoods imply moves to higher-quality schools – a principal mechanism through which residential mobility programs can affect educational outcomes.

In section II, we review existing literature on the association between neighborhoods and educational outcomes. In section III, we present the details of the MTO program. Section IV discusses our data sources, and section V outlines our econometric approach. Sections VI and VII present our results, and section VIII concludes.

II. Existing Literature

The impact of neighborhoods on children's outcomes is subject to wide debate.[2] From a theoretical perspective, residential mobility and the sorting of individuals into neighborhoods is a key factor in the production of human capital (Benabou, 1993; Fernandez, 2002). Some researchers argue that early childhood environments, in combination with individual attributes and family background, influence subsequent outcomes much more than environmental conditions in later childhood or adolescence (Bouchard, 1997; Duncan et al., 1998; Shonkoff and Phillips, 2000). Developmental theory and studies of school failure suggest that arguments concerning the importance of early influences may be particularly relevant for educational achievement (Slavin, Karweit, and Wasik, 1993). Others believe that disadvantaged neighborhoods may have adverse effects on adolescent development by depriving youth of positive peer influences, adults who provide role models and actively monitor neighborhood events, and school, community, and healthcare resources, as well as by exposing them to violence (Sampson, Raudenbush, and Earls, 1997).

In contrast to theories about the deleterious effects of disadvantaged neighborhoods, "relative deprivation" models argue that poor families may actually fare better in low-income neighborhoods; in high-income neighborhoods, these families may face discrimination or may experience resentment. These models predict that children in low-income families living in high-income neighborhoods will exhibit worse outcomes, including low educational attainment, behavioral problems, and diminished mental health (Wood, 1989; Marsh and Parker, 1984; Collins, 1996).

---

[2] For reviews, see Mayer and Jencks (1989), Jencks and Mayer (1990), Brock and Durlauf (2001), Duncan and Raudenbush (2001), Ellen and Turner (1997), Leventhal and Brooks-Gunn (2000), and Sampson, Morenoff, and Gannon-Rowley (2002).

The bulk of the empirical research to date studying neighborhood effects and youth educational outcomes uses non-experimental data, typically linking developmental studies of children to Census data on local area characteristics. For example, studies focusing on the reading achievement and vocabulary outcomes of five- to six-year-olds have generally found that more affluent neighborhoods are associated with higher achievement in comparison with middle income neighborhoods, even after controlling for family sociodemographic characteristics (Chase-Lansdale and Gordon, 1996; Chase-Lansdale, Gordon, Brooks-Gunn, and Klebanov, 1997; Duncan, Brooks-Gunn, and Klebanov, 1994; Kohen et al., 2000). Researchers focusing on older youth have found that higher neighborhood socioeconomic status is associated with higher combined reading and math scores (Halpern-Felsher et al., 1997; Ainsworth, 2002) and greater likelihood of high school graduation (Aaronson 1998).[3] While these non-experimental studies are suggestive, the causal link between neighborhoods and educational outcomes is not clear; observationally-equivalent families selecting to live in different neighborhoods may be different on unobserved characteristics – characteristics that may also influence educational outcomes for their children. Duncan, Boisjoly and Harris (2001) and Solon, Page, and Duncan (2000) show that correlations between neighboring children in their achievement scores and subsequent educational attainment are small once family background is controlled for, suggesting only a limited role for neighborhood factors.

Researchers have attempted to handle concerns about unobservable differences between individuals living in different neighborhoods by using the quasi-experiment of court-ordered remedial programs, in which federal courts have required HUD to provide funding for rental assistance and housing counseling services in order to reduce racial segregation in publicly-

---

[3] For reviews of the literature, see Jencks and Mayer (1990), Brooks-Gunn, Duncan, and Aber (1997), Furstenberg et al. (1999), and Leventhal and Brooks-Gunn (2000).

assisted housing. In an influential study, Rosenbaum (1995) argued that in Chicago's Gautreaux program, residential location was essentially determined by quasi-random waitlist ordering, so that families who moved to suburban locations were comparable to those who moved to other in-city locations. He found that children in suburban neighborhoods had higher satisfaction with teachers and had better attitudes about school, and that high school dropout rates were much lower for suburban children – five percent compared with 20 percent among those in city neighborhoods. Despite the influence of the Gautreaux study, the sample sizes are small and the response rates are low, allowing for the possibility of substantial bias.[4]

In a more recent study of children moving out of public housing in Chicago due to Hope VI demolitions, Jacob (2004) found no effect on children's test scores, and found only small changes in neighborhood circumstances despite departure from public housing. Currie and Yelowitz (2000) found that children in public housing projects were less likely to be held back in school than children in similarly poor families without access to public housing and speculate that this resulted from public housing providing better living conditions than these families would have had in the absence of the public housing.

The MTO research platform addresses the selection problem using a randomized design described in detail in section III. Early MTO work based on about 350 Baltimore children ages 5-12 found large changes in neighborhood circumstances for the experimental group relative to the control group and positive effects on reading and math test scores over the first four years after random assignment (Ludwig, Ladd, and Duncan, 2001). A study of 168 children ages 6-10 at the MTO site in New York did not find effects on test scores for the experimental versus

---

[4] Regarding the quasi-randomness of placements in Gautreaux, Votruba and Kling (2004) find that placements depended upon various factors about the families that housing counselors used in the placement process.

control group overall after three years – although it did find positive effects on test scores for a sample of male youth (Leventhal and Brooks-Gunn, 2004).

While the experiment cannot provide a direct method for distinguishing between different mechanisms through which neighborhoods affect children, it can provide more precise estimates of the impact of neighborhoods on educational and other outcomes. This paper uses data on over 5000 children ages 6-20 at all five MTO sites, looking at medium-term outcomes four to seven years after random assignment, in order to help solidify our understanding of these effects.

III. The Moving to Opportunity Experiment

The MTO demonstration program was designed to assess the impact of providing families living in subsidized housing with the opportunity to move to neighborhoods with lower levels of poverty. Families were recruited for the MTO program from public housing developments in Boston, Baltimore, Chicago, Los Angeles and New York. HUD primarily targeted developments located in census tracts with 1990 poverty rates of at least 40 percent. Program eligibility requirements included residing in a targeted development, having very low income that met the Section 8 income limits of the public housing authority, having a child under eighteen, and being in good standing with the housing authority. From 1994-1997, 4248 eligible families were randomly assigned to one of three groups: a control group (n = 1310), an experimental treatment group (n = 1729), and a Section 8 treatment group (n = 1209).[5]

Each family assigned to the Section 8 group received a housing voucher or certificate that could be used to rent an apartment in the private market, under the standard terms of the federal Section 8 housing program. Each family in the "experimental" group received a similar voucher

---

[5] Families were initially randomly assigned in an 8:3:5 ratio of experimental:Section 8:control group families. The initial ratios were chosen to minimize minimum detectable effects of experimental impacts based on forecasted voucher utilization, and were adjusted over time in response to actual utilization.

or certificate, but that could only be used to rent an apartment in a tract with a poverty rate of less than ten percent (based on 1990 Census data). In order to help the experimental group families comply with this geographic restriction, local non-profits offered these families mobility counseling.[6] The geographic restriction on the experimental group's voucher applied only for the first year, after which the voucher could be used in any tract. Control group families were not offered housing vouchers, but they could continue to live in public or subsidized housing as long as they remained eligible. Treatment group families who did not use their vouchers within the required time period could also remain in public housing.[7] Families residing in public housing or using vouchers to rent apartments in the private market are generally required to pay 30 percent of their adjusted income in rent.

Forty-seven percent of the experimental group families and 59 percent of the Section 8 group families used the program housing voucher to "lease-up," or move to a new apartment. We refer to the families who moved using a voucher as treatment "compliers." By randomly assigning families to different voucher groups, the demonstration was designed to introduce an exogenous source of variation in neighborhood conditions.

---

[6] The non-profits helped families locate and visit units in low-poverty neighborhoods, negotiate rents, and navigate the credit review process (Goering et al., 1999).
[7] Under the Section 8 program, families typically had a maximum of 120 days to search for an apartment. In order to provide MTO families with more time to locate a suitable apartment, HUD allowed the local public housing authorities to delay the issuance of certificates and vouchers for the experimental group to provide these families with a larger window (approximately six months) in which to locate an apartment (Feins, Holin, and Phipps, 1994).

<center>**IV. Sample and Data**</center>

A.   *Sample and Data Sources*

        This paper focuses on test score data collected in 2002 for MTO children who were school age or slightly older (ages six to 20 as of December 31, 2001) at the time of interview.[8] The age range of the sample allows us to examine the impact of neighborhoods on educational achievement and to test the hypothesis of stronger effects for younger rather than older children. Most of our information about educational outcomes comes from data collected in collaboration with Abt Associates and HUD four to seven years after the families entered the MTO program. One adult and up to two children from each family were selected for this data collection. Interviewers administered a battery of achievement tests to the sample children and interviewed those children who were at least 8 years old. The interview asked children about their schools, neighborhoods, friends, health, behavior, and activities. Interviewers also asked adults about their children's behavior, health, schooling, and activities.

        The interview and test score data were collected in two main phases. During the first phase, interviewers attempted to locate and interview all 4248 families and successfully obtained data for 80 percent. Almost all of the interviews were conducted in person using a computer-assisted interview system, with some out-of-state interviews conducted by telephone.

        In the second phase, 30 percent of families without complete data were randomly selected for continued data collection efforts. During the second phase, data were collected from about 49 percent of this subsample. The interviews attempted during the second phase are representative of all non-completes at the end of the first phase, so we can estimate the overall effective response rate (ERR) as the sum of the first phase response rate (R1) plus the subsample response

---

[8] For additional information about these data, see Orr et al. (2003). The child sample was selected from MTO participants ages five to 19 as of May 31, 2001. For this paper, we restrict the sample slightly to children ages six to 20 as of December 31, 2001. The advantage of the latter age definition is that it more closely approximates the age of the children at the time of the interviews, which were conducted primarily during the first half of 2002.

rate (R2) multiplied by the first phase's non-response rate: ERR = R1 + R2*(1 - R1). For the

MTO study, the overall effective response rate was 90 percent. For our child sample, the

effective response rate was 85 percent for achievement test scores (n = 5074 for complete math

and reading scores), 89 percent for child self-reported survey data (n = 4609), and 85 percent for

adult reports about behavior problems (n = 5248).

The surveys completed by families when they applied for the MTO program provide

some baseline information about the children. A regression of achievement test completion on

baseline characteristics and treatment status indicates that the likelihood of having test score data

is not related to treatment status; however, we are more likely to have test score data for children

who did not have learning problems at baseline, who were from the Chicago site, and whose

parents were still in school or did not have a high school diploma at baseline.

B.   *Baseline Characteristics*

Table 1 presents selected baseline characteristics of children for whom we were able to

obtain achievement test scores. The table shows the means for the control (column i),

experimental (column ii), and Section 8 groups (column vi). As panel A shows, the sample is

roughly equally divided between boys and girls. The mean age of the sample at the end of 2001

is just over 13 years old. The sample consists mainly of minority children: approximately two-

thirds are non-Hispanic African-Americans and about 30 percent are Hispanic (black or non-

black). The majority of the sample children are from female-headed households. We used a

series of t-tests to check the statistical significance of differences in 50 characteristics (items

shown in Table 1 as well as the other baseline covariates controlled for in our analyses) between

the control group mean and each treatment group mean. These t-test results show just a small

fraction of variables with differences that are significant at the .05 level; hence, our analytic

sample generally appears to be balanced on observable characteristics across treatment and control groups.

In addition to showing the means for the overall treatment groups, Table 1 shows the means for treatment compliers (columns iii and vii) and non-compliers (columns iv and viii). Experimental group compliance rates are higher for Los Angeles and lower for Chicago than for the other sites. Compared to those from noncomplier families, experimental group children from complier families are more likely to have parents who were younger, never married, on AFDC, still in school, very dissatisfied with their neighborhoods, had less social contact with neighbors, had a household member who had recently been victimized, and who were more optimistic about finding apartments in other parts of the city. Families that successfully leased-up through the program also tended to have teenage children and to have fewer members.

C.   *Data on Neighborhoods and Schools*

The MTO demonstration gave families the opportunity to move to new neighborhoods. For children, new neighborhoods were expected to translate into new schools as well. To assess the extent to which offering families housing vouchers affected the neighborhoods and schools of MTO children, we constructed address and school histories. Residential addresses from baseline until data collection in 2002 were compiled from several sources including contacts with the families, the National Change of Address system, and credit reporting bureaus. Street addresses were geocoded and linked to 1990 and 2000 Census tract data. We linearly interpolate the data for inter-census years and extend this linear trend to extrapolate post-2000 years. We hypothesized that neighborhoods have a cumulative impact on children; thus, we created neighborhood "exposure" measures that reflect the average of the characteristics of all of the

neighborhoods the children lived in between randomization and follow-up, weighting each neighborhood by residential duration.

To construct a school history for each child, interviewers asked the adult for the names and grades of all schools the child had attended since randomization. The names and addresses of the schools allowed us to link the schools to school-level information about student enrollment and school type from the National Center for Education Statistics' (NCES) Common Core of Data (CCD) and Private School Survey (PSS). Additional school-level information was obtained from state education departments and from the National Longitudinal School-Level State Assessment Score Database (NLSLSASD).[9] Interviews with the children provided another source of information about the schools. Children were asked about their school's climate including its safety and the level of disruptions by other students.

D.   *Achievement Test Scores*

Our primary measures of educational achievement are the reading and math scores of MTO children from the Woodcock Johnson-Revised (WJ-R) battery of tests administered by the interviewers (Woodcock and Johnson, 1989, 1990). The test scores have the advantage of being direct measures of children's reading and math achievement, and unlike other performance measures such as grades, are defined consistently regardless of school attended. We chose the WJ-R for the evaluation because it can be used across a wide range of ages, has good internal reliability (high .80s to low .90s on tests), has demonstrated concurrent validity with other commonly used achievement tests (typically in the .60s and .70s for older children), and has been standardized on a nationally representative sample (Woodcock and Mather, 1989, 1990). The WJ-R has been used in national studies such as the Panel Study of Income Dynamics' Child

---

[9] The NLSLSASD is funded by the U.S. Department of Education and contains school-level test score information for U.S. public schools (www.schooldata.org).

Development Supplement (PSID-CDS; Hofferth et al., 1998) and the Head Start Family and Child Experiences Survey (FACES). Regarding the importance of WJ-R scores for predictions, our analysis of the PSID-CDS found that the correlation between scores in 1997 and 2002 were reasonably high, between .5 and .6, for black students ages 8-17 in 2002. WJ-R is highly predictive of whether students are in gifted, normal, or learning disabled classes, and strongly correlated with other tests of reading and math (McGrew et al., 1991).

A child's broad reading score is the average of the child's scores on two subtests: letter-word identification (items vary from matching a picture and word to reading a word correctly) and passage comprehension (items vary from identifying a picture associated with a phrase to filling in a missing word in a passage). The broad math score is the average of the math calculation (a self-administered test ranging from addition to calculus problems) and the applied problems (practical problems that require analysis and simple calculations) subtest scores. We also report a simple average of the children's combined broad reading and math scores.

In analyzing the children's test scores, two issues came to our attention. The first is that different interviewers appear to be associated with systematically higher or lower test scores, even after controlling for child characteristics. Details of this analysis are given in the Appendix. In order to adjust for these "interviewer effects," we first estimate interviewer fixed effects using a linear regression model (with a separate model for each test score) that controls for our standard covariates and tract fixed effects.[10] We then adjust each score by removing the component of the score attributed to the interviewer effect. Results presented use the adjusted scores.

---

[10] We control for tract fixed effects because the interviews conducted by an individual are not randomly distributed with respect to location.

A second issue is that MTO children ages five through eight scored close to the national average on the WJ-R, considerably higher than one would expect given these children's demographic characteristics. Although the scores are high, we believe the scores do provide information about academic achievement. For example, individual covariates such as age, behavior problems, and participation in gifted classes are strongly predictive of scores. Thus, we believe the data are still appropriate for drawing comparisons based on the relative levels of scores of the control and treatment groups.

Performance on the WJ-R can be reported using several different metrics. We use the WJ-R's "W" scale as our underlying metric because these scores reflect an absolute measure of performance and have the attractive property of being equal-interval.[11] To facilitate interpretation of results, we transform the W scores to z-scores that have a mean of zero and standard deviation of one for the control group.

E.   Measures of Behavior Problems and Schooling

In addition to our primary test score outcomes, we examine the effect of neighborhoods on behavior problems and schooling. Interviewers asked the adults whether the children exhibited specific behavioral problems. These problems are a subset of those used for the National Longitudinal Study of Youth (NLSY). We define our measure of behavioral problems as the fraction of eleven problems that the adult reported as "sometimes" or "often" true for the child. The survey also gathered information from the adults on other schooling outcomes such as

---

[11] The W score is a scaled score based on a Rasch model estimated by McGrew, Werder, and Woodcock (1991). A W score of 500 represents individual achievement similar to the average person beginning the fifth grade. An item with 500-level difficulty is one that a fifth grader will answer correctly 50 percent of the time. Someone with a score of 520 will on average have gotten 500-level items correct 90 percent of the time, 510-level items correct 75 percent of the time, and 520-level items correct 50 percent of the time. The difference between 550 and 530 (20 points) and between 520 and 500 (20 points) indicates the same relative odds that the higher-scoring individual will answer an item correctly in comparison to the lower-scoring individual.

grade retention, school suspensions, and any special classes taken. To assess how engaged the

children were with learning, we asked them directly about how hard they work at school,

tardiness, hours spent reading, etc.

V. Econometric Models

A. *Estimation of the Effect of Being Offered a Housing Voucher*

We hypothesized that moves to lower poverty neighborhoods would lead to improved

educational outcomes for children. Our basic strategy for identifying the effects of

neighborhoods is to compare the educational outcomes of children whose families were offered

housing vouchers to those whose families were not offered vouchers. The random assignment of

families to voucher (treatment) and non-voucher (control) groups allows us to interpret

differences in outcomes as the effects of being offered the treatment, the "intent-to-treat" (ITT)

effects.

We estimate the ITT effects using a simple regression framework:

$$(1) \qquad Y = Z\pi_1 + X\beta_1 + \varepsilon_1,$$

in which Y is the outcome of interest, Z is an indicator for assignment to a treatment group, and X

is a series of baseline covariates. The coefficient π_1 on the indicator for treatment assignment

captures the ITT estimate for the outcome. In a randomized experiment, the unbiased estimation

of π_1 does not require the inclusion of covariates (X) in the model. However, we include

covariates in our model to gain additional precision and to control for any chance differences

between the groups. We use separate regressions to estimate the effects for the experimental and

Section 8 treatments. Sample weights allow us to account for the sampling of children from each

family, the subsampling of children for the second phase of interviewing, and the changes in the

ratios by which families were randomly assigned.[12] To account for correlations in the data between siblings, we cluster by family and report Huber-White standard errors.

The ITT estimates provide us with measures of the average impacts of being offered a voucher. Using these ITT estimates and information on compliance rates, one can estimate the magnitude of the impact on those who complied with the treatment (i.e., moved using a program voucher). Assuming that families in the treatment group were not affected if they did not use the voucher, the magnitude of the "treatment-on-treated" (TOT) effect is essentially the ITT divided by the fraction that complied with the treatment (Bloom, 1984). Thus, the 47 percent compliance rate for children in our experimental group implies that the TOT effects are approximately twice as large as our ITT estimates. To estimate TOT effects for specific outcomes, we adjust for covariates by using a two stage least squares regression with treatment status as the instrumental variable for treatment compliance.[13]

B. Estimation of Effects by Age

We hypothesized that the effect of neighborhoods on educational outcomes would be stronger for younger rather than older children. To examine effects for different age groups, we divided the child sample into three roughly-equal groups: ages six to ten, 11 to 14, and 15 to 20 (as of December 31, 2001). Using a regression model containing interactions between the treatment indicator and three age-group dummies (G_1, G_2, G_3), we estimate the effects for each age group:

---

[12] For the analyses, each child is weighted by the product of (a) the inverse probability of being selected from among the children in his family (if more than two) for the sample, (b) the inverse probability of being selected for the second phase of interviewing if part of the subsample, and (c) the inverse probability of his family being assigned its randomization status. The last component of the weight is necessary because the ratios by which families were randomly assigned to treatment and control status was altered at different sites in response to higher than anticipated lease-up rates. The revised ratios were designed to minimize the minimum detectable effects (MDEs). These weights were constructed to eliminate the association between treatment status and time or cohort; to preserve the overall proportions of control, Section 8, and experimental group families; and to preserve the proportions by site-time period. For additional information on construction of the weights, see Orr et al. (2003), Appendix B.
[13] For additional discussion of TOT estimation, see Katz, Kling and Liebman (2001).

(2) $Y = G_1Z\pi_{21} + G_2Z\pi_{22} + G_3Z\pi_{23} + X\beta_2 + \varepsilon_2.$

The coefficients π_{21}, π_{22}, and π_{23} on the interactions between the treatment indicator and age

groups capture the ITT effects on outcome Y for the three age groups. (Main effects of the age

groups are controlled for by X.)

To test the hypothesis of linear age effects, we use a regression model that includes the

treatment status indicator and an interaction of the treatment indicator with the child's age in

years (A):

(3) $Y = Z\pi_3 + AZ\gamma_3 + X\beta_3 + \varepsilon_3.$

The coefficient γ_3 on the interaction between treatment and age provides an estimate of the

treatment effect for each additional year of age and a test of the null hypothesis of no linear

interaction.

C. Exploratory Analyses of Effects by Subgroup

In addition to examining whether the effects of neighborhoods differed by age group, we

explore whether the effects differed by gender, race and ethnicity, and educational risk factors.

Analyses of crime and mental health outcomes for MTO youth suggest more beneficial effects

for girls (Kling, Liebman, and Katz, 2005; Kling, Ludwig, and Katz, 2005). Race and ethnicity

could potentially have reduced the impact of MTO by restricting the residential options of

treatment families due to racial segregation or affecting their choice of neighborhoods due to

same group or linguistic preferences. In addition, school peer effects could differ as well if these

effects are stronger within than across racial and ethnic groups (Hoxby, 2000). On the other

hand, if the residential choices of African-Americans were particularly conscribed prior to

enrollment, then one might expect the increased opportunities for mobility that MTO offered to

have produced stronger impacts for African-Americans. There is also some suggestion, albeit

contested, from the school voucher literature that vouchers may have more positive impacts for non-Hispanic African-American students than for white students (Peterson and Howell, 2004; Krueger and Zhu, 2004).

Effects could also differ by level of academic risk. Higher performing students might be able to make greater use of the educational resources of higher income neighborhoods. Similarly, students with fewer behavioral problems might find it easier to adapt to the norms of a new school. On the other hand, the negative effects of high-poverty neighborhoods could be particularly important for those students at greatest risk and the opportunity to move might have the greatest impact on these students. We examine results by the presence of behavioral, emotional, or learning problems at baseline as reported by the head of household and by baseline characteristics predictive of low, moderate, and high test scores.[14]

Regarding subgroups defined by city of original residence, we briefly summarize estimates of effects on academic achievement by site in our results section for comparability to other studies. Estimates for specific MTO sites are challenging to interpret because the convenience sample of sites is small and the sites differ on many dimensions, including their private housing markets, public housing stock, job markets, racial and ethnic composition, school systems, transportation systems, and crime levels. For these reasons, we focus our analysis primarily on results that pool data across sites. For those interested in the effects by site, detailed results are available on the web.

---

[14] To determine baseline characteristics predictive of test scores, we estimated predicted values of test scores using a model estimated from control group data with the covariates discussed in Table 1 and its notes. Then we created an indicator for scores in the lower, middle, and upper third of the distribution of predicted scores. To avoid overfitting with our moderate sample size and large number of covariates, we used a jackknife procedure in which separate models were estimated for each observation using all observations in the control group not including the observation for which the score was being predicted.

To determine the effects of the program for specific subgroups, we used the regression model shown in equation 1 and limited the estimation sample to the subgroup. As the subgroups were not pre-specified, these results should be viewed as exploratory.

VI. Effects on Mobility, Neighborhood, and School Characteristics

The MTO demonstration was designed to lead similar families to reside in different neighborhoods. Table 2 examines the impact of the experimental and Section 8 treatments on the neighborhood and school environments of children. The first row of Table 2 shows the estimates for the average poverty rate of the census tracts MTO children lived in between randomization and 2002: the left panel presents the control means, the middle panel presents the analytic Ns and ITT estimates for the experimental group, and the right panel shows the analytic Ns and ITT estimates for the Section 8 group. On average, controls lived in high poverty neighborhoods with poverty rates of 46 percent (column i). Column (vi) shows the estimated ITT effect of the experimental treatment. For all children combined, the offer of a restricted housing voucher led to living in neighborhoods with poverty rates averaging 12.6 percent lower than that of controls (33.0 percent poverty versus 45.6 percent). For those who used the restricted vouchers, this translates to a reduction in neighborhood poverty rate of about 25 percent relative to the rate if they had not used the vouchers (the TOT effect). Children in families that moved using an experimental voucher resided an average of 3.1 years (range of 1.3 to 4.6 years for the 25th to 75th percentile) at their new addresses. The differences in poverty rates are largest soon after random assignment and then decline over time.[15] Using a separate regression, we estimated the ITT effect of the Section 8 treatment on poverty rates (column xi). The effect was somewhat

---

[15] There are three main reasons for this. First, the census tracts to which MTO compliers initially moved had higher poverty rates on average in 2000 than in 1990. Second, subsequent moves by treatment compliers tended to be to tracts with higher poverty rates. Third, many control group families eventually left their original housing projects and these moves tended to be to census tracts with lower poverty.

smaller than the effect for the experimental group, reducing the average poverty rate by 10.6 percent. As the rest of Panel A shows, both treatment offers resulted in children living in neighborhoods with lower male unemployment, more college-educated adults, and fewer minorities. Effects on neighborhoods were generally stronger for the experimental treatment than for the Section 8 treatment.

Panel B focuses on residential mobility and school switching between randomization and follow-up. Children in both the experimental and Section 8 groups experienced more residential moves than controls, who themselves moved on average at least once. For the experimental group, the voucher offer also led to more moves of at least ten miles from the baseline address.[17] Treatment group children spent less time than controls attending schools in the five main urban districts associated with the MTO sites, however, there were no statistically significant differences between treatment and control groups in terms of the number of schools attended or mid-grade school changes.

Panel C presents the estimated effects of MTO on children's school peers and school climate. The treatment offers led to greater attendance of schools with higher average state exam scores and with smaller shares of minority and free lunch eligible students (i.e., from families with incomes below 130 percent of the poverty level). The experimental treatment offer did not, however, lead students to attend schools with lower pupil-teacher ratios. Furthermore, the average change in school performance on state exams was fairly modest, with the schools of the experimental group ranking on average at the 19th percentile compared to the 15th percentile for the control group. Further analyses, not shown in the table, indicate that the difference in school

---

[17] Results not shown in the table indicate that the fraction with at least one residential move since random assignment was 70 percent in the control group, 82 percent in the experimental group, and 86 percent in the Section 8 group.

scores is attributable to differences in student demographics.[18] As with neighborhoods,

estimated effects on school peers tended to be greater for the experimental group than for the

Section 8 group. Although the administrative data on schools indicates differences in school

peers between the treatment and control groups, the children's perceptions of their schools'

climates generally do not differ between the treatment and control groups.

To provide a better sense of what the average changes shown in Table 2 represent in

terms of the types of neighborhoods and schools of treatment compliers at follow-up, Table 3

shows the share of treatment compliers in neighborhoods and schools above or below different

threshold characteristics. For comparison purposes, the table also shows estimates of the control

complier means. Control compliers are those children in the control group whose families would

have complied with the treatment if offered it; the neighborhoods and schools of the control

compliers represent the counterfactual of what the treatment compliers would have experienced

in the absence of the treatment. Although we cannot directly identify the control compliers, we

can estimate the characteristics of control compliers under the assumption that the distribution

for noncompliers in the treatment and control groups is the same. The shares for the treatment

compliers are observed, and the difference in the shares for treatment and control compliers is

the TOT effect.

The first row of Table 3 shows that while less than 5% of control compliers are estimated

to be living in tracts with a poverty rate of less than 10%, over 25% of experimental treatment

---

[18] Our school test score data are aggregate data and thus it was not possible to assess the quality of the schools using
a true value-added analysis of individual student test score gains. In supplemental analyses, we did however rank the
schools based on the residual of their average scores after controlling for the percentage (using fourth order
Legendere polynomials) of students who were free lunch eligible, reduced lunch eligible, African-American,
Hispanic, and Asian. The average residual ranking of control schools was at the 50th percentile and the difference
between the experimental treatment group and controls was not statistically significant. Thus, the difference in test
score rankings observed between the experimental treatment and control group appears to be a reflection of
demographic differences in the students attending the schools. The lack of differences in scores after controlling for
student demographics is perhaps not surprising insofar as MTO was designed to move families to less poor
neighborhoods and thus the school characteristics impacted would be expected to be correlated with the
socioeconomic status of the neighborhood and the students.

compliers were living in these types of tracts. While roughly half of the control compliers were still living in high poverty neighborhoods (poverty rates of at least 40%), this was true for only than 10% of experimental treatment compliers (see fourth row). The distributions indicate that although the experimental treatment affected the likelihood of a child living in a neighborhood or attending a school that was above the 50th percentile in rank or had a majority of white non-Hispanics, the treatment induced only a small share of treatment complier children who would otherwise not have lived in these types of neighborhoods or attended these types of schools to do so: 15% of experimental group complier children for tracts above the 50th percentile rank and 8% of children for schools above the 50th percentile.

Figure 1 uses kernel density estimates to display the distribution of selected 2002 neighborhood and school characteristics for experimental compliers in comparison to control compliers.[19] As illustrated by the upper left-hand graph in Figure 1, the experimental treatment led to a distinct shift in terms of neighborhood poverty levels. The bottom two graphs show the distribution of state percentile ranks of neighborhood poverty (with higher ranks indicating less poverty) and school exam scores (with higher ranks indicating higher scores). The two graphs help to compare the changes MTO induced for neighborhoods versus schools. The experimental treatment led to a substantial shift in the distribution of neighborhoods in terms of poverty rank but a more modest change in the distribution of school ranks.

---

[19] The density distribution for the overall control group is the sum of the distributions for the control compliers and control noncompliers. Although we cannot directly observe the distributions for the control compliers and noncompliers, we can estimate the distribution for control compliers by subtracting the distribution for the treatment noncompliers from the overall control complier distribution under the assumption that the control noncomplier distribution and the treatment noncomplier distributions are the same. The overall distribution for the treatment group (f_1) can be decomposed into the fraction (π_1) due to the density for compliers (f_{11}) and the fraction ($1-\pi_1$) due to the density for noncompliers (f_{10}), where all three distributions are directly observable. For controls, only the overall distribution (f_0) is observed. Imbens and Rubin (1997) show that under the assumption that treatment noncompliers have the same distribution of poverty rates as the control noncompliers ($f_{10} = f_{00}$), one can subtract the treatment noncomplier density from the control overall density to obtain an estimate of the control complier density: $f_{01} = (f_0 - (1-\pi_1)f_{10})/\pi_1$.

In summary, the offer of a voucher led families to live in neighborhoods that were substantially less poor, had more educated residents and had somewhat fewer minority residents. The offer also led children to attend schools that performed somewhat better on state exams. However, the treatment did not generally lead families to move to white suburban neighborhoods or lead children to attend top performing schools.

VII. Effects on Educational Outcomes

A.   *Effects on Educational Outcomes, Overall and by Age Group*

We had hypothesized that moves to lower poverty neighborhoods would lead to improved educational outcomes for children. Panel A of Table 4 presents the estimated effects of the experimental and Section 8 treatments on test scores, our primary outcomes. The first row of the table presents the results for the combined reading and math scores. By construction, the mean of the normalized scores for the control group (column i) is zero, with a standard deviation of one. To examine scores by age, we divided the sample into three age groups of roughly equal size: ages 6 to 10, 11 to 14, and 15 to 20. As the control means for the specific age groups in columns ii through iv show, scores rise with age but do so more slowly for the oldest group. The control mean for the youngest children is -.857, or more than a standard deviation below the mean for the middle age group of .281. The coefficient of .018 in column v represents the ITT effect of experimental treatment on reading and math for all ages combined, and is less than two hundredths of a standard deviation. The standard error of .030 (in parentheses in column v) indicates that the ITT estimate is not statistically significant and that we had sufficient statistical power to detect a true effect as small as .084 standard deviations (or 2.8 times the standard error of the estimate) 80 percent of the time at the .05 level of significance (Bloom, 1995).

We had hypothesized that the effects of the intervention would be stronger for younger rather than older children. Using a separate regression, we estimate the ITT effects of experimental treatment for each age group (see columns vi through viii). The treatment effect on the combined reading and math scores is not statistically significant for any of the age groups nor is the coefficient on the linear age interaction (shown in column ix) statistically significant. The coefficient on the linear age interaction of .0016 implies that a ten-year age difference, such as the difference between the effect for 18-year-olds versus 8-year-olds, is associated with only a .016 increase (or less than two hundredths of a standard deviation) in the magnitude of the ITT effect. The last five columns of the table present parallel results for the Section 8 treatment and similarly show no evidence of effects on achievement scores or of an interaction between treatment effects and age.

Panel B of Table 4 shows estimates of the effects of the treatment on parental reports of behavior problems and problems at school, including grade repetition, suspensions, and requests from the school for someone to come in to talk about problems the child was having. Table 5 contains estimates of ITT effects for educational mediators such as school engagement (e.g., child self-report of paying attention in class, tardiness) and special classes (remedial or advanced). As with test scores, the estimated ITT effects and the coefficients on the age interactions are generally not statistically significant for either the experimental or Section 8 treatments. The point estimates of the treatment effects are also small. The one exception is that the experimental treatment appears to be associated with more behavior problems for youth ages 11 to 14 (Table 4). Findings are essentially unchanged regardless of whether or not they are adjusted for interviewer effects, and regardless of whether or not covariates are included in the estimation. While it is possible that the reference group against which parent-reported behavior problems are assessed is itself affected by the MTO intervention, we do not believe this is

biasing our results against finding treatment effects. Our reasoning is that any changes in the reference group were most salient initially, but studies two to three years after random assignment found some significant reductions in behavior problems for MTO children in the experimental group relative to the control group in Boston (Katz, Kling, and Liebman, 2001) and in New York (Leventhal and Brooks-Gunn, 2003) rather than the increases we had hypothesized would be associated with parent-reported ratings by experimental group parents using a new and better-behaved reference group.

B.   *Exploratory Analysis of Effects by Subgroups*

In addition to the main analyses, we also conducted exploratory analyses of educational effects by subgroup. Table 6 presents these results. We estimated the effects of the treatment on reading scores (columns i-iii), math scores (columns iv-vi), and the behavior problems index (columns vii-ix) by gender; by race and ethnicity; by baseline characteristics predictive of low, moderate, and high test scores; and by problems at baseline. For example, the first two rows show ITT estimates for girls and boys. The results are consistent in sign with the significant beneficial treatment effects for girls and adverse effects for boys found by Kling, Liebman and Katz (2005) in analysis of MTO youth ages 15-20 across the spectrum of education, risky behavior, mental health, and physical health – but the effects in Table 6 pooling all ages are not statistically significant for any of the contrasts in panel A except for the increase in behavior problems for experimental group boys relative to control group boys. Detailed estimates by gender and age are available on the web. These estimates do not show any statistically significant effects by gender for combined reading and math scores overall or for any age group. The effect on the combined broad reading and math scores for girls was positive (ITT = .051; SE = .038) while the estimate for boys was just below zero (ITT = -.008, SE = .042). The tables further

show that the experimental treatment appears to have had similar effects on the school and neighborhood characteristics of both girls and boys, however, older girls appear to have experienced more moves since randomization than older boys.

In fact, for achievement test scores, none of the ITT estimates for the subgroups shown in Table 6 are significant except for non-Hispanic African-American children for whom the experimental treatment appears to be associated with better reading skills with average scores eight tenths of a standard deviation higher than for the control group. Although the ITT effect of the Section 8 treatment for African-American children is not statistically significant, it is similar in magnitude, at seven tenths of a standard deviation, to the ITT estimate for the experimental treatment. In analyses not shown in the table, we find that the reading effects for African-Americans are concentrated in the Baltimore and Chicago sites, and our sample includes virtually no other racial or ethnic groups at these sites. In the other three sites (with multi-ethnic samples), we do not find differences in effects by race or ethnicity. Because of the exploratory process of this analysis and the large number of subgroups examined, the chances of having at least one subgroup ITT estimate result in a t-statistic greater than 2 is quite high even if the null hypothesis of no effect were true. Given the lack of evidence for an effect on African-Americans in the three sites with both African-Americans and Hispanics and the probability of finding at least one large t-statistic when examining a large number of subgroups, we caution that the statistically significant finding for African-Americans may be due to the sampling variability.

Higher values on the behavior problems index indicate more behavior problems as reported by the adult. For the experimental group, most of the ITT estimates on the index are not statistically significant. However, for boys and for children with baseline characteristics predictive of higher test scores, the experimental treatment appears to be associated with more

behavior problems. For the Section 8 group, none of the estimated effects are statistically distinguishable from zero.

In analyses shown in the web appendix, we estimated the effects on combined reading and math scores by site. These results reveal a positive effect for the Baltimore experimental group (ITT = .129, SE = .058) and a large negative effect on scores for the New York Section 8 group (ITT = -.214, SE = .068). For the eight other site-treatment groups, the estimated ITT effects were not statistically significant and ranged in magnitude from -.034 to .020 for the experimental group and from -.013 to .068 for the Section 8 group. We also examined the relationship between effects on scores and effects on potential mediators such as neighborhood and school characteristics across the sites, with detailed results given in the web appendix. There was no clear pattern. Although the Baltimore experimental treatment was associated with more positive estimated effects on mediators than the New York Section 8 group, other site-treatments with estimated neighborhood and school impacts comparable to Baltimore (such as LA and Boston experimental) show no impacts on combined reading and math scores. For example, the estimated effects on school ranks for the Baltimore, Boston and LA experimental groups are .063, .041 and .062, respectively, and on neighborhood poverty are -.113, -.093, and -.153. None of the experimental groups were associated with effects on special education or help, however, the Section 8 groups in Boston and LA were associated with less special education/help and in New York with more.

VIII. Conclusion

The Moving to Opportunity experiment provides a test of the consequences for poor children when their families are offered the chance to move from high-poverty neighborhoods to more affluent ones. Roughly half of the families offered this chance took it up, producing

sizeable differences in neighborhood conditions of experimental and control families during the four- to seven-year period after the program began.

We find that families offered housing vouchers in the MTO demonstration moved on average to residential neighborhoods that were substantially less impoverished, and sent their children to schools that were of modestly-higher quality. We did not find evidence of improvements in reading scores, math scores, behavior or school problems, or school engagement, overall or for any age group. Developmental theory suggests more rapid cognitive development among younger children and greater ability to adapt to new social environments. This theory suggests that environments (be they positive or negative) may have a greater impact on younger than older children. The lack of impacts was particularly surprising in the case of young children in the MTO sample, most of whom were preschoolers at the time that their families moved.

The overall results are in contrast to early MTO results for Baltimore. Ludwig et al. (2001) found that results on test scores drawn two to three years after baseline from the Baltimore MTO site suggested that achievement gains might indeed be larger for younger children than older children. Our longer-run data continue to show statistically significant test score impacts in the Baltimore site for elementary school-aged children. But the children for whom Ludwig et al. found positive impacts, ages 5-11 at random assignment, did not sustain their gains in our follow-up data. Rather, the gains, particularly for reading, were again concentrated among the elementary school-aged children at the time of the four- to seven-year follow-up – the younger siblings of the children in the Ludwig et al. study. It will be several more years before these children, whose formal education has entirely taken place after enrollment in MTO, can be assessed in adolescence to examine the persistence of effects. This site-specific effect may be the product of the Baltimore-area schools or some other structural

process, or may have resulted by chance. In ad hoc analyses we examined the relationship of estimated effects for test scores and mediators across site and treatment groups but we found no clear pattern of effects.

Why were there virtually no favorable treatment effects on children's test scores, school success or engagement, or problem behaviors, even among children who were preschoolers at the time of their moves -- despite improvements in neighborhood poverty rates? We focus our discussion on children in families randomly assigned to the experimental group, for whom improvements in neighborhood conditions and child test scores were expected to be greatest.

A first possible explanation of the null achievement results is that experimental group families undertook steps that undid some of the possible advantages of their placements in middle-class neighborhoods. Subsequent moves led them to reside in neighborhoods four to seven years after baseline that were considerably less affluent than their original placement neighborhoods. At the same time, control families, all of whom had indicated their eagerness to move by signing up for the chance to win the mobility lottery, made residential moves outside of the MTO demonstration that improved their neighborhood conditions. At follow-up, control children who had moved were living in neighborhoods that on average had poverty rates 17.6 percentage points lower than their original neighborhoods.

Despite these dilutions of the neighborhood "treatment," there were still huge differences in average neighborhood conditions between the baseline and follow-up surveys – neighborhood poverty rates for experimental movers were estimated to average about 20 percent in 2002 -- about half the average rate of control families. In the case of children in families assigned to the Section 8 treatment group, the comparable poverty rate improvement amounted to 11 percentage points.

By any accounting, the neighborhood "treatment" produced by the MTO offer was substantial, and almost certainly more substantial than any improvements that might be produced by in-place community improvement programs. At the same time, the neighborhood improvements experienced by the experimental group did not involve moving to truly affluent neighborhoods. Previous nonexperimental studies have suggested that neighborhood effects on academic achievement are largely found in comparisons of children in affluent areas to those in middle-income neighborhoods.

A related possible explanation for the null results is that although experimental families moved to less impoverished areas, most did not move to racially or ethnically-integrated neighborhoods. Three-fifths of experimental families that moved under the terms of the program still lived in neighborhoods with 80 percent or more minority populations. Discrimination may limit the availability of high-quality schools or other public services in minority neighborhoods, although we had no way of testing whether this was the case. It appears that it may take a program like Gautreaux (Rubinowitz and Rosenbaum, 2000), which defined target neighborhoods in terms of both poverty and race, to induce permanent moves to neighborhoods that are both more affluent and more integrated. Since Gautreaux lacked a compelling control group, we cannot say whether or not its "treatment" produced larger academic gains for children in participating families. The lack of integration may have been slightly more pronounced for MTO children's new schools relative to their new neighborhoods due to the number of white non-Hispanic children in some urban areas who attend private schools.

It is also important to note that MTO occurred in a different school context than Gautreaux. School finance equalization has reduced the resource disparities between school districts in poor and non-poor neighborhoods (Card and Payne, 2002) and today school choice is built into many urban school systems, offering more educational options to low-income families.

Over 30% of the control group children in Chicago and Los Angeles were attending magnet schools. In addition, school choice may have allowed MTO movers to continue to send their children to schools in their old neighborhoods. Evidence from a qualitative study of MTO families (Popkin et al., 2001) suggests that some parents took advantage of the option of sending their children to schools close to their original neighborhoods, near the homes of relatives. Perhaps they did so because they were more comfortable with their children's original schools or they preferred schools in neighborhoods with which they were familiar.

We suspect that one of the most important factors behind the stagnant test scores is that the improvements in neighborhood conditions enjoyed by experimental group families were accompanied by much more modest improvements in the schools attended by their children -- as expressed, for example, in terms of the ranks within the state of tract poverty rates and school test scores. Participation in MTO did not result in attendance at high-performing schools. Even among children in experimental group families who moved in conjunction with the program, only 14 percent were in schools ranked above the state median in test scores. Furthermore, the differences in school test score performance appear to reflect differences in peer characteristics such as the proportion of free lunch eligible students and may not be indicative of higher quality schools. We found no differences in the pupil-teacher ratios for the experimental and control group and when youth themselves were asked to characterize the climate and resources of their schools, virtually no experimental-control group differences were found. The lack of positive effects on self-reported measures such as school climate could reflect a lack of improvement in the educational climate of the schools attended by the treatment group but could also potentially be due to changes in frame of reference, particularly for the youngest children who may never have attended schools in the original neighborhoods.

The lack of effects on achievement test scores despite modest gains in the peer characteristics of the experimental group's schools is consistent with the results from Cullen, Jacob, and Levitt's (2003) study of school choice in Chicago which suggests that attending a higher performing high school may not, in the short term, produce gains in reading or math. On the other hand, Gould, Lavy, and Paserman (2004) estimated the effects on Ethiopian immigrants in Israel of being assigned (quasi-randomly) to a school with higher or lower average test scores. They found that assignment to schools with higher average test scores was associated with lower high school drop out rates and higher passing rates on matriculation exams. Using per capita income as an alternative measure of school environment they found similar results.

Even if there had been no improvements in school quality rather than the modest improvements we observe, the MTO treatment appears to have affected the educational environment of children in other ways. Some studies of elementary school students suggest that student achievement may be affected by the achievement of peers. Hoxby (2000), using variation in class composition as an instrumental variable, identified peer effects for Texas elementary school students, effects that were stronger within than across racial groups. Thus peer effects are one mechanism through which MTO could have lead to higher achievement scores. MTO could also have affected achievement levels through community and family mechanisms. It is clear that MTO moved families into more advantaged communities. These communities may have offered children more educated adult role models, higher educational norms, and greater community resources. As noted previously, although MTO's effects on neighborhoods were quite substantial, MTO did not move children into affluent communities.

Another likely factor behind the lack of achievement impacts is that while MTO-related neighborhood advantages appeared to improve the mental health of mothers, they failed to translate into other kinds of family advantages that might have promoted children's well-being.

31

Adults moving in conjunction with the MTO program were similar to their control-group counterparts in their employment status, welfare receipt, family income, parenting practices, and connections to their children's schools and to the parents of their children's friends (Kling, Liebman, and Katz 2005) .

Additionally, potential gains associated with neighborhood improvements may have been offset by the disruption of relocating and changing schools or districts. A number of studies have found an association between moves and poor academic performance and high school drop out (Ingersoll, Scamman, and Eckerling, 1989; Haveman, Wolfe and Spaulding, 1991). Although this association may in large part be due to differences between the students who move and those who do not, the moves themselves are hypothesized to negatively affect school performance, perhaps by disrupting social relationships (Pribesh and Downey, 1999). In theory, midyear school changes may be especially disruptive, however, Hanushek, Kain and Rivkin's (2001) study of school switching in Texas did not detect a difference between school year and summer switches. The MTO experimental treatment group experienced more moves than controls but did not significantly differ on our measures of number of schools attended or mid-grade school changes.

It appears that interventions focused exclusively on neighborhoods rather than on factors directly related to the child, family, and school are unable to solve the myriad problems of children growing up in poverty. From a policy perspective, residential mobility programs such as Section 8 (now called Housing Choice Vouchers) and even the more dramatic MTO experimental treatment do not appear to have large impacts on the academic problems of children who live in public housing in high-poverty neighborhoods. At the same time, the evaluation of MTO suggests other important areas in which residential mobility programs could have important impacts. The MTO experimental voucher offer was associated with positive

health outcomes for female adults and youth (Kling, Liebman, and Katz, 2005) and fewer criminal arrests among female youth; perhaps most important to the participating families, families offered these vouchers achieved their primary initial objective of dramatically reduced exposure to drugs and violence (Kling, Ludwig, and Katz, 2005).

Beyond these policy implications, what lessons does the MTO experiment hold with respect to the apparent consensus in the literature of neighborhood research that neighborhood conditions indeed matter for children's achievement (Brooks-Gunn, Duncan, and Aber, 1997)? A first noteworthy point is that MTO is highly selective, and includes children that represent only a subsample of those used in most neighborhood effects studies. Very few of the children in national studies such as the National Longitudinal Survey of Youth or the Panel Study of Income Dynamics live in conditions as bad as those experienced by the children in MTO families, and neighborhood effects estimated for a representative population may differ markedly from effects estimated for children in poor urban neighborhoods. But sample differences are unlikely to resolve the discrepancy, since most theories of contextual effects on children's development lead to the expectation that improvement in neighborhood conditions should matter *the most* for children living in the worst neighborhoods (Jencks and Mayer, 1990).

There are two more likely explanations. First, previous literature using nonexperimental data found effects of residence in affluent neighborhoods in comparison to middle income neighborhoods. MTO did not result in many families moving to affluent neighborhoods. Second, the neighborhood effects estimated in nonexperimental studies may be overstated. Difficult-to-measure differences in families living in good and bad neighborhoods could produce the spurious appearance of neighborhood effects (Manski, 1993; Moffitt, 2001). Because it caused large changes in neighborhood poverty but few changes in either school quality or family conditions, and because its treatment was randomly assigned, MTO provides vital data on the

likely achievement impacts of moving families from poor to less poor neighborhoods. These

impacts, if they exist, are small.

Appendix. Achievement Test Scoring

The test scores used in this paper have been adjusted for potential interviewer effects. The Woodcock Johnson-Revised tests used in the MTO study indicate the level of achievement within a very wide range, as opposed to many tests given in schools, which test proficiency at a particular threshold appropriate for specific grade levels. These same tests have also been used in other large social science studies, such as the Panel Study of Income Dynamics Child Development Supplement, the Los Angeles Family and Neighborhoods Survey, and the Welfare, Children and Families Three City Study. In order to adapt to the achievement level of each individual and to avoid confounding reading skills with other skills, the tests involve considerable interaction between the sample children and the interviewer conducting the test.

There are two subtests on which the Broad Reading (BR) score is based, Letter-Word Identification (LW) and Passage Comprehension (PC), and two on which the Broad Math (BM) score is based, Applied Problems (AP) and Calculation (CA). If the child did not begin with six consecutive correct answers, easier items were then asked to establish a "basal" level of performance. The test score is based on the number of correct answers, imputing correct answers for all items below the basal level. The items increase in difficulty until a child gives six consecutive incorrect responses, which establishes a "ceiling" of performance, at which point the test is stopped. Thus, while administering the test, the interviewer must score the items.

There are two types of interviewer effects that we suspect are most likely. First, interviewers read items aloud during the test (LW: 1-5 but not 6-57; PC: items 1-30 but not 31-43; AP: all items 1-60; CA: none of 1-58), and the reading and pronunciation skills of the interviewers varied. Second, there is some interviewer judgment required in scoring. For example, many LW items ask for pronunciation of words, such as "sufficient," but correct pronunciation is subject to interpretation. One PC item is: "A good composition has an interesting introduction and a strong conclusion. The body is ____ the beginning and the end." Correct answers are "between" or "in between" and examples of incorrect answers are "interesting," "supporting," and "both." While the interviewers were instructed in training to only score the item as correct if the child said "between" or "in between," it may have been the case that some interviewers were more inclusive and marked items as correct if the response seemed correct to them. The CA test consists of math problems in a workbook and involves little interviewer interaction or judgment.

We have several pieces of statistical evidence suggesting that some interviewers may systematically score respondents higher and some interviewers may score respondents lower on the reading and math tests. There is wide dispersion in the mean scores by interviewer, though interpretation of this statistic is confounded with the systematic assignment of interviewers to neighborhoods that they were most familiar with and where they could best locate sample members. Our main analysis of interviewer effects relies on differences between the test scores of different interviewers who tested sample members in the same census tract. In order to examine a sufficient number of tests per interviewer, we pool data for all 5223 children ages six through 20 who were tested in the MTO study. Specifically, we analyze the regressions of test scores on interviewer indicator variables conditional on census tract fixed effects and on individual characteristics. Interviewer coefficients are estimated relative to the interviewer conducting the largest number of tests in each of the five main MTO sites. Essentially, we assume that interviewers are as good as randomly assigned to children within census tracts, even though they are systematically assigned to children across tracts.

Simple summary measures of potential interviewer effects are the F-statistics on the 126 interviewer coefficients, which are: BR=3.0; BM=3.2; LW=2.5; PC=4.7; AP=4.1; CA=2.2. All p-values are less than .0001. This pattern is consistent with a hypothesis that the CA test (requiring the least interviewer reading or judgment) would have smaller interview effects – although its p-value is also less than .0001. Bonferroni-adjusted p-values for the most extreme interviewer, who reported systematically high scores, are .04 on BM, and less than .01 on PC and AP.

We conducted a similar set of analyses on the PC-LW and AP-CA differences, testing the hypothesis that interviewers should not have a systematic association with differences on two tests of reading or two tests of math for the same individual. The F-statistics on the 126 interviewer coefficients are: PC-LW=3.7, AP-CA=3.1. An extreme interviewer (who reported the largest differences) had Bonferroni-adjusted p-values of .0001 on PC-LW and .01 on AP-CA.

Based on this evidence, we conclude there is a reasonable possibility that scores may have differed systematically by interviewer. Although all interviewers conducted interviews with sample members of all three treatment groups, the proportions differed. The geographical mix also differed, with some interviewers mainly interviewing controls and experimental and Section 8 group non-compliers still living in inner-city areas.

In order to investigate the sensitivity of the treatment results to potential test score effects, we computed an adjusted test score. Using the logic described above, we estimated the interviewer coefficient conditional on census tract fixed effects and individual characteristics, and calculated the estimated interviewer effect as the deviation of the interviewer from the site mean. The adjusted score is simply the unadjusted W-score minus the estimated interviewer effect.

References

Aaronson, Daniel. 1998. "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes." *Journal of Human Resources* 33: 915-946.

Ainsworth, James W. 2002. "Why Does it Take a Village? The Mediation of Neighborhood Effects on Educational Achievement." *Social Forces* 81(1): 117-152.

Benabou, Roland. 1993. "Workings of a City: Location, Education, and Production." *Quarterly Journal of Economics* 108(3): 619-652.

Bloom, Howard S. 1984. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review* 8(April): 225-246.

Bloom, Howard S. 1995. "Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs." *Evaluation Review* 19(5): 547-556.

Bouchard, Thomas J. Jr. 1997. "IQ Similarity in Twins Reared Apart: Findings and Responses to Critics." In *Intelligence, Heredity, and Environment*, eds. Robert J. Sternberg and Elena Grigorenko, Cambridge: Cambridge University Press, 126-162.

Brock, William A. and Steven N. Durlauf. 2001. "Interactions-based Models." In *Handbook of Econometrics, Volume 5*, eds. James J. Heckman and Edward E. Leamer, Amsterdam: North-Holland, 3297-3380.

Brooks-Gunn, Jeanne, Greg J. Duncan, and J. Lawrence Aber, eds. 1997. *Neighborhood Poverty: Contexts and Consequences for Children (Volume 1)*. New York: Russell Sage Foundation.

Card, David and A. Abigail Payne. 2002. "School Finance Reform, the Distribution of School Spending, and the Distribution of SAT Scores," *Journal of Public Economics* 83(1): 49-82.

Chase-Lansdale, Lindsay P. and Rachel A. Gordon. 1996. "Economic Hardship and the Development of Five- and Six-Year-Olds: Neighborhood and Regional Perspectives." *Child Development* 67: 3338-3367.

Chase-Lansdale, Lindsay P., Rachel A. Gordon, Jeanne Brooks-Gunn, and Pamela K. Klebanov. 1997. "Neighborhood and Family Influences on the Intellectual and Behavioral Competence of Preschool and Early School-Age Children." In *Neighborhood poverty: Context and Consequences for Children (Volume 1)*, eds., Jeanne Brooks-Gunn, Greg J. Duncan, and J. Lawrence Aber, New York: Russell Sage Foundation Press, 79-118.

Collins, Rebecca L. 1996. "For Better or Worse: The Impact of Upward Social Comparison on Self-Evaluations." *Psychological Bulletin* 119: 51-69.

Cullen, Julie B., Jacob, Brian A., and Steven Levitt. 2003. "The Effect of School Choice on Student Outcomes: Evidence from Randomized Lotteries." National Bureau of Economic Research Working Paper 10113, November.

Currie, Janet, and Aaron Yelowitz. 2000. "Are Public Housing Projects Good for Kids?" *Journal of Public Economics* 75: 99-124.

Duncan, Greg J., Johanne Boisjoly, and Kathleen Mullan Harris. 2001. **"**Sibling, Peer, Neighbor and Schoolmate Correlations as Indicators of the Importance of Context for Adolescent Development." *Demography*, 38(3): 437-447.

Duncan, Greg J., Jeanne Brooks-Gunn, and Pamela Klebanov. 1994. "Economic Deprivation and Early-Childhood Development." *Child Development* 65(2): 296-318.

Duncan, Greg J. and Stephen W. Raudenbush. 2001. "Neighborhoods and Adolescent Development: How Can We Determine the Links?" In *Does it Take a Village? Community Effects on Children, Adolescents, and Families*, eds., Alan Booth and Nan Crouter, State College, PA: Pennsylvania State University Press, 105-136.

Duncan, Greg J., W. Jean Yeung, Jeanne Brooks-Gunn, and Judith R. Smith. 1998. How much does childhood poverty affect the life chances of children? *American Sociological Review* 63: 406-423.

Ellen, Ingrid and Margery A. Turner. 1997. "Does Neighborhood Matter? Assessing Recent Evidence." *Housing Policy Debate* 8(4): 833-866.

Feins, Judith D., Mary Joel Holin, and Antony A. Phipps. 1994. *Moving to Opportunity for Fair Housing Demonstration Program Operations Manual*. Report prepared for the U.S. Department of Housing and Urban Development by Abt Associates Inc., Cambridge, MA.

Fernandez, Raquel. 2002. "Sorting, Education, and Inequality." In *Advances in Economics and Econometrics: Theory and Applications: Eighth World Congress,* eds., Mathias Dewatriport, Lars P. Hansen, and Steven J. Turnovsky, Cambridge: Cambridge University Press.

Furstenberg, Frank F., Jr., Thomas D. Cook, Jacquelynne Eccles, Glen H. Elder, Jr., and Arnold J. Sameroff, eds. 1999. *Managing to Make it: Urban Families and Adolescent Success*. Chicago: University of Chicago Press.

Goering, John, Joan Kraft, Judith Feins, Debra McInnis, Mary Joel Holin, and Huda Elhassan. 1999. *Moving to Opportunity for Fair Housing Demonstration Program: Current Status and Initial Findings*. Report prepared for the U.S. Department of Housing and Urban Development by Abt Associates Inc., Cambridge, MA.

Gould, Eric D., Lavy, Victor and M. Daniele Paserman (2004). Immigrating to Opportunity: Estimating The Effect of School Quality Using a Natural Experiment On Ethiopians in Israel. *Quarterly Journal of Economics,* 119(2): 489-526.

Halpern-Felsher, Bonnie L., James P. Connell, Margaret Beale Spencer, J. Lawrence Aber, Greg J. Duncan, Elizabeth Clifford, Warren E. Crichlow, Peter A. Usinger, Steven P. Cole, LaRue Allen, and Edward Seidman. 1997. "Neighborhood and Family Factors Predicting Educational Risk and Attainment in African-American and European-American Children and Adolescents." In *Neighborhood Poverty: Context and Consequences for Children (Volume 1)*, eds. Jeanne Brooks-Gunn, Greg J. Duncan, and J. Lawrence Aber, New York: Russell Sage Foundation Press, 146-173.

Haveman, Robert, Wolfe, Barbara, and James Spaulding. 1991. "Childhood Events and Circumstances Influencing High School Completion." *Demography*, 28, 133-157.

Heckman, James J. 2000. "Policies to Foster Human Capital." *Research in Economics*, 54(1): 3-56.

Hofferth, S., Davis-Kean, P.E., Davis, J., Finkelstein, J. 1998. *The Child Development Supplement of the Panel Study of Income Dynamics, 1997 User Guide*. Survey Research Center Institute for Social Research, The University of Michigan. Ann Arbor, MI.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation," National Bureau of Economic Research Working Paper 7867, August.

Imbens, Guido W. and Rubin, Donald B. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies,* 1997, 64(221), pp. 555-575.

Jacob, Brian A. 2004. "Public Housing, Housing Voucher, and Student Achievement: Evidence from Public Housing Demolitions in Chicago." *American Economic Review* 94(1): 233-258.

Jencks, Christopher and Susan E. Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." In *Inner-City Poverty in the United States*, eds., Laurence E. Lynn, Jr. and Michael G.H. McGeary, Washington, DC: National Academy Press, 111-186.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics*, 116(2): 607-654.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2005. "Experimental Analysis of Neighborhood Effects." NBER Working Paper 11577, August.

Kling, Jeffrey R., Jens Ludwig, and Lawrence F. Katz. 2005. Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment. *Quarterly Journal of Economics* 120(1): 87-130.

Kohen, Dafna E., Jeanne Brooks-Gunn, Tama Leventhal, and Clyde Hertzman. 2002. Neighborhood income and physical and social disorder in Canada: Associations with young children's competencies. *Child Development* 73: 1845-1860.

Krueger, Alan B. and Pei Zhu. 2004. Another Look at the New York City School Voucher Experiment. *American Behavioral Scientist*. 47(5): 658-698.

Leventhal, Tama and Jeanne Brooks-Gunn. 2000. "The Neighborhoods They Live In: The Effects of Neighborhood Residence on Child and Adolescent Outcomes." *Psychological Bulletin* 126(2): 309-337.

Leventhal, Tama and Jeanne Brooks-Gunn. 2003. "New York City Site Findings: The Early Impacts of Moving to Opportunity on Children and Youth," in *Choosing a Better Life: Evaluating the Moving to Opportunity Social Experiment*, eds., John Goering and Judith Feins, Washington, DC: Urban Institute Press, 213-244.

Leventhal, Tama and Jeanne Brooks-Gunn. 2004. "A Randomized Study of Neighborhood Effects on Low-Income Children's Educational Outcomes." *Developmental Psychology* 40(4): 488-507.

Lippman, Laura, Shelly Burns, and Edith McArthur. 1996. *Urban Schools: The Challenge of Poverty and Location (NCES 96-184).* Washington, DC: National Center for Education Statistics.

Ludwig, Jens, Helen F. Ladd, and Greg J. Duncan. 2001. "Urban Poverty and Educational Outcomes." *Brookings-Wharton Papers on Urban Affairs,* 147-201.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60: 531-542.

Marsh, Herbert W. and John W. Parker. 1984. "Determinants of Student Self-Concept: is it Better to be a Relatively Large Fish in a Small Pond Even if You Don't Learn to Swim as Well?" *Journal of Personality and Social Psychology* 47: 213-231.

Mayer, Susan E. and Christopher Jencks. 1989. "Growing Up in Poor Neighborhoods: How Much Does it Matter?" *Science* 243: 1441-1445.

McGrew, Kevin S., Judy K. Werder, and Richard A. Woodcock. 1991. *WJ-R Technical Manual.* Allen, TX: DLM.

Moffitt, Robert A. 2001. "Policy Interventions, Low-Level Equilibria, and Social Interactions." In *Social Dynamics*, eds., Steven N. Durlauf and H. Peyton Young, Cambridge: MIT Press, 45-82.

Orr, Larry, Judith D. Feins, Robin Jacob, Eric Beecroft, Lisa Sanbonmatsu, Lawrence F. Katz, Jeffrey B. Liebman, and Jeffrey R. Kling. 2003. *Moving to Opportunity: Interim Impacts Evaluation.* Washington D.C.: U.S. Department of Housing and Urban Development.

Peterson, Paul E. and William G. Howell. 2004. Efficiency, Bias, and Classification Schemes. A Response to Alan B. Krueger and Pei Zhu. *American Behavioral Scientist*. 47(5): 699-717.

Popkin, Susan J., Laura E. Harris, and Mary K. Cunningham. 2001. *Families in Transition: A Qualitative Analysis of the MTO Experience.* Washington, D.C.: The Urban Institute.

Rosenbaum, James E. 1995. "Changing Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program." *Housing Policy Debate* 6(1): 231-269.

Rouse, Cecilia, Jeanne Brooks-Gunn, and Sara McLanahan. 2005. "Introducing the Issue." *The Future of Children.* School Readiness: Closing Racial and Ethnic Gaps. 15(1): 5-13.

Rowe, David C. 1994. *The Limits of Family Influence: Genes, Expression, and Behavior*. New York: Guilford Press.

Rubinowitz, Leonard S. and James E. Rosenbaum. 2000. *Crossing the Class and Color Lines: From Public Housing to White Suburbia*. Chicago: The University of Chicago Press.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multi-level Study of Collective Efficacy." *Science* 277: 918-924.

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing Neighborhood Effects: Social Processes and New Directions in Research." *Annual Review of Sociology* 28: 443-78.

Shonkoff, Jack P. and Deborah Phillips, eds. 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.

Slavin, Robert E. , Nancy L. Karweit and Barbara A. Wasik. 1993. Preventing Early School Failure: What Works? *Educational Leadership*, 50(4): 10-18.

Solon, Gary, Marianne E. Page, and Greg J. Duncan. 2000. "Correlations Between Neighboring Children in Their Subsequent Educational Attainment." *Review of Economics and Statistics* 82: 383-392.

U. S. Department of Education. 2003a. *The Nation's Report Card: Trial Urban District Assessment Mathematics Highlights 2003 (NCES 2004-458).* Washington, DC: National Center for Education Statistics.

U. S. Department of Education. 2003b. *The Nation's Report Card: Trial Urban District Assessment Reading Highlights 2003 (NCES 2004-459).* Washington, DC: National Center for Education Statistics.

Votruba, Mark E., and Jeffrey R. Kling. 2004. "Effects of Neighborhood Characteristics on the Mortality of Black Male Youth: Evidence from Gautreaux." Princeton IRS Working Paper 491, July.

Wood, Joanne V. 1989. "Theory and Research Concerning Social Comparisons and Personal Attributes." *Psychological Bulletin* 106: 231-248.

Woodcock, Richard W. and M. Bonner Johnson. 1989, 1990. *Woodcock-Johnson Psycho-Educational Battery - Revised*. Itasca, IL: Riverside.

Woodcock, Richard W. and Nancy Mather. 1989, 1990. WJ-R Tests of Achievement: Examiner's Manual. In R.W. Woodcock and M.B. Johnson's *Woodcock-Johnson Psycho-Educational Battery - Revised*. Itasca, IL: Riverside.
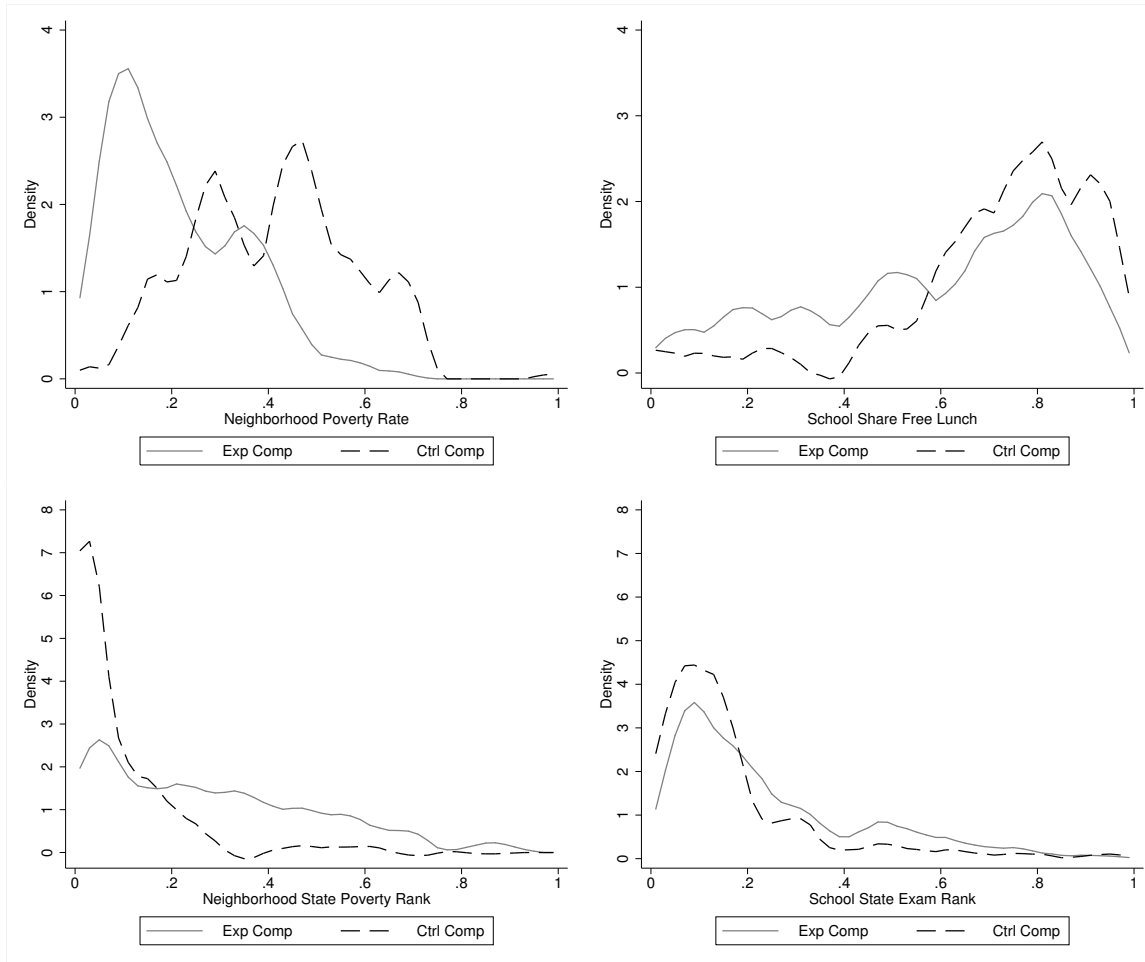
Notes: Exp Comp = experimental compliers. Ctrl Comp = control compliers. Sample is restricted to children ages 6 to 16 for whom WJ-R test score data were available. Kernel densities estimates are based on an Epanechnikov kernel and halfwidth of .030. For the experimental group, we directly observe the distributions for the overall group, for treatment compliers, and for treatment noncompliers. For the control group, we do not observe who would have complied with the treatment if offered it. However, under the assumption that control noncompliers have the same distribution of characteristics as the experimental noncompliers, we estimate the control complier distribution by subtracting the experimental noncomplier density from the overall control group density (Imbens and Rubin, 1997). Neighborhood is defined as the Census tract in which the child lived in 2002 and school as the school attended in 2002. Tract characteristics are from the 2000 Census and school characteristics from the NCES's Common Core of Data, the Private School Survey, and the National Longitudinal School-Level Assessment Score Database. Information on free lunch program participation was generally not available for Illinois. For comparability, the Kernel density estimates for neighborhood poverty rate and share free lunch were restricted to children with valid data on both measures. School exam scores were generally not available for older children in Baltimore and New York. For comparability, density estimates of state poverty rank and state exam rank exclude children ages 14 and older for Baltimore and New York and only include children with valid data on both measures. Higher rankings represent neighborhoods with lower poverty and schools with higher test scores.

TABLE 1 — SELECTED BASELINE CHARACTERISTICS

Variable	Control Mean (i)	Experimental Mean (ii)	CP Mean (iii)	NCP Mean (iv)	CP-NCP (v)	Section 8 Mean (vi)	CP Mean (vii)	NCP Mean (viii)	CP-NCP (ix)
A. Child demographics									
Age in years (as of 12/31/01)	13.1	13.3	13.1	13.5	-.3	13.3	13.0	13.8	-.9*
Male	.52	.49	.47	.50	-.03	.50	.49	.52	-.03
Hispanic ethnicity	.31	.29	.27	.30	-.03	.29	.25	.35	-.10*
Non-Hisp. African-American	.63	.66	.66	.66	.00	.65	.70	.57	.13*
Non-Hisp. other race	.03	.04	.05	.03	.02	.04	.03	.05	-.02
Baltimore site	.13	.14	.16	.13	.03	.15	.19	.09	.10*
Boston site	.20	.18	.16	.20	-.04	.20	.17	.25	-.08*
Chicago site	.23	.26	.18	.33	-.15*	.26	.29	.21	.08
Los Angeles site	.19	.17	.25	.11	.14*	.15+	.18	.10	.09*
New York site	.24	.24	.25	.23	.02	.23	.16	.35	-.19*
B. Child health problems									
Weighed <6 lbs at birth†	.16	.15	.12	.18	-.06	.17	.16	.19	-.03
Hospitalized prior to age one†	.22	.19	.19	.19	-.00	.15+	.14	.17	-.03
Problems with school/play	.05	.07+	.06	.09	-.03	.06	.06	.07	-.01
Problems req. medicine/equip	.08	.09	.09	.10	-.01	.10	.09	.12	-.03
C. Child education									
Read to more than once/day†	.32	.26	.28	.24	.04	.25	.21	.32	-.11
Attended gifted classes or did advanced work‡	.17	.15	.16	.14	.01	.16	.16	.17	-.01
School asked someone to come in about problems‡	.25	.26	.25	.26	-.01	.26	.25	.27	-.03
Behavior/emotional probs‡	.06	.09	.08	.11	-.03	.10+	.10	.11	-.01
Learning problems‡	.16	.17	.16	.18	-.02	.17	.13	.21	-.07*
Expelled in past two years‡	.08	.11	.10	.12	-.01	.11	.09	.13	-.04
D. Adult & HH characteristics									
Adult is male	.02	.01	.01	.02	-.01	.02	.01	.02	-.02
Adult never married	.64	.61	.65	.58	.07*	.63	.65	.58	.07
Adult was teen parent	.26	.28	.29	.28	.01	.28	.31	.24	.07
Adult works	.23	.26	.25	.27	-.03	.23	.23	.22	.00
Adult on AFDC	.79	.78	.82	.75	.08*	.79	.83	.74	.09*
Adult has h.s. diploma	.36	.39	.38	.40	-.02	.38	.38	.38	-.00
Hhld member victimized by crime in past 6 months	.42	.44	.48	.40	.08*	.41	.41	.41	.01
Getting away from gangs or drugs a reason for moving	.78	.77	.78	.76	.01	.74	.76	.72	.04
Schools a reason for moving	.51	.50	.54	.47	.07	.56	.57	.55	.02
N (children)	1574	2067	964	1103		1433	860	573	

Notes: AFDC = Aid to Families with Dependent Children. CP = complier. NCP = non-complier. HH = household. † = applies only to ages zero to five at baseline. ‡ = applies only to ages six to 17 at baseline. + = difference between treatment and control mean is statistically significant at the five percent level. * = difference between treatment compliers and non-compliers is statistically significant at the five percent level. Variables presented in this table are covariates included in the regression models; age as of December 2001 is included in the model as a sixth order Legendre polynomial. In addition to the covariates shown, the regression models also control for child's age at baseline (ages six to 17 versus zero to five), adult characteristics (age categories, in school, has GED), household characteristics (car, disabled member, teenage children, household size), neighborhood characteristics (resided in at least five years, very dissatisfied with, safe at night, has friends there, has family there, adult chats with neighbors, adult would tell neighbor if saw child getting into trouble), and moving (moved three or more times during past five years, previously applied for Section 8 assistance, very sure would find new apartment).

TABLE 2 — EFFECTS ON SCHOOL AND NEIGHBORHOOD CONTEXT

	Control Mean				Experimental versus Control					Section 8 versus Control				
	All Ages (i)	Age 6-10 (ii)	Age 11-14 (iii)	Age 15-20 (iv)	N (v)	ITT All (vi)	ITT Age 6-10 (vii)	ITT Age 11-14 (viii)	ITT Age 15-20 (ix)	N (x)	ITT All (xi)	ITT Age 6-10 (xii)	ITT Age 11-14 (xiii)	ITT Age 15-20 (xiv)
A. Census tract characteristics														
Average tract poverty rate [ADDR]	.456	.475	.444	.450	3635	-.126* (.008)	-.134* (.012)	-.120* (.011)	-.123* (.011)	3004	-.106* (.008)	-.117* (.012)	-.106* (.010)	-.095* (.010)
Share of tract males who are unemployed [ADDR]	.203	.205	.194	.210	3546	-.033* (.006)	-.033* (.010)	-.022* (.008)	-.043* (.009)	2937	-.033* (.006)	-.033* (.010)	-.032* (.008)	-.032* (.010)
Average share of tract who are minority [ADDR]	.905	.908	.907	.900	3635	-.075* (.008)	-.072* (.013)	-.081* (.011)	-.073* (.013)	3004	-.032* (.009)	-.026* (.011)	-.043* (.012)	-.027* (.013)
Average share of tract with college degree† [ADDR]	.123	.118	.125	.125	3635	.059* (.005)	.065* (.007)	.057* (.006)	.056* (.007)	3004	.031* (.004)	.037* (.006)	.033* (.006)	.022* (.006)
B. Residential mobility and school changes														
Moved more than 10 miles from baseline address [ADDR]	.119	.111	.123	.123	3453	.061* (.021)	.070* (.032)	.057* (.028)	.057* (.028)	2850	.016 (.023)	.015 (.032)	.013 (.030)	.019 (.031)
Number of moves since RA [ADDR]	1.180	1.245	1.173	1.121	3635	.333* (.059)	.347* (.095)	.367* (.080)	.285* (.076)	3004	.532* (.071)	.626* (.112)	.611* (.097)	.360* (.085)
Average share of schools in main baseline school districts‡ [SCH]	.885	.892	.878	.886	3512	-.085* (.015)	-.111* (.024)	-.063* (.023)	-.081* (.022)	2858	-.050* (.016)	-.058* (.024)	-.032 (.024)	-.058* (.026)
School closer to baseline than 2002 address, age <= 16 [ADDR] [SCH]	.274	.277	.262	.301	2457	-.054* (.025)	-.102* (.034)	.003 (.035)	-.088 (.053)	2039	-.049 (.026)	-.101* (.035)	-.023 (.037)	.012 (.061)
Number of schools attended since RA§, age <= 16 [SCH]	2.105	1.447	2.403	2.917	2944	.062 (.039)	.061 (.047)	.086 (.058)	.003 (.101)	2432	.081 (.045)	.080 (.053)	.095 (.071)	.046 (.119)
Mid-grade school change since RA¶, age <= 16 [SCH]	.171	.116	.193	.247	2944	-.003 (.017)	-.005 (.023)	.011 (.026)	-.041 (.044)	2432	.019 (.020)	.035 (.029)	.006 (.029)	.015 (.054)

C. School peers, pupil-teacher ratio and climate														
Average share of students eligible for free lunch [SCH], excl. IL	.739	.807	.754	.655	2479	-.064* (.009)	-.082* (.014)	-.060* (.013)	-.051* (.014)	2201	-.035* (.009)	-.062* (.015)	-.022 (.015)	-.022 (.015)
Average share of minority students [SCH]	.912	.929	.904	.903	3489	-.046* (.007)	-.055* (.011)	-.036* (.012)	-.048* (.012)	2840	-.029* (.009)	-.016 (.010)	-.025 (.013)	-.046* (.017)
Average percentile rank of schools on state exams [SCH], excl. age > 13 for MD, NY	.148	.134	.151	.172	2742	.044* (.007)	.052* (.010)	.040* (.009)	.035* (.014)	2219	.020* (.007)	.019* (.009)	.030* (.011)	.003 (.013)
Pupil-teacher ratio [SCH], excl. MA	18.48	18.41	18.53	18.51	2681	-.06 (.18)	-.40 (.33)	-.17 (.30)	.40 (.26)	2149	-.19 (.21)	-.04 (.36)	-.22 (.34)	-.28 (.30)
School climate index, higher is better# [SR]	.646	.685	.649	.610	2871	-.010 (.011)	-.042* (.020)	-.015 (.015)	.026 (.022)	2367	.009 (.012)	-.014 (.021)	.017 (.017)	.017 (.024)
Gangs in school or neighborhood [SR]	.545	.446	.592	.550	3061	-.060* (.022)	-.054 (.047)	-.116* (.033)	-.005 (.034)	2523	-.019 (.025)	-.011 (.052)	-.060 (.037)	.020 (.038)

Notes to Table 2: ADDR = address history from tracking file linked to Census tract data. Average tract characteristics are the average for a child's residences from randomization through 2001, weighted by duration at each address. Values for non-census years were interpolated or extrapolated from 1990 and 2000 Census data. Share of males unemployed is for the child's address in 2002 based on 2000 Census data. SCH = school history data linked by school name to school demographics from the National Center for Education Statistics and to average scores on state exams from the National School-Level State Assessment Score Database. SR = child self-report, available only for children at least eight years old as of May 31, 2001. † = At least an associates or bachelors degree. ‡ = The main five baseline school district are Baltimore City, Boston, Chicago, Los Angeles Unified, and New York City school districts. § = Number of schools (grades 1 through 12) attended since RA. ¶ = Any mid-grade school change was defined as attending more than one school for the same (unrepeated) grade. # = The school climate index consists of agreeing or strongly agreeing that "discipline is fair," "I feel safe at [this] school," and "teachers are interested in students," and disagreeing or strongly disagreeing with the statements that "there is a lot of cheating on tests and assignments" and "disruptions by other students get in the way of my learning." n/a = not applicable. Intent-to-treat (ITT) estimates for "all ages" from equation (1) and for age groups from equation (2), using covariates as described in Table 1 and weights described in section V. * = statistically significant at the five percent level. Standard errors, adjusted for heteroskedasticity, are in parentheses.

TABLE 3 — DISTRIBUTION OF NEIGHBORHOOD AND SCHOOL CHARACTERISTICS
FOR TREATMENT COMPLIERS

	Experimental versus Control				Section 8 versus Control			
	CCM	TCM	TOT	(SE)	CCM	TCM	TOT	(SE)
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
A. Current Neighborhood Characteristics [ADDR]								
Percent poor in census tract (E-C n = 2820; S-C n = 2351)								
share in tracts with less than 10% poor	.048	.270	.221*	(.030)	.042	.078	.036	(.020)
... with less than 20% poor	.202	.586	.384*	(.049)	.164	.297	.133*	(.042)
... with less than 30% poor	.362	.748	.385*	(.055)	.358	.578	.219*	(.046)
... with less than 40% poor	.488	.905	.417*	(.054)	.538	.794	.256*	(.045)
... with less than 50% poor	.703	.970	.268*	(.045)	.708	.935	.227*	(.036)
Percent of tract residents who are white non-Hispanics (E-C n = 2820; S-C n = 2351)								
share in tracts more than 10% white non-Hispanic	.354	.512	.159*	(.053)	.287	.305	.019	(.043)
... more than 20% white non-Hispanic	.281	.395	.114*	(.049)	.214	.219	.005	(.039)
... more than 30% white non-Hispanic	.182	.334	.152*	(.046)	.148	.192	.044	(.036)
... more than 40% white non-Hispanic	.132	.261	.129*	(.042)	.104	.152	.048	(.032)
... more than 50% white non-Hispanic	.123	.213	.089*	(.039)	.085	.101	.015	(.031)
Statewide neighborhood poverty rank (higher rank indicates less poverty; E-C n = 2781; S-C n = 2323)								
share in tracts ranking above the 10th %ile	.265	.733	.468*	(.053)	.280	.502	.222*	(.046)
... ranking above the 20th %ile	.123	.572	.448*	(.041)	.084	.266	.182*	(.038)
... ranking above the 30th %ile	.038	.434	.396*	(.036)	.049	.133	.084*	(.027)
... ranking above the 40th %ile	.032	.282	.251*	(.031)	.036	.092	.056*	(.023)
... ranking above the 50th %ile	.029	.182	.152*	(.028)	.032	.064	.032	(.019)
B. Current or Most Recent School Characteristics [SCH]								
Percent of students who are free lunch eligible, excluding Chicago (E-C n = 1999; S-C n = 1783)								
share in schools with less than 10% eligible for free lunch	.023	.047	.024	(.017)	.024	.041	.016	(.015)
... with less than 20% eligible	.036	.112	.075*	(.022)	.017	.080	.063*	(.025)
... with less than 30% eligible	.069	.179	.109*	(.029)	.070	.120	.051	(.028)
... with less than 40% eligible	.074	.252	.178*	(.035)	.098	.161	.063*	(.031)
... with less than 50% eligible	.144	.352	.208*	(.041)	.197	.243	.046	(.035)

Percent of students who are white Non-Hispanic (E-C n = 2824; S-C n = 2303)								
share in schools more than 10% white non-Hispanic	.281	.411	.130*	(.044)	.241	.285	.044	(.037)
... more than 20% white non-Hispanic	.171	.310	.139*	(.041)	.158	.208	.051	(.033)
... more than 30% white non-Hispanic	.129	.252	.122*	(.038)	.108	.168	.061	(.032)
... more than 40% white non-Hispanic	.100	.192	.092*	(.033)	.087	.116	.030	(.028)
... more than 50% white non-Hispanic	.073	.155	.082*	(.031)	.055	.087	.032	(.026)
School rank on statewide exams, excluding age > 13 for MD, NY (E-C n = 2406; S-C n = 1970)								
share in schools ranking above the 10th %ile	.575	.741	.166*	(.053)	.538	.607	.068	(.043)
... ranking above the 20th %ile	.215	.439	.224*	(.047)	.192	.275	.083*	(.035)
... ranking above the 30th %ile	.143	.291	.147*	(.040)	.119	.161	.042	(.029)
... ranking above the 40th %ile	.087	.201	.114*	(.034)	.082	.103	.022	(.024)
... ranking above the 50th %ile	.061	.137	.076*	(.029)	.067	.076	.009	(.020)

Notes: Sample restricted to children ages 6 to 16. Distribution cutpoints are overlapping. CCM = control complier mean estimated as equal to the treatment complier mean minus the treatment-on-treated effect. TCM = unadjusted treatment complier mean. TOT = treatment-on-treated effect estimated using two stage least squares with assignment to a treatment group serving as an instrumental variable for treatment compliance. ADDR = address history from tracking file linked to 2000 Census tract data. For sample consistency, neighborhood characteristics are restricted to children for whom WJ-R scores were available. SCH = school history data linked by school name to data on student demographics from the National Center for Education Statistics and on state test scores from National Longitudinal School-Level State Assessment Score Database. Free school lunch program participation was generally not available for Chicago. California reports the number of children receiving free lunch rather than the number eligible. School exam rankings for Maryland and New York exclude children age 13 years and older as high school exam scores were generally not available at these sites. E-C n = number of observations included in the Experimental - Control comparison. S-C n = number of observations included in the Section 8-Control comparison. * = statistically significant at the 5 percent level. Standard errors for the TOT effect, adjusted for heteroscedasticity, are in parentheses.

TABLE 4 — EFFECTS ON TEST SCORES AND ON BEHAVIOR AND SCHOOL PROBLEMS

	Control Mean				Experimental versus Control					Section 8 versus Control				
	All Ages (i)	Ages 6-10 (ii)	Ages 11-14 (iii)	Ages 15-20 (iv)	ITT All Ages (v)	ITT Ages 6-10 (vi)	ITT Ages 11-14 (vii)	ITT Ages 15-20 (viii)	Linear Age Inter-action (ix)	ITT All Ages (x)	ITT Ages 6-10 (xi)	ITT Ages 11-14 (xii)	ITT Ages 15-20 (xiii)	Linear Age Inter-action (xiv)
A. Test Scores														
Combined broad reading and math scores [WJR]	.000	-.857	.281	.590	.018 (.030)	-.009 (.054)	.034 (.044)	.029 (.049)	.0016 (.0081)	-.000 (.032)	-.046 (.055)	-.031 (.047)	.073 (.056)	.0064 (.0081)
Broad reading score [WJR]	.000	-.776	.224	.568	.032 (.031)	.016 (.058)	.055 (.048)	.027 (.050)	-.0005 (.0084)	.031 (.034)	.009 (.062)	.008 (.051)	.072 (.058)	.0011 (.0088)
Broad math score [WJR]	.000	-.872	.323	.561	-.009 (.030)	-.037 (.052)	-.004 (.045)	.017 (.051)	.0023 (.0081)	-.037 (.032)	-.083 (.055)	-.074 (.047)	.046 (.056)	.0069 (.0080)
B. Behavior and school problems														
Index of eleven behavior problems† [PR]	.270	.285	.284	.239	.017 (.011)	.006 (.018)	.035* (.018)	.009 (.017)	.0005 (.0024)	.006 (.012)	.005 (.018)	.011 (.019)	.001 (.019)	-.0007 (.0025)
Ever repeated a grade in school [PR]	.221	.174	.226	.265	.024 (.016)	.002 (.027)	.044 (.030)	.024 (.029)	.0026 (.0039)	-.020 (.017)	-.010 (.030)	-.007 (.032)	-.042 (.030)	-.0035 (.0040)
Suspended from school in past two years [PR]	.166	.114	.208	.176	-.001 (.013)	-.016 (.020)	.018 (.026)	-.007 (.024)	.0003 (.0032)	-.010 (.015)	-.009 (.021)	.012 (.029)	-.033 (.025)	-.0042 (.0033)
School asked to come in and talk about problems‡ [PR]	.300	.294	.332	.262	.011 (.018)	.016 (.032)	.016 (.031)	-.004 (.034)	.0004 (.0052)	.008 (.021)	.003 (.034)	.016 (.036)	.005 (.038)	.0034 (.0057)

Notes: WJR = Woodcock Johnson-Revised battery of tests. PR = parental report. † = The behavior problem index is the fraction of the following eleven items that the adult respondent reported were "often" or "sometimes" true for the child: "cheats or tells lies," "bullies or is cruel or mean to others," "hangs around with kids who get into trouble," "is disobedient at school," "has trouble getting along with teachers," "has difficulty concentrating, cannot pay attention for long," "is restless or overly active, cannot sit still," "has a very strong temper and loses it easily," "disobedient at home," "has trouble getting along with other children," and "withdrawn, does not get involved with others." ‡ = during past two years the school asked someone to come in to talk about problems the child was having with schoolwork or behavior. Intent-to-treat (ITT) estimates for all ages from equation (1) and estimates for age groups from equation (2), using covariates as described in Table 1 and weights described in section V. Linear age interaction is the estimated additional treatment effect from equation (3) of an additional year of age. * = statistically significant at the five percent level. Standard errors, adjusted for heteroskedasticity, are in parentheses.

TABLE 5 — EFFECTS ON SCHOOL ENGAGEMENT AND SPECIAL CLASSES

	Control Mean				Experimental versus Control					Section 8 versus Control				
	All Ages (i)	Ages 8-10 (ii)	Ages 11-14 (iii)	Ages 15-18 (iv)	ITT All Ages (v)	ITT Ages 8-10 (vi)	ITT Ages 11-14 (vii)	ITT Ages 15-18 (viii)	Linear Age Inter-action (ix)	ITT All Ages (x)	ITT Ages 8-10 (xi)	ITT Ages 11-14 (xii)	ITT Ages 15-18 (xiii)	Linear Age Inter-action (xiv)
A. School Engagement														
Always pays attention in class [SR]	.548	.598	.560	.487	.022 (.023)	.077 (.042)	-.025 (.032)	.051 (.041)	-.0060 (.0077)	.026 (.025)	.031 (.045)	-.007 (.037)	.074 (.045)	-.0000 (.0085)
Works hard in school [SR]	.573	.669	.584	.477	-.000 (.023)	.064 (.041)	-.015 (.032)	-.028 (.041)	-.0113 (.0078)	.001 (.024)	.041 (.044)	-.024 (.036)	.008 (.044)	-.0103 (.0085)
Reads at least five hours/week excluding schoolwork [SR]	.261	.254	.250	.283	-.004 (.021)	-.033 (.038)	-.011 (.029)	.029 (.037)	.0096 (.0071)	-.000 (.023)	-.035 (.042)	.003 (.032)	.023 (.043)	.0073 (.0081)
Late less than once a month [SR]	.482	.592	.504	.358	.013 (.023)	-.038 (.044)	.040 (.032)	.018 (.039)	.0055 (.0078)	.017 (.026)	.004 (.050)	.064 (.037)	-.041 (.043)	-.0072 (.0085)
B. Special Classes														
Class for gifted students or did advanced work in past two years [PR]	.115	.120	.117	.109	.019 (.014)	.008 (.027)	.025 (.021)	.018 (.026)	.0006 (.0052)	.002 (.015)	.004 (.030)	.007 (.021)	-.006 (.026)	-.0013 (.0050)
Special class or help for learning, behavioral or emotional problems in past two years [PR]	.255	.250	.266	.241	.019 (.018)	.053 (.041)	.008 (.029)	.006 (.030)	-.0034 (.0065)	.013 (.020)	.043 (.041)	.022 (.032)	-.024 (.033)	-.0067 (.0071)

Notes: SR = child self-report. PR = parental report. Intent-to-treat (ITT) estimates for all ages from equation (1) and estimates for age groups from equation (2), using covariates as described in Table 1 and weights described in section V. Linear age interaction is the estimated additional treatment effect from equation (3) of an additional year of age. * = statistically significant at the five percent level. Standard errors, adjusted for heteroskedasticity, are in parentheses.

TABLE 6 — EFFECTS BY SUBGROUP ON ACHIEVEMENT TEST SCORES AND BEHAVIOR PROBLEMS

	Broad Reading Z-Score [WJR]			Broad Math Z-Score [WJR]			Behavior Problems Index‡ [PR]		
	CM (i)	E-C ITT (ii)	S-C ITT (iii)	CM (iv)	E-C ITT (v)	S-C ITT (vi)	CM (vii)	E-C ITT (viii)	S-C ITT (ix)
A. Gender									
Girls	.103	.060 (.038)	.012 (.043)	.050	.028 (.041)	-.037 (.044)	.237	-.010 (.014)	-.007 (.015)
Boys	-.096	-.002 (.045)	.033 (.046)	-.046	-.029 (.040)	-.038 (.041)	.301	.041* (.014)	.025 (.016)
B. Race and ethnicity									
African-American non-Hispanic	-.058	.084* (.036)	.069 (.042)	.003	-.005 (.035)	-.045 (.040)	.290	.011 (.014)	-.004 (.015)
Hispanic	.035	-.033 (.053)	-.031 (.052)	-.060	-.034 (.051)	-.043 (.048)	.244	.020 (.018)	.034 (.018)
C. BL characteristics predictive of test scores†									
Predictive of low test scores	-.283	.072 (.054)	.070 (.057)	-.263	.027 (.049)	-.032 (.055)	.304	.006 (.019)	-.008 (.020)
Predictive of moderate test scores	-.050	-.038 (.043)	-.042 (.050)	-.048	-.077 (.044)	-.057 (.051)	.271	.008 (.018)	-.007 (.018)
Predictive of high test scores	.371	.017 (.045)	-.043 (.048)	.350	-.053 (.045)	-.090 (.048)	.232	.032* (.016)	.018 (.019)
D. BL problems, ages six to 20									
Behavioral, emotional or learning problems	.012	.067 (.093)	.145 (.112)	.174	-.084 (.093)	-.050 (.116)	.319	.048 (.031)	.051 (.033)

Notes: BL = baseline. WJR = Woodcock Johnson-Revised battery of tests. PR = parental report. † = subgroups by characteristics predictive of low, moderate, and high test scores created as described in section V. ‡ = The behavior problem index is the fraction of the following eleven items that the adult respondent reported were "often" or "sometimes" true for the child: "cheats or tells lies," "bullies or is cruel or mean to others," "hangs around with kids who get into trouble," "is disobedient at school," "has trouble getting along with teachers," "has difficulty concentrating, cannot pay attention for long," "is restless or overly active, cannot sit still," "has a very strong temper and loses it easily," "disobedient at home," "has trouble getting along with other children," and "withdrawn, does not get involved with others." CM = Control mean. E-C = experimental versus control group comparison. S-C = Section 8 group versus control group comparison. Intent-to-treat (ITT) estimates from equation (1) restricting to the subgroup of interest and using covariates as described in Table 1 and weights described in section V. Estimates are for children ages six to 20 as of December 31, 2001, except for the behavior/learning problem and gifted subgroups, which are only applicable to children ages 11 to 20 in 2001 and at least six years old at baseline. The effects of the experimental treatment on broad reading are estimated using the scores for 1880 girls, 1803 boys, 2405 African-Americans, 1069 Hispanics, 1196 with characteristics predictive of low scores, 1237 with characteristics predictive of moderate scores, 1254 with characteristics predictive of high scores, and 409 with behavior, emotional, or learning problems in the experimental or control group. Effects of being offered a Section 8 voucher are estimated using the scores of 1520 girls, 1518 boys, 1857 African-Americans, 973 Hispanics, 978 with characteristics predictive of low scores, 1037 with characteristics predictive of moderate scores, 1023 with characteristics predictive of high scores, and 344 with behavior, emotional, or learning problems in the Section 8 or control group. * = statistically significant at the five percent level.