Title
Neighborhood effects on use of African-American Vernacular English

Journal
Proceedings of the National Academy of Sciences of the United States of America, 112(38)

ISSN
0027-8424

Authors
Rickford, John R
Duncan, Greg J
Gennetian, Lisa A
et al.

Publication Date
2015-09-22

DOI
10.1073/pnas.1500176112

Peer reviewed

Neighborhood effects on use of African-American Vernacular English

John R. Rickford[a,1], Greg J. Duncan[b], Lisa A. Gennetian[c], Ray Yun Gou[d], Rebecca Greene[a], Lawrence F. Katz[e,d], Ronald C. Kessler[f], Jeffrey R. Kling[g,d], Lisa Sanbonmatsu[d], Andres E. Sanchez-Ordoñez[h], Matthew Sciandra[d], Ewart Thomas[h], and Jens Ludwig[i,d]

[a]Department of Linguistics, Stanford University, Stanford, CA 94305; [b]School of Education, University of California, Irvine, CA 92697; [c]Institute of Human Development and Social Change, New York University, New York, NY 10003; [d]National Bureau of Economic Research, Cambridge, MA 02138; [e]Department of Economics, Harvard University, Cambridge, MA 02138; [f]Department of Health Care Policy, Harvard Medical School, Boston, MA 02115; [g]Congressional Budget Office, Washington, DC 20515; [h]Department of Psychology, Stanford University, Stanford, CA 94305; and [i]Harris School of Public Policy, University of Chicago, Chicago, IL 60637

Edited by Sara S. McLanahan, Princeton University, Princeton, NJ, and approved July 28, 2015 (received for review January 8, 2015)

African-American Vernacular English (AAVE) is systematic, rooted in history, and important as an identity marker and expressive resource for its speakers. In these respects, it resembles other vernacular or nonstandard varieties, like Cockney or Appalachian English. But like them, AAVE can trigger discrimination in the workplace, housing market, and schools. Understanding what shapes the relative use of AAVE vs. Standard American English (SAE) is important for policy and scientific reasons. This work presents, to our knowledge, the first experimental estimates of the effects of moving into lower-poverty neighborhoods on AAVE use. We use data on non-Hispanic African-American youth (n = 629) from a large-scale, randomized residential mobility experiment called Moving to Opportunity (MTO), which enrolled a sample of mostly minority families originally living in distressed public housing. Audio recordings of the youth were transcribed and coded for the use of five grammatical and five phonological AAVE features to construct a measure of the proportion of possible instances, or tokens, in which speakers use AAVE rather than SAE speech features. Random assignment to receive a housing voucher to move into a lower-poverty area (the intention-to-treat effect) led youth to live in neighborhoods (census tracts) with an 11 percentage point lower poverty rate on average over the next 10–15 y and reduced the share of AAVE tokens by ∼3 percentage points compared with the MTO control group youth. The MTO effect on AAVE use equals approximately half of the difference in AAVE frequency observed between youth whose parents have a high school diploma and those whose parents do not.

neighborhood effects | segregation | language | African-American Vernacular English | code switching

Language is in many respects a socially constructed behavior, jointly influenced by exposure, identity, and peer group influence (1). One's speech patterns are shaped not only by one's family, but also by one's broader regional and social environment. For example, people who immigrate from non-English-speaking countries to the United States at an early age wind up speaking English with nearly the same proficiency as those who were born in the United States, even though their older siblings and parents do not (2, 3). Less clear is whether different local social environments within a city, state, or country exert causal effects on the use of dialects such as African-American Language (4) or African-American Vernacular English (AAVE), which is the most vernacular variety of African-American English and is used across the country (5–7).

This work presents what, to our knowledge, is the first study of how much social environments—neighborhoods—exert a causal effect on the use of AAVE. Previous research in sociolinguistics has documented substantial variation in AAVE use by socioeconomic class, defined by using various combinations of occupational status, education, and income or residence quality (8, 9). There are theoretical reasons to believe any or all of these measures shape AAVE use by neighborhood (as discussed further below and in *SI Appendix*). However, this correlation may not reflect the causal effect of neighborhood environments on language and could instead be driven by the effects of unmeasured person- or family-level variables that jointly determine both residential location and speech patterns. Causal inference about the effects of neighborhoods on speech is more convincing if based on a study that uses a randomized experimental design to assign similar families to live in different types of neighborhood contexts.

Evidence for neighborhood effects on AAVE use is relevant for understanding the degree to which future changes in neighborhood economic and racial segregation may affect the vitality and use of this dialect (10, 11). This is a topic of importance to sociolinguists, because vernaculars have benefits as in-group markers and expressive resources (12). Such evidence is also relevant for understanding how changes in segregation will affect disparities in other life outcomes because previous studies suggest that AAVE use could affect children's school success—at least given the way schools currently operate—and that AAVE speakers are often victims of what Baugh calls "linguistic profiling" (13)—discrimination in the workplace, housing markets,

SOCIAL SCIENCES

Significance

We provide, to our knowledge, the first experimental evidence of neighborhood effects on the use by low-income minority youth of African-American Vernacular English (AAVE). Rising U.S. residential economic segregation may be contributing to growing differences within the population in AAVE use, which has benefits to in-group solidarity and identity but is associated with discrimination in schools and workplaces and so may exacerbate the disadvantages of youth growing up in high-poverty areas. To the extent that the association between AAVE use and income represents a causal effect of AAVE use, our illustrative calculations suggest that neighborhood effects on speech could increase lifetime earnings by approximately $18,000 (∼3–4% of lifetime income).

Author contributions: J.R.R., G.J.D., L.A.G., R.Y.G., R.G., L.F.K., R.C.K., J.R.K., L.S., M.S., and J.L. designed research; J.R.R., G.J.D., L.A.G., R.Y.G., R.G., L.F.K., R.C.K., J.R.K., L.S., A.E.S.-O., M.S., E.T., and J.L. performed research; J.R.R., R.Y.G., R.G., J.R.K., L.S., A.E.S.-O., M.S., E.T., and J.L. analyzed data; and J.R.R., G.J.D., L.F.K., R.C.K., J.R.K., L.S., M.S., E.T., and J.L. wrote the paper.

Conflict of interest statement: The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: A restricted-access version of the data used in this paper will be provided to the U.S. Department of Housing and Urban Development (HUD). At the time of writing, HUD plans to make the data available to responsible researchers through a data license (details at www.huduser.org/portal/research/pdr_data-license.html). HUD expects to arrange for alternative archiving of the data in the future.

[1]To whom correspondence should be addressed. Email: rickford@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1500176112/-/DCSupplemental.

and schools (7, 14–18). Of course, efforts to modify how social institutions interact with people using different dialects are independently important, regardless of any relationship between neighborhood segregation and AAVE use.

Our study capitalizes on a unique opportunity to understand neighborhood effects on speech (AAVE use) by incorporating sociolinguistic measures of speech and language patterns into the long-term follow-up of participants in a large-scale government residential-mobility experiment called Moving to Opportunity (MTO). We believe this type of language measurement has never before been incorporated into a large, randomized social experiment. We provide estimates of the causal effect of changes in neighborhood environments on speech patterns by using MTO's randomly assigned variation in opportunities for poor families to move to low-poverty areas.

MTO

The MTO experiment was designed and carried out by the U.S. Department of Housing and Urban Development (HUD). Eligibility was limited to families with children living in public housing projects located in selected distressed neighborhoods in five cities (Baltimore, Boston, Chicago, Los Angeles, and New York City). Approximately one-quarter of eligible families applied to MTO. From 1994 to 1998, MTO enrolled a total of 4,604 low-income, mostly minority families who were then randomly assigned by lottery to:

(*i*) A "low-poverty voucher (LPV) group" that was offered housing search assistance and housing vouchers that could only be used to relocate to a low-poverty census tract (one with a 1990 poverty rate <10%); or

(*ii*) A "control group" that received no special assistance, although some families can (and did) move by themselves.

(The MTO experiment also included a traditional voucher group that was offered housing vouchers that they could use to move to a new private-market apartment of their choice; for budgetary reasons, we did not collect speech data from that group.)

Not all households assigned to the LPV group relocated through the MTO program. In our analyses, we compare average speech patterns of all families assigned to the LPV group (regardless of whether they moved through MTO) vs. all controls. In the medical literature, this comparison is the basis of "intention-to-treat" (ITT) analysis and preserves the key strength of MTO's experimental design. As we show below, MTO generated large differences in the socioeconomic composition of neighborhoods experienced by otherwise similar groups of low-income families, and so helps overcome the self-selection problem that plagues nonexperimental studies of neighborhood effects on speech and other outcomes.

In earlier work examining various outcomes of youth in MTO, Ludwig et al. found no detectable effects on youth schooling and youth physical health of being in a family assigned to the LPV group (19). But they found gender-based differential effects of the LPV on other youth outcomes, with girls doing better on some measures than their control-group counterparts and boys doing worse relative to their controls (see also ref. 20). A more recent study of MTO youth by Chetty et al. that incorporates longer-term follow-up data shows sizable impacts of MTO moves to less-disadvantaged neighborhoods on earnings during adulthood of youth who were relatively young when their families moved in MTO (21).

The present work examines the degree to which this mobility experiment changes the speech patterns of MTO youth.

Results

In this study, we focus on MTO youth rather than adults, because propensity for second-language or dialect acquisition is more pronounced for youth than adults (22–24). Because our focus in this paper is on use of AAVE, we restrict attention to non-Hispanic African-American youth ($n = 629$). Table 1 presents the average baseline characteristics for the youth in our study sample by randomized group (LPV and control; see also *SI Appendix*, Table S3). The study sample is very economically disadvantaged; at baseline, four of five households were receiving cash welfare. The average poverty rate in the baseline neighborhoods (census tracts) was almost 60%, and >70% of residents were African-American. Statistical tests fail to reject the null hypothesis that the average baseline characteristics are the same for the LPV and control groups, confirming that MTO randomization was carried out correctly.

All families in the LPV group were offered the opportunity to move to a low-poverty neighborhood. However, of non-Hispanic African-American youth assigned to the LPV group who are in our speech analysis sample, only 52% of their families used an MTO voucher to relocate to a low-poverty census tract. Fig. 1 shows that random assignment to the LPV rather than control group generated larger differences in neighborhood disadvantage than in neighborhood racial composition. One year after random assignment, LPV group youth lived in neighborhoods with an average poverty rate 21 percentage points lower than the control group mean of 56%. This change equals 1.27 SDs of the tract poverty distribution within the control group (and 1.71 SDs in the national tract poverty distribution; *SI Appendix*, Table S6).

The size of the MTO effect on neighborhood poverty decreased over time, and the difference in poverty rates in the neighborhoods in which participants were living just before the beginning of the MTO long-term survey fielding period (10–15 y after random assignment) is only 4 percentage points (approximately a quarter of a control group SD). However, over the full course of the 10- to 15-y follow-up period, the average poverty rate difference is 11 percentage points (control mean 43%, $P < 0.05$). (See *SI Appendix*, Table S6 for additional results.)

MTO also changed other measures of neighborhood economic composition that are correlated with poverty and that could also be relevant for the language environment. For example, within the MTO control group, we find that AAVE use by youth is correlated with parent education (*SI Appendix*, Table S14), and MTO generated large changes in the share of adults in a family's neighborhood with a high school or college degree (*SI Appendix*, Table S6). Additionally, MTO had effects on the characteristics of the schools that these youth attended, although they are typically smaller in proportional terms than are MTO's effects on neighborhood characteristics (*SI Appendix*, Table S7).

All youth in the speech collection pool were asked to respond to an engaging open-ended question (a question about either the happiest or the scariest moment in their life) and to answer an open-ended question at the end of our interview about whether the respondent had anything else to say about their neighborhood or housing programs more generally. Written informed consent was obtained before beginning interviews. The intent was to elicit informal speech, but speech would still be relatively formal given the interviewer's status as a stranger recording an interview. These speech samples were transcribed by trained linguistic listeners (see *SI Appendix* for details). The unit of observation in our data is the "token"—that is, an occurrence of a selected speech variable or feature in which speakers have the option of using AAVE rather than Standard American English (SAE).

Tokens were coded for use of AAVE rather than SAE for 10 language features (five grammatical and five phonological) that have been shown in previous research to distinguish the two dialects (5, 8, 25). Of the 14,191 tokens we analyzed in our dataset, 1,492 (11%) represented grammatical features and 12,699 (89%) were phonological (pronunciation) features (*SI Appendix*, Table S2). Our key dependent variable is whether a given token is realized by a grammatical or phonological AAVE

Table 1. Baseline characteristics by treatment group

Characteristics	Control mean	LPV mean
Youth characteristics		
Male	0.554	0.494
Age in whole years as of December 31, 2007	16.615	16.643
Non-Hispanic African-American	1.000	1.000
Health problems that limited activity	0.047	0.053
Health problems that required special medicine or equipment	0.070	0.076
Household characteristics		
Adult was employed at baseline	0.245	0.258
Receiving Aid to Families with Dependent Children	0.818	0.853
Neighborhood characteristics		
Census tract poverty rate	0.588	0.584
Census tract share black	0.719	0.716
Primary move reason was to get away from drugs and gangs	0.535	0.535
Primary move reason was better schools for children	0.164	0.189
Randomization site		
Baltimore	0.202	0.167
Boston	0.128	0.087
Chicago	0.374	0.338
Los Angeles	0.148	0.239*
New York	0.148	0.169

All values represent shares (except age). Values are calculated by using sample weights to account for changes in random assignment ratios across randomization cohorts, survey sample selection, two-phase interviewing, and language sample selection. The sample is non-Hispanic African-American youth speakers (ages 13–20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only whose speech samples included at least one analyzable language token ($n = 629$ youth). *$P < 0.05$ on an independent group t test of the difference between the LPV group and the control group.

variant rather than its SAE variant. To assess the reliability of our measure, we estimated that AAVE use rates have a correlation of 0.39 among siblings ($P < 0.001$) and that AAVE use among the first and second halves of each youth's tokens are significantly correlated with each other—see *SI Appendix* for details. The validity of our measure is suggested by the fact that AAVE use rates follow patterns reported in previous studies that measure AAVE use in other ways—for example, higher among African-Americans than Hispanics and higher among youth whose parents have less schooling.

The first row of Table 2 presents our key results for the effects of the MTO experiment on AAVE use, showing that assignment to the LPV rather than control group causes a decline in AAVE use of 2.8 percentage points (95% confidence interval + 0.1 to −5.7 percentage points; $P = 0.056$). The control mean is 48.5% of tokens using AAVE. This ITT estimate is based on a model that controls only for baseline characteristics measured before MTO random assignment. To provide some context for the size of this effect, within the control group, the difference in AAVE prevalence between youth whose head of household does vs. does not have a high school diploma equals 5.4 percentage points (see *SI Appendix*, Table S14 for details). So the MTO effect on AAVE use is approximately half the difference in AAVE use between children whose parents graduated high school and those who did not.

When we split our analysis by language feature, we find a statistically significant effect on phonological tokens, but not grammatical tokens, although we cannot reject the null hypothesis that the two effect sizes are the same (*SI Appendix*, Table S11).

Also of interest is how gender moderates MTO's effects on AAVE use. Previous research in sociolinguistics finds a "conservative tendency" in speech patterns among women, who "show a lower rate of stigmatized variants and a higher rate of prestige variants than men," and who seem to adopt prestige forms at a higher rate than men (ref. 1, pp. 266–267 and 274; see also refs. 5 and 26–28). Even for young children, boys use AAVE variants during spontaneous discourse at approximately twice the rate of girls (29, 30). However, in the MTO data, we find that overall rates of AAVE use within the control group are similar for boys vs. girls (48.9% vs. 48.2%; Table 2). The estimated MTO effect on reducing AAVE use is larger for females compared with males (−4.5 vs. −1.1 percentage points), but we cannot reject the null that they are the same ($P = 0.245$; *SI Appendix*, Table S9).

Age patterns in vernacular use have also been of great interest to sociolinguists. In observational studies, younger children (e.g., age 13 or younger) seem more likely to change speech patterns after their families move to a new area than older youth (refs. 23; 24; and 31, p. 176). All else equal, that would lead us to predict that children who are relatively younger at baseline should exhibit more pronounced changes in AAVE use. Conversely, being relatively older at the time of our MTO long-term data collection may moderate MTO's effects, because previous research suggests that AAVE use increases during adolescence but begins to decline as youth enter college or the labor market (25, 31–34).

Table 2 highlights the challenge of testing for age effects in the MTO data: We see signs that older age (17+) at the time of data collection is associated with larger increases in AAVE use compared with those <17, but this difference could be due to the fact that youth of different ages at the time of our survey also had different average baseline ages. More generally, the MTO design does not make it possible to disentangle the independent effects of age at baseline, age at the time of our survey, and duration of exposure to the MTO experiment; the latter is just the difference between the first two factors, so we cannot vary one while holding the other two constant.

Subgroup analyses suggest that household heads' motivation for enrolling in MTO may be an important moderator of MTO effects on AAVE use. Table 2 shows that, for youth whose parents reported at baseline that they signed up for MTO to either get away from drugs and gangs in their current (baseline) neighborhoods or to access better schools, MTO treatment assignment reduces AAVE use by 5.2 percentage points (95% confidence interval −8.7 to −1.7 percentage points; $P = 0.003$). For the rest of the MTO youth sample, the effect of MTO treatment assignment on

SOCIAL SCIENCES

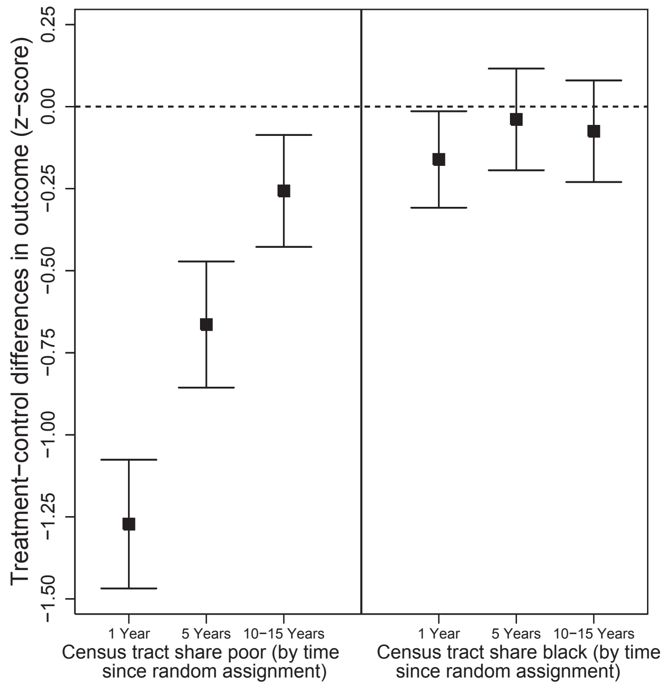

Fig. 1. MTO effects on neighborhood conditions. Impact on each outcome of assignment to the LPV group for non-Hispanic African-American youth whose language sample from the MTO long-term survey was analyzed. The squares represent the ITT estimate for the effect of being assigned to the LPV group, rather than control, for the outcomes listed on the *x* axis: neighborhood (census tract) share poor and share black at the address where the youth was living 1, 5, and 10–15 y after random assignment (the 10- to 15-y address is where the youth was living as of May 31, 2008, just before the beginning of the long-term survey fielding period). Share poor and black are z-scores, standardized by the control group mean and SD. The box whiskers represent the 95th percent confidence interval around the estimates. Census tract characteristics are based on interpolated data from the 1990 and 2000 decennial Censuses as well as the 2005–2009 American Community Survey.

AAVE use is a positive and statistically insignificant 3 percentage points. The difference between the estimated MTO effects for the two subgroups is 8.3 percentage points ($P = 0.007$; *SI Appendix*, Table S10). However, given the number of subgroups analyzed (*SI Appendix*, Tables S9 and S10), this result could be a "false positive" and thus should be viewed as only suggestive.

The *SI Appendix* shows that our results are qualitatively robust to a number of different decisions about how we define our sample and carry out our analyses. For example, the results are similar when we calculate average marginal effects using probit or logit models instead of linear regression, or limit the sample to youth whose language samples are above some threshold size, or collapse the data and carry out analysis at the person, rather than token, level.

Discussion

To our knowledge, ours is the first study to use data from a randomized experiment to examine whether moving into a more economically advantaged neighborhood causes a decline in the rate at which speakers use AAVE vs. SAE. On average, youth in the control group in our sample used AAVE in ~49% of the speech tokens that we collected and analyzed 10–15 y after baseline. Random assignment to the LPV group, members of which were given the opportunity to use a housing voucher to move out of a high-poverty public housing project into a less-distressed neighborhood, reduced AAVE prevalence by ~3 percentage points.

To take advantage of the key strengths of MTO (the randomized experimental design and large sample), we collected speech data from as large a sample as possible. However, given constraints on budget and how long we could spend with respondents for the multipurpose MTO study, the speech samples we collected were relatively short (*SI Appendix*). Relative to previous sociolinguistic research, the amount of speech recorded per person is lower, but the number of respondents is much greater. Because our focus is on comparing group averages (speech patterns of people who were, vs. were not, randomized to have the chance to move to a less-distressed neighborhood), the total volume of speech used in our main estimates is quite large (n = 14,191 tokens total).

It is important to keep in mind that the MTO study sample is not representative of all American households. Although the families that signed up for MTO are generally similar to other urban minority samples in high-poverty urban areas that have been studied in the "neighborhood effects" literature (35, 36), the families living in high-poverty urban areas are much more disadvantaged on average than other American families. Of course how to help these very disadvantaged families is of particular policy concern.

Given our research design, we cannot isolate what specific features of the neighborhood social environment are responsible for the observed effects on AAVE use, although we do see that MTO generated larger changes in the LPV group's neighborhood poverty rate than racial composition. MTO also led LPV youth to live in areas with more highly educated adults, who are more likely to work in occupations ranked higher on the "linguistic market" in requiring or rewarding standard rather than vernacular use (31, 37). In addition, LPV youth attended schools with somewhat fewer black and minority students, and so may have experienced less peer group pressure to retain the vernacular and avoid "acting white" (38). Potential effects of social class and network differences between the control and LPV neighborhoods on vernacular use are discussed at greater length in *SI Appendix*, taking the sociolinguistic literature into account.

Our MTO data are informative about the types of neighborhood changes induced by this intervention, which could be different from the impacts of even more dramatic mobility interventions—for example, moving low-income minority families like those in MTO out to very affluent, predominantly white suburbs. Our data also do not allow us to determine how much of the neighborhood effect on AAVE use is mediated by neighborhood effects on language use (or other things) in the home.

Although the youth in the LPV group do demonstrate an ability to use more SAE-like speech in a formal setting (an interview with a stranger), our experiment should not be interpreted as indicating that they have abandoned AAVE more generally. We have no evidence on these youths' informal vernacular use with family members and friends, and it is likely, if not certain, that their informal speech contains higher frequencies of AAVE features. At the same time, virtually all discussion in the linguistics literature of the issue of developing bidialectal competence in AAVE and other vernacular speakers has emphasized the feasibility and value of extending speakers' repertoire to include command of SAE when needed, rather than replacing competence in the vernacular with competence in the standard (for a summary, see ref. 39). That extension of bidialectal competence for MTO youth in the SAE domain is what the experiment discussed in this paper has demonstrated.

We found that youth in families that enrolled in MTO because they cared the most about crime or school quality in their baseline public housing projects experienced the largest reductions in AAVE speech. One possible explanation is suggested by the ethnographic work of Anderson (40), who notes that within high-poverty, inner-city neighborhoods, residents who identify as "street" are less able or willing to engage in "code switching" to SAE (pp. 35–36). A quarter century earlier, Labov (41) had noted that African-Americans who were not active participants

Table 2. MTO effects on youth AAVE use

Baseline characteristics used for subgroup analysis	Control mean	AAVE variant used in token: LPV vs. control ITT effect: Coefficient (SE)	*P* value	No. of tokens (no. of youth)
Overall	0.485	−0.028∼ (0.015)	0.056	14,191 (629)
By gender				
Female	0.482	−0.045* (0.020)	0.030	7,347 (307)
Male	0.489	−0.011 (0.021)	0.589	6,844 (322)
By age in 2008				
Age <17 in 2008[†]	0.477	0.014 (0.026)	0.572	4,459 (210)
Age 17+ in 2008[‡]	0.489	−0.047* (0.017)	0.007	9,732 (419)
By household head's primary reason for wanting to move				
To get away from drugs and gangs or for better schools for the children	0.494	−0.052* (0.018)	0.003	9,754 (440)
Another reason	0.465	0.030 (0.025)	0.224	4,437 (189)

*$P < 0.05$, ∼$P < 0.10$ on two-tailed *t* test. Robust SEs are shown in parentheses. LPV vs. control ITT effects were estimated by using an ordinary least squares regression model controlling for the baseline covariates in *SI Appendix*, Table S3, using person-level survey weights, and clustering by family ID. Subgroup analyses were run as an interaction with the treatment group indicator. Tokens represent each instance where the speaker used 1 of the 10 language features. Tokens were analyzed for whether the speaker used the AAVE or the SAE variant for that token. The sample is all tokens from the speech samples of non-Hispanic African-American youth (ages 13–20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only (n = 14,191 tokens from n = 629 youth).

[†]Baseline age: mean 2.4, range 0–5; mean exposure, 12.5 y.

[‡]Baseline age: mean 6.1, range 2–11; mean exposure, 12.5 y.

in the street culture of adolescent Harlem gangs or peer groups, partly because of parental pressures, were also less consistent users of AAVE. It may be that youth in families most concerned about the potentially adverse effects of the "code of the street" participated less in vernacular peer groups and were most prepared to switch away from AAVE when moving to lower-poverty areas.

Our estimates have implications for how neighborhood effects on AAVE use may shape the long-term life outcomes of these youth. As noted above, given the current structure of American society, AAVE use is associated with adverse schooling, housing and labor market outcomes. Many researchers and policymakers are appropriately concerned about and are working to change how individuals and organizations interact with people who use vernaculars. However, in the meantime, changes in neighborhood economic and racial segregation may change school or economic outcomes by changing AAVE speakers' inclination and ability to code switch to SAE.

One challenge for assessing this hypothesis is that there are no plausibly causal estimates of the relationship between AAVE use and different key life outcomes. However, as described in *SI Appendix*, we used data from the MTO control group to regress different outcomes against AAVE use and some basic demographic controls. This exercise is necessarily speculative, because at least part of the association between AAVE use and different long-term life outcomes may be due to omitted variables. With this caveat in mind, our illustrative calculations suggest that the effects on AAVE use of spending 10–15 y in a neighborhood with a poverty rate ∼11 percentage points lower than the youth would otherwise have (the MTO ITT effect) could increase annual earnings by approximately $350. Because only approximately half of the LPV group youth's families used their MTO vouchers to move to low-poverty areas, the effect on those who actually moved through MTO (the treatment-on-the-treated effect) could be as high as approximately $700 per year, or approximately $18,000 in present value over the youth's entire working career (3–4% of lifetime income). The results we present here thus may provide at least a partial explanation for the recent findings of long-term beneficial effects of approximately $3,500 per year on the adult earnings of MTO youth who were under age 13 at baseline (21).

Our finding that MTO moves change the frequency of AAVE use by youth is consistent with previous observational research in sociolinguistics showing that lower- and working-class status is correlated with the highest frequency of AAVE use within the African-American community (see, for example, ref. 8). Various studies have shown that the level of income segregation in American neighborhoods has been increasing since 1970 and that concentrated poverty (the share of poor families living in census tracts with poverty rates of 40% or more) has increased as well, including during the 2000s (42–44). Our findings raise the possibility that rising U.S. residential economic segregation may be contributing to growing differences within the population in AAVE use in a manner that could further exacerbate the economic disadvantages of youth growing up in high-poverty areas. Although efforts to eliminate discrimination in schools, labor markets, and criminal justice settings are of critical importance, policies to reverse the trend toward increased economic segregation may also play a role in shaping black-white inequality in language and hence in life outcomes.

Materials and Methods

Our research team subcontracted with the Institute for Social Research at the University of Michigan to collect in-person data on outcomes for the MTO study sample in 2008–2010, or 12 y after baseline on average (range 10–15 y). The effective response rates for these long-term follow-up surveys were 90% for the adult household heads in MTO and 89% for all youth who were selected for the survey (who were between the ages of 10 and 20 at the end of 2007). Among the non-Hispanic African-American youth on whom we focus in this paper (ages 13–20 only), response rates were similar across randomized MTO groups (*SI Appendix*, Table S1), as were the characteristics of the interviewers working with each group (*SI Appendix*, Table S4). We successfully collected and transcribed speech data from 71% of the youth who were eligible for language data collection (68% for the LPV group and 73% for the control group), which leaves us with a final sample size of n = 629. (See *SI Appendix* for additional details). Our study was reviewed and approved by the federal Office of Management and Budget and the Institutional Review Boards at HUD, the National Bureau of Economic Research, the University of Chicago, the University of Michigan, and Northwestern University.

The five grammatical language features that were coded as AAVE or SAE are as follows: (*i*) use of "ain't" rather than standard negators like "aren't," "isn't," and "hasn't"; (*ii*) multiple negation, involving the use of negative indefinites like "never," "nothing," or "no one" in addition to a negated auxiliary verb like "shouldn't"; (*iii*) absence of third singular present tense –"s"

SOCIAL SCIENCES

(as in "He walkØ" for "He walks"); (*iv*) absence of copula or auxiliary "is" or "are" (as in "They Ø happy" for "They are happy"); and (*v*) "was"-leveling (as in "They was nice" for "They were nice"). The five phonological language features were as follows: (*i*) consonant cluster reduction (as in "fas" for "fast"); (*ii*) r-deletion or vocalization after a vowel (as in "mothuh" for "mother"); (*iii*) DH-stopping (as in "dis" for "this"); (*iv*) TH-stopping (as in "wit" for "with" or "mout" for "mouth"); and (*v*) "ai" monophthongization (as in "rad" for "ride" or "ah" for "I"). A token can be a single phoneme or pronunciation segment (e.g., pronouncing "them" with either an initial "th" or with a "d") or a grammatical form (e.g., using the full, contracted, or deleted form of "is" in "She is∼'s∼Ø cold"). Note that a single phrase like "She should do nothing" can be examined for multiple examples of AAVE use—for example, both double negation ("She 'shouldn't' do 'nothing'" and use of "th" vs. "t" in "nothing"), so contributing two tokens to the analysis sample. Coding of both r-deletion and DH-stopping was generally capped at a maximum of 10 tokens per speaker.

To estimate the effects of the offer to use an LPV voucher, known as the ITT effect, we regress AAVE use (*Y*) on an indicator (*Z*) for whether youths' families were assigned to the LPV group instead of the control group and a set of baseline covariates (*X*) to improve statistical precision:

$$Y = Z\beta_1 + X\beta_2 + \varepsilon \quad [1]$$

We use linear regression to estimate Eq. **1**. We cluster SEs at the family level to account for the nonindependence of tokens taken from the same individual and from youth in the same families. The results are not sensitive to dropping the baseline covariates.

To approximate the effects of actually moving with an MTO voucher, under the assumption that assignment to the LPV group in MTO only affects the language use of those who move through MTO, we also report results dividing the ITT effect by the share of those assigned to the LPV group who relocate using an MTO voucher (*SI Appendix*, Table S12).

ACKNOWLEDGMENTS. The survey data collection effort was led by Nancy Gebler of the University of Michigan's Survey Research Center under subcontract to our research team, while initial transcription of the language samples analyzed here was carried out by Holly Craig and her team at the University of Michigan. Additional transcription and coding was carried out at Stanford University under the direction of Drs. Rickford and Greene; we thank Carolyn G. Chiu, Devon A. Holman, Cameron W. Jeffers, Ellie Redding, Trent Woodward, and Avery Youngblood. Thanks also to Jeffrey Grogger for helpful conversations. This research was supported by HUD Grant C-CHI-00808, which assisted with the design of the data collection protocol for the long-term MTO study and reviewed this manuscript prior to submission to ensure the manuscript did not violate the confidentiality of MTO program participants, but had no other involvement with the study. This work was also supported by National Science Foundation Grants SES-0527615 and 1125795; the National Institute for Child Health and Human Development Grants R01-HD040404, R01-HD040444, and R21-HD062870; Centers for Disease Control Grant R49-CE000906; National Institute of Mental Health Grant R01-MH077026; National Institute for Aging Grants P30-AG012810, R01-AG031259, and P01-AG005842-22S1; the National Opinion Research Center's Population Research Center (through National Institute of Child Health and Human Development Grant R24-HD051152-04); University of Chicago's Center for Health Administration Studies; U.S. Department of Education/Institute of Education Sciences Grant R305U070006; the John D. and Catherine T. MacArthur Foundation; the Smith Richardson Foundation; the Spencer Foundation; the Annie E. Casey Foundation; the Bill & Melinda Gates Foundation; and the Russell Sage Foundation. MTO data were provided by HUD. Any errors and all opinions are our own. The views expressed here are those of the authors and should not be interpreted as those of HUD, the Congressional Budget Office or the U.S. Government.

1. Labov W (2001) *Principles of Linguistic Change: Social Factors* (Blackwell, Oxford).
2. Bleakley H, Chin A (2010) Age at arrival, English proficiency, and social assimilation among U.S. immigrants. *Am Econ J Appl Econ* 2(1):165–192.
3. Kuziemko I (2014) Human capital spillovers in families: Do parents learn from or lean on their children? *J Labor Econ* 32(4):755–786.
4. Lanehart SL, ed (2015) *Oxford Handbook of African American Language* (Oxford Univ Press, Oxford).
5. Labov W (1972) *Language in the Inner City: Studies in the Black English Vernacular* (Univ of Pennsylvania Press, Philadelphia).
6. Green LJ (2002) *African American English: A Linguistic Introduction* (Cambridge Univ Press, Cambridge, U.K.).
7. Rickford JR (1999) *African American Vernacular English: Features and Use, Evolution, and Educational Implications* (Blackwell, Oxford).
8. Wolfram WA (1969) *A Sociolinguistic Description of Detroit Negro Speech* (Center for Applied Linguistics, Washington, DC).
9. Labov W, Cohen PS, Robins C, Lewis J (1968) *A Study of the Non-Standard English of Negro and Puerto-Rican Speakers in New York City* (US Regional Survey, Philadelphia), Final Report, Cooperative Research Project 3228.
10. Labov W (2010) Unendangered dialect, endangered people: The case of African American Vernacular English. *Transforming Anthropol* 18(1):15–27.
11. Rickford JR (2010) Geographical diversity, residential segregation, and the vitality of African American Vernacular English and its speakers. *Transforming Anthropol* 18(1): 28–34.
12. Ryan EB (1975) *Why Do Low-Prestige Varieties Persist? Language and Social Psychology*, eds Giles H, St. Clair RN (University Park, Baltimore), pp 145–157.
13. Baugh J (2003) Linguistic profiling. *Black Linguistics: Language, Society, and Politics in Africa and the Americas*, eds Makoni C, Smitherman G, Ball AF, Spears AK (Routledge, New York), pp 155–168.
14. Rickford JR (1999) Language diversity and academic achievement in the education of African American students: An overview of the issues. *Making the Connection: Language and Academic Achievement Among African American Students*, eds Adger CT, Christian D, Taylor OL (Center for Applied Linguistics, Washington, DC), pp 1–29.
15. Fischer MJ, Massey DS (2004) The ecology of racial discrimination. *City Community* 3(3):221–241.
16. Lippi-Green R (2012) *English with an Accent: Language, Ideology, and Discrimination in the United States* (Routledge, London), 2nd Ed.
17. Terrell SL, Terrell F (1983) Effects of speaking Black English upon employment opportunities. *ASHA* 25(6):27–29.
18. Grogger J (2011) Speech patterns and racial wage inequality. *J Hum Resour* 46(1): 1–25.
19. Ludwig J, et al. (2013) Long-term neighborhood effects on low-income families: Evidence from Moving to Opportunity. *Am Econ Rev Pap Proc* 103(3):226–231.
20. Kessler RC, et al. (2014) Associations of housing mobility interventions for children in high-poverty neighborhoods with subsequent mental disorders during adolescence. *JAMA* 311(9):937–948.
21. Chetty R, Hendren N, Katz LF (2015) *The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.* (National Bureau of Economic Research, Cambridge, MA), NBER Working Paper 21156.
22. Lenneberg EH (1967) *Biological Foundations of Language* (Wiley, New York).
23. Trudgill P (1986) *Dialects in Contact* (Blackwell, Oxford).
24. Siegel J (2012) *Second Dialect Acquisition* (Cambridge Univ Press, Cambridge, U.K.).
25. Van Hofwegen J, Wolfram W (2010) Coming of age in African American English: A longitudinal study. *J Sociolinguist* 14(4):427–455.
26. Abrahams RD (1973) The advantages of black English. *Language, Society, and Education: A Profile of Black English*, ed DeStefano JS (Charles A. Jones, Worthington, OH), pp 96–106.
27. Labov W (1990) The intersection of sex and social class in the course of linguistic change. *Lang Var Change* 2(2):205–254.
28. Eckert P (1989) The whole woman: Sex and gender differences in variation. *Lang Var Change* 1(3):245–267.
29. Washington JA, Craig HK (1994) Dialectal forms during discourse of poor, urban, African American preschoolers. *J Speech Hear Res* 37(4):816–823.
30. Washington JA, Craig HK (1998) Socioeconomic status and gender influences on children's dialectal variations. *J Speech Lang Hear Res* 41(3):618–626.
31. Chambers JK (2003) *Sociolinguistic Theory: Linguistic Variation and its Social Significance* (Blackwell, Oxford), 2nd Ed.
32. Rickford JR, Price M (2013) Girlz II women: Age-grading, language change and stylistic variation. *J Sociolinguist* 17(2):143–179.
33. Baugh J (1996) Dimensions of a theory of econolinguistics. *Towards a Social Science of Language: Papers in Honor of William Labov*, eds Guy GR, Feagin C, Schiffrin D, Baugh J (John Benjamins, Amsterdam), pp 397–419.
34. Wolfram W, Van Hofwegen J (2016) On the utility of composite indices in longitudinal language study. *Panel Studies of Language Variation and Change*, eds Buchstaller I, Wagner S (Routledge, London).
35. Ludwig J (2012) The long-term results from the Moving to Opportunity residential mobility demonstration. *Cityscape (Washington, DC)* 14(2):1–28.
36. Sampson RJ, Sharkey P, Raudenbush SW (2008) Durable effects of concentrated disadvantage on verbal ability among African-American children. *Proc Natl Acad Sci USA* 105(3):845–852.
37. Sankoff D, Laberge S (1978) The linguistic market and the statistical explanation of variability. *Linguistic Variation: Models and Methods*, ed Sankoff D (Academic, New York), pp 239–250.
38. Fordham S, Ogbu JU (1986) Black students' school success: Coping with the "burden of 'acting white.'". *Urban Rev* 18(3):176–206.
39. Rickford JR, Sweetland J, Rickford AE, Grano T (2013) *African American, Creole, and Other Vernacular Englishes in Education: A Bibliographic Resource* (Routledge, New York).
40. Anderson E (1999) *Code of the Street: Decency, Violence, and the Moral Life of the Inner City* (W.W. Norton, New York).
41. Labov W (1973) The linguistic consequences of being a lame. *Lang Soc* 2(1):81–115.
42. Kneebone E, Nadeau C, Berube A (2011) *The Re-Emergence of Concentrated Poverty: Metropolitan Trends in the 2000s* (The Brookings Institution, Washington, DC), Metropolitan Opportunity Series Paper No 25.
43. Watson T (2009) Inequality and the measurement of residential segregation by income in American neighborhoods. *Rev Income Wealth* 55(3):820–844.
44. Reardon SF, Bischoff K (2011) Income inequality and income segregation. *AJS* 116(4): 1092–1153.

SUPPORTING INFORMATION APPENDIX

NEIGHBORHOOD EFFECTS ON USE OF AFRICAN-AMERICAN VERNACULAR ENGLISH

August 17, 2015

John R. Rickford[a,1], Greg J. Duncan[b], Lisa A. Gennetian[c], Ray Yun Gou[d], Rebecca Greene[a], Lawrence F. Katz[e,d], Ronald C. Kessler[f], Jeffrey R. Kling[g,d], Lisa Sanbonmatsu[d], Andres E. Sanchez-Ordoñez[h], Matthew Sciandra[d], Ewart Thomas[h], and Jens Ludwig[i,d]

[a] Department of Linguistics, Stanford University, Stanford, CA 94305; [b] School of Education, University of California, Irvine, CA 92697; [c] Institute of Human Development and Social Change, New York University, New York, NY 10003; [d] National Bureau of Economic Research, Cambridge, MA, 02138; [e] Department of Economics, Harvard University, Cambridge, MA 02138; [f] Department of Health Care Policy, Harvard Medical School, Boston, MA 02115; [g] Congressional Budget Office, Washington, DC 20515; [h] Department of Psychology, Stanford University, Stanford, CA 94305; and [i] Harris School of Public Policy, University of Chicago, Chicago, IL 60637.

[1] Corresponding Author: Stanford University, Department of Linguistics, Stanford, CA 94305; (650) 725-1565, rickford@stanford.edu

Contents

1. Moving to Opportunity (MTO) Demonstration Design and Study Sample

The U.S. Department of Housing and Urban Development's (HUD) Moving to Opportunity (MTO) demonstration was authorized by the U.S. Congress in the Housing and Community Development Act of 1992 (for more background on MTO, see refs. 1 and 2).

MTO enrolled families between 1994 and 1998 in five cities: Baltimore, Boston, Chicago, Los Angeles, and New York. To be eligible, families had to have at least one child under age 18 and live in public housing developments or project-based assisted housing in high-poverty areas, defined as a census tract in which more than 40 percent of the population was living in poverty in 1990. (Census tracts are geographic areas defined by the U.S. Census Bureau that typically contain 2,500 to 8,000 residents, with boundaries that were originally drawn to be "homogenous with respect to population characteristics, economic status, and living conditions" (3)). The Public Housing Authorities (PHAs) in each city conducted outreach to all eligible households through fliers, tenant associations, and other means, and all those interested received the opportunity to apply for this special program. At orientation meetings, families were told they would be randomly assigned to one of three groups if they applied. Those heads of households who remained interested after the briefing were screened for Section 8 housing voucher eligibility, completed the MTO baseline survey, and signed an enrollment agreement.

A total of 4,604 eligible households enrolled in MTO, representing around one-quarter of the population of MTO-eligible families (4, 5). Eligible applicants were randomly assigned to one of three groups:

1. The *MTO Low-Poverty Voucher* (LPV) group[*] received Section 8 rental assistance certificates or vouchers that they could use only in census tracts with 1990 poverty rates below 10 percent. In each city, a nonprofit organization under contract to the PHA provided mobility counseling to help LPV group families locate and lease suitable housing in a low-poverty area. Families had to initially sign a one year lease. After one year, families were able to use their voucher to renew their lease or relocate without any special MTO-imposed constraints on their moves. Families assigned to the LPV group were required to abide by all of the regular rules and requirements of the Section 8 certificate and voucher programs, including having a limited amount of time to search for housing and lease-up before they lost the rights to their subsidy, being required to contribute 30 percent of their adjusted income toward rent (the same rent requirement as in public housing), and prohibitions on rental assistance to households engaging in certain types of criminal activity.

2. The *MTO Traditional Voucher* (TRV) group[†] received regular Section 8 certificates or vouchers that were not subject to any special location restrictions under the MTO program. These families received no special mobility counseling in MTO beyond what is usually offered by local housing authorities to housing-voucher recipients.

---

[*] In previous discussions of the MTO results, the LPV group has also been referred to as the experimental group.

[†] As above, in other discussions of MTO, the TRV group has been referred to as the Section 8 (only) group.

3. The *MTO control group* received no certificates or vouchers through MTO, but continued to be eligible for project-based housing assistance and whatever other social programs and services to which the families would otherwise be entitled.

Assignment rates within the five MTO cities, or demonstration sites, were adjusted during the implementation of MTO to compensate for the fact that the lease-up rate for the two MTO voucher groups turned out to be higher than had been anticipated. The sample weights used in the quantitative analyses presented in the text and below adjust for differences among sites and over time in the random assignment ratio (see below and (1) for additional details about the sample weights).

2. Prior Sociolinguistic Literature Related to Neighborhood Effects on Dialect Use

In order to understand what social and psychological dimensions occasioned by the move to a low-poverty neighborhood via the MTO program might have affected the use of African-American Vernacular English (AAVE), it is helpful to consider what the previous sociolinguistics literature has to say about various features of the neighborhood social environment such as the socioeconomic status of neighborhood residents, members of a given study subject's specific social network, attitudes and identity, and their relation to dialect acquisition and use both for AAVE and other dialect varieties. These different neighborhood features are conceptually distinct in the sense that moving to a more affluent area might expose youth to more affluent and highly-educated neighborhood adults but may or may not change the set of people with whom they actually interact on a regular basis (network), and such a move might, but need not, change their sense of identity or attitudes towards dialect (AAVE) versus Standard American English (SAE) use. This previous literature relies on observational (non-experimental) data but is nonetheless informative about which aspects of MTO-induced moves may be relevant for causally affecting AAVE use among respondents, and guides our examination of candidate mediating measures.

2.1. Socioeconomic status

Socioeconomic status (SES), or class, as measured by occupational status (often in combination with education), income, and sometimes with additional factors like residency (number of rooms, rented vs. owned) and/or area, was one of the first social or neighborhood dimensions to be significantly correlated with sociolinguistic variation, in Labov's (6) ground-breaking study of English in the Lower East Side of New York City.

The most thorough study of sociolinguistic variation by SES among African-Americans is Wolfram's (7) study of Detroit, which revealed significant stratification by social class by head of household in relative frequencies of phonological and especially grammatical features of AAVE. His measures of AAVE included six of the language features used in this study (consonant cluster reduction, r-deletion, TH-stopping, multiple negation, 3rd singular –s absence, and copula absence). His measure of SES was an index of multiple measures: occupation, educational attainment and residency or living conditions. For *occupation*, Wolfram's index included 7 levels, ranging from professionals at the top (lawyers, doctors); then high school teachers and executive assistants; semi-professionals (auto salesmen, postal clerks); technicians (factory foremen, electricians) in the middle; skilled workmen (carpenters, policemen); semi-skilled workmen (gas station attendants, taxi-drivers); and unskilled workers (odd-job and heavy labor workers) at the bottom, with the unemployed (although not mentioned) presumably ranking with or even below them. *Education* levels ranged from graduate degrees through college (graduation; one year or more), high school (graduation vs. some), junior high, and less. *Residency* involved a combination of the number of rooms per occupied unit (from 10.5+ to 1-4.4) and the percentages of sound houses per census block, adjusted up or down by income levels per census tract.

Wolfram's quantitative data revealed that the lower an African-American speaker's SES or social class was, the higher the frequency of AAVE use was, as shown here for one phonological variable, consonant cluster reduction (p. 60), and one grammatical variable, copula and auxiliary *is/are* absence (p. 169):

SES	**Consonant Cluster Reduction**	**Copula Absence**
Lower Working	84.2%	56.9%
Upper Working	79.2%	37.3%
Lower Middle	65.9%	10.9%
Upper Middle	51.0%	4.7%

Although sociolinguistic studies of variation by social class elsewhere in the U.S. and the world (8) sometimes differ from Wolfram in the number and nature of the indices they employ beyond occupation, they generally find, as Milroy (9) notes, that "higher social class groups have relatively high linguistic scores (that is, they approximate closer to standardized varieties)" (p. 11). Moreover, upwardly mobile Working Class members (so classified if their current occupations are higher than their father's or their own prior occupation) "use fewer non-standard variants than the stable group (S) in the stable class of their origin" and also fewer than the stable members of the Lower Middle Class above them (ref. 10, p. 64, drawing on ref. 6). This is especially relevant to our MTO study given that MTO seems to have facilitated socioeconomic mobility for youth who were less than 13 years old at randomization (11).

However, many components of social class and mobility do not necessarily translate directly into different ways of talking, and it is important for both public policy and social science to better understand which aspects of neighborhoods influence dialect acquisition and use, and which social and psychological mediators are likely implicated in these "neighborhood effects." For example, with respect to *occupation*, a typical component of class status, it is known that the standard variety is expected or rewarded to a greater extent in some occupations (e.g. lawyers and teachers) than others (e.g. manual laborers). This phenomenon was first described for France as the "marché linguistique" (12) and for Canada and elsewhere as the "linguistic market" (13).

Parental *education* and *income*, other common components of class measurement, can be reflected in youths' use of the standard variety and educational attainment because schools in lower-poverty neighborhoods tend to have teachers who are more highly qualified and who, like parents themselves, are more likely to model and require standard speech and discourage use of the vernacular (14). For instance, in the village of Cane Walk, Guyana, while youth (under age 18) used less vernacular or creole English than seniors (over age 55) did in both the lower income Estate Class [EC=Working Class] and the somewhat higher income Non-Estate Class [NEC=Lower Middle Class], speakers in the intermediate 18-55 EC group were linguistically similar to the seniors, while the intermediate 18-55 NEC group were linguistically similar to the youth (15). The reason for this is that, among EC members, the seniors and intermediate age groups had both been limited to elementary education. By contrast, among NEC members, it was only the seniors who had, historically, been so limited, with *both* the intermediate and youngest NEC age groups enjoying access to secondary education and the opportunities and motivations for using or approximating to Standard English it provided.

2.2. Social networks

Milroy (9) was the first to show that social networks, especially dense and multiplex ones, have a powerful norm-enforcing effect, and that such networks tend to be especially prevalent in poor and working class neighborhoods, where they work to reinforce the vernacular. Connections between networks and AAVE use were first demonstrated by Labov (16), who showed that active participation in Harlem street culture through peer groups like the Jets and Cobras was associated with the most consistent use of the African-American vernacular. In other studies, AAVE use by African-Americans in Philadelphia was negatively correlated with the amount of contact these speakers had with whites (17); and Edwards (18) found a similar correlation in Detroit using a broader measure, the Vernacular Culture Index, of relative involvement with inner city neighborhood groups versus white friends. Milroy and Milroy (19) use these and other examples to forge a more general connection between social class and networks, noting inter alia that in inner-city Belfast, "speakers whose ties to the localized network are weakest…approximate least closely to vernacular norms" (p. 9).

Social network factors like these, plus the effect of schooling, may have led to the LPV group experiencing less peer group pressure to retain the vernacular and avoid "acting white" (20). As a respondent in one study (14) noted, "So we gotta have our survival mechanism within our community. And our language is it. It lets us know that we all in this thing together" (p. 184). Rickford (21) interviewed a youth in California who noted "Over at my school, if they – first time they catch you talkin' white, they'll never let it go. Even if you just quit talking like that, they'll never let it go!" (p. 192).

2.3. Attitudes and identity

It is also worth emphasizing that social class and network differences in speech result not just from differential access to or affinity for the standard, but also from differential ideologies about and affinity for the vernacular. Such ideologies are most evident in conflict models of class that see working class and middle class groups as having competing values and orientations. As Milroy and Milroy (19) note, citing refs. 22 and 23, among others, "in Belfast, in New York City, and (no doubt) elsewhere, young men are ridiculed by their peers if they use middle-class forms" (p. 4). Not only do vernacular forms mark working and lower class speakers' identity with and commitment to their geographical neighborhoods, communities of practice (24) or class positions if they embrace them, but they also provide warmth, familiarity and solidarity for such speakers. Ryan (25) notes that low prestige varieties persist around the world because of their value as markers of solidarity and group identity, and Carmichael (26) reports that the use of post-vocalic *r* by speakers who moved away from their Chalmette neighborhood in New Orleans after Hurricane Katrina for several years, was, whether they eventually relocated or not, significantly correlated with their awareness and positive orientation to Chalmatian r-lessness, which was absent from the neighborhood to which they moved. More specific to AAVE is Rickford and Rickford's (27) conclusion that speakers (two adults and a student) who were observed using AAVE in a primarily low income Black and Hispanic school setting in California did so (p. 222):

> because it is the language in which comfortable informal conversation takes place daily for them . . . because it resonated for them,…capturing a vital core of experience that had to be expressed *just so*;…because to have used Standard English might have marked the relationships between the participants as more formal or distant than the speaker wanted. For these individuals, not to have used Spoken Soul [=AAVE] might have meant they were not who or what or where they were and wanted to be.

Although we have no specific ethnographic report or direct measure of this, control group youth may have had more positive attitudes and orientations towards AAVE than youth in the LPV group, especially those who successfully moved to and remained in low-poverty areas.

In summary, the prior sociolinguistic literature provides reason to believe that MTO-induced changes in neighborhood socioeconomic or racial composition could change AAVE use among MTO participants, as could changes in the social networks, attitudes, and identities of these participants. Isolating the independent effects of these different candidate mediators is difficult given the MTO research design, which may change multiple mediators simultaneously. Below we present more information about which of these candidate mediators were changed, and by how much, as a result of assignment to the MTO LPV group.

3. Materials and Methods

3.1. Data sources

The HUD-sponsored evaluation of the MTO program included a baseline survey conducted just prior to randomization and an "interim MTO study," which gathered uniform data across all five sites and examined outcomes for MTO adults and youth at 4-7 years after random assignment (1, 28–31).

More than a decade after randomization and the baseline survey, our research team was engaged by HUD to follow up with MTO families to assess a variety of outcomes. These data were collected for our research team by the Survey Research Center (SRC) of University of Michigan's Institute for Social Research from June 2008 to April 2010, on average 12.7 years after randomization (range 10.0 to 15.3). The sample frame included one adult from each family in the LPV and control groups, as well as youth who were living in the baseline households and were ages 10-20 at the end of 2007. For budgetary reasons, we randomly sampled two-thirds of adults in the TRV group, who were also interviewed a few months later, on average, than the two other randomly assigned MTO groups. The data collection plan for our long-term follow-up study of MTO families was reviewed and approved by the federal Office of Management and Budget and the Institutional Review Boards at HUD, the National Bureau of Economic Research, the University of Chicago, the University of Michigan, and Northwestern University.

Target respondents were traced and, when contacted, offered $50 to complete a survey about health, economic conditions, and other outcomes, drawing mostly on questions from existing national studies. MTO adults in their surveys were also asked to report a full residential history between the time of random assignment and the survey, which we draw on to create residential histories for the youth in our study sample. (The full set of survey instruments for MTO adults and youth is available at www.mtoresearch.org.) Written informed consent was obtained before beginning interviews.

Trained interviewers using Computer-Assisted Personal Interviewing on laptop computers administered the survey primarily in the respondent's homes, with the session scheduled at the respondent's convenience. Interviewers were blinded to MTO group assignments. (As discussed further below, the socio-demographic characteristics of interviewers were similar for those that interviewed members of the different randomized MTO groups).

After 75-80% of the sample was interviewed in the initial phase of fieldwork, a probability subsample of 35% of remaining hard-to-reach cases were selected for further recruitment efforts (32). The latter interviews were up-weighted to adjust for the random sub-sampling of hard-to-reach cases.

To account for two-phase sampling, we calculated effective response rates (ERR) (32). Response rates were calculated using American Association of Public Opinion Research definition RR1w (33). Specifically, the response rate calculations account for the change over time in the MTO random assignment ratios as well as the two-phase survey sampling design of the long-term evaluation. The weights equal the product of the random assignment ratio weight, the survey and language sampling weights, and the two-phase sampling weight (equal to 1 for youth interviewed in Phase 1, equal to 1/0.35 for youth who were randomly selected for the Phase 2 survey sample, and equal to 0 for youth who were not randomly selected for the Phase 2 survey sample). The ERR is equal to the weighted number of interviews divided by the weighted survey sample frame total minus the weighted number of decedents.

Our focus here is primarily on MTO youth because our goal is to learn more about neighborhood effects on speech patterns, and previous sociolinguistic research suggests that speech patterns may be more sensitive to changes in social environments when experienced at relatively younger ages (34–37).[‡] Additionally, because our key speech measure of interest is use of African-American Vernacular English (AAVE), we focus on MTO youth who are non-Hispanic African-Americans.

As shown in Table S1, the ERR for our long-term survey of all MTO youth in the LPV and control groups was 90%, and equal to 91% for non-Hispanic African-Americans. For our target population the ERR was 91% for both the control group and the LPV group, and we cannot reject the null hypothesis that the ERRs are the same across randomized groups.

For budget reasons, our research team randomly selected n=980 non-Hispanic African-American youth from the LPV and control groups from whom to collect data on language. We focused data collection on these two of the three MTO groups in order to maximize the contrast in average neighborhood conditions between the groups of youth whose language patterns would be compared. Table S1 shows that 78% of non-Hispanic African-American youth who were eligible

‡ While there is some debate about whether there is a "critical period" for language acquisition (after which it is impossible to acquire a certain level of fluency), and what the age at which the critical period would end (arguments range from age 4 to 13), there seems to be much less disagreement about there being a "sensitive period" earlier in life during which language acquisition is easier.

for speech data collection consented to participate in the language component of the survey and have their responses audio-recorded. A total of 75% had their language successfully audio-recorded, and 71% of those had their audio recordings successfully transcribed by our sociolinguistics team and are included in our final analysis sample of n=629 youth. (We cannot reject the null hypotheses that the ERRs are the same in the LPV and control groups for the overall sample or for any of the following subsamples: selected for the language items on the survey, consented to audio recording, audio successfully recorded, or analyzed for this paper.)

The usual approach in sociolinguistic studies is to collect extensive speech data from a relatively small number of study subjects. But to take advantage of the key strength of MTO (its randomized experimental design), we collected speech data from as large a study sample as possible. Given the constraints on our data-collection budget and on the amount of time we could spend with respondents for the multi-purpose long-term MTO follow-up study, the speech samples we collected were relatively short (about 170 words per person on average). So relative to previous sociolinguistic research our language per person is lower, but our number of respondents is much greater.

The final speech measures that ISR collected from MTO respondents were selected by Dr. John Rickford and his team of sociolinguists. This selection was also guided by two pre-tests that ISR carried out with smaller samples of adults and youth with socio-demographic characteristics similar to those of the MTO study sample. As part of the final long-term youth survey, all of the survey respondents who were selected for language data collection were asked to respond to an open-ended question: “We are interested in the types of experiences people have had. What is the *happiest* moment in your life that you can remember?” ISR interviewers probed and prompted respondents (“Tell me more,” “What happened?” “What else happened?” “Where did that happen?” “When did this happen?” “What did you do?” “How did you feel?” “And?...”) with the goal of recording a narrative at least two minutes long (the software used to administer the survey interviews had a stopwatch function that interviewers started as the respondent began speaking). A small share of respondents (about 5 percent) was asked instead “What is the *scariest* moment in your life that you can remember?” Our original design planned for asking half the sample about their happiest moment and half their scariest moment, but we switched to asking all respondents about their happiest moment shortly into the survey fieldwork period when ISR reported that responding to the scariest moment prompt was upsetting to some respondents and made some interviewers uncomfortable. Replicating an item from the MTO interim surveys, all respondents were also asked: “Thank you for your participation in the Moving to Opportunity Study. Is there *anything else* that you would like to tell me about your neighborhood, or experiences, or any suggestions that you might have for improving housing programs?”

Both open-ended items were intended to attenuate the effects of formal observation and elicit casual speech, in which AAVE dialect use is likely to be more common compared to more formal speech (38–40). But, as sociolinguists readily admit, “methods…for overriding the constraints of the formal interview are only substitutes for the real thing, and give us only fragments of the vernacular [the style in which minimum attention is paid to speech]” (ref. 41, p. 115). We must therefore assume, as noted in the main paper, that the style recorded in this MTO study is essentially the style in which youth interact with a stranger.

After the open-ended "happiest moment" question, respondents in our speech-collection exercise were randomly assigned to one of two subgroups, one of which would be asked to do a read-aloud exercise and the other an elicited-imitation exercise. Analysis of the language samples from these later exercises is not included in this paper.

Transcription of these audio recordings, using the CLAN (Computerized Language Analysis) transcription program (42) was carried out by the research laboratory of Dr. Holly Craig of the University of Michigan. Transcription reliabilities are established by independent observers, who re-transcribe 10% of responses. Scoring reliability in Dr. Craig's language laboratory tends to be high, 90% or better (43). These transcriptions, as well as the original audio recordings, were then shared with Dr. Rickford's team at Stanford for quantitative, variationist analysis (defined as analysis of linguistic and social constraints on linguistic variation in language samples, e.g. recorded speech).

Figure S1 shows the cumulative probability function of the number of words in the speech samples of the 629 youth in our main analysis sample. The average word count was roughly 170 words; over 90% of the sample spoke 50 or more words, some three-quarters spoke 100 or more words, and about one-third spoke 200 or more words.[§]

3.2. Measures

In our main analyses the key dependent variable is the relative frequency of African-American Vernacular English (AAVE) vs. corresponding Standard American English (SAE) variants in speech recorded from each respondent. The features that we coded and tabulated for this paper in the speech samples from MTO youth included five grammatical and five phonological features widely used in analyses of sociolinguistic variation in AAVE (7, 21, 38–40, 44–50).

The grammatical features were:

- use of *ain't* rather than SAE negators like *aren't, isn't, hasn't* or *didn't* (as in AAVE *He ain't here* for SAE "He isn't here" or *She ain't do it* for "She didn't do it");
- multiple negation, with negation marked both on the auxiliary verb and on accompanying indefinite particles or pronouns like *never* or *nobody* (as in *He don't talk to nobody* for "He doesn't talk to anybody");
- absence of third singular present tense *–s* (as in *He walkØ* for "He walks");
- copula and auxiliary *is* or *are* absence (as in *He Ø happy* for "He is happy" or *They Ø walking* for "They are walking");
- *was*-leveling (as in *They was nice* for "They were nice").

The phonological features were:

- word or syllable final consonant cluster reduction (as in *fas'* for "fast");
- r-deletion after a vowel (as in *mothuh* for "mother");

[§] The word counts used to compute the average among analyzed youth speakers are for illustrative purposes only. In some cases the word counts were generated by the CLAN software and are accurate, but in other cases only a rough word count was manually generated to determine if the speech samples were long enough to include the speaker in analysis that was limited to speakers with greater than a certain number of words. Those manual word counts were stopped at 100 words, which is why Figure S1 shows a vertical line at 100 in the cumulative distribution function.

- DH-stopping (as in *dis* for “this”);
- TH-stopping (as in *wit* for “with” or *mout* for “mouth”);
- *ai* monophthongization, which involves producing a single, long *aa*, rather than a diphthong that moves from *a* to *i* (as in *raad* for “ride”).[**]

Table S2 shows how the 14,191 tokens in our analysis sample are distributed across the different grammatical and phonological features for which we coded the MTO youth speech data. About 89% of all tokens were related to phonological features. The distribution of tokens that were spoken and transcribed across different phonological and grammatical features was generally similar for youth who were randomly assigned to the control versus LPV group in the MTO experiment. Only one of the ten pair-wise comparisons is statistically significant (P<0.05): tokens that provide an opportunity for use of double negation were slightly more common for the control group (2.5% of all tokens) than for the LPV group (1.4%). For r-deletion and DH-stopping, the most common features, coding was generally capped at a maximum of 10 tokens per speaker for that feature.

As we note in the main paper, the norm in sociolinguistics is to collect very large amounts of language from a relatively modest number of people. However, our study is very different – to exploit the strength of the MTO randomized social experiment in facilitating estimation of causal neighborhood effects on speech, we collected speech data from a large number of people, which in turn necessarily required obtaining relatively shorter speech samples. This collection of moderate-length speech samples from a large study sample is one of the methodological innovations of our study, but this innovation means that the existing sociolinguistic literature does not provide us with the tools that we need to determine the validity and reliability of the type of AAVE measure that we analyze in this paper. To the extent to which any linguistic research addresses this question, some studies indicate that in general listeners are able to distinguish a speaker’s race even from relatively modest amounts of speech (51, 52).

Because the previous research on the exact statistical properties of our measures is so limited, we have carried out the original tests of the reliability and validity of our AAVE measures that we discuss below. To establish the *reliability* of our AAVE measure, we calculate the correlation in AAVE measures between study subjects from the same MTO household – we would expect these to be positively correlated if our AAVE measure is reliable given the influence of the home

[**] Note that our list includes five features (copula absence, third singular present –*s* absence, multiple negation, and *ain’t*) from the set of six that Renn and Terry (49) used to construct their subset measure of AAVE and to show that it was as effective as Washington and Craig’s (48) 34-feature Dialect Density Measure in assessing adolescent style shift. However, we should add that Renn and Terry’s *ain’t* measure tabulated only instances in which *ain’t* was used for SAE “isn’t.” Moreover, like Washington and Craig, who divided the number of occurrences or tokens of the vernacular variant by the number of utterances used by each speaker, Renn and Terry divided the number of vernacular variants by the number of utterances and (separately) by the number of words used by each speaker. By contrast, we followed the majority practice in AAVE studies and quantitative sociolinguistics by dividing the number of occurrences of variants of interest (in our case the AAVE variants) out of all the tokens of the variable (AAVE and SAE) in which those variants could have occurred. This practice follows the “accountability principle” adumbrated by Labov (80): “THAT ANY VARIABLE FORM (a member of a set of alternative ways of ‘saying the same thing’) SHOULD BE REPORTED WITH THE PROPORTIONS OF CASES IN WHICH THE FORM DID OCCUR IN THE RELEVANT ENVIRONMENT, COMPARED TO THE TOTAL NUMBER OF CASES IN WHICH IT MIGHT HAVE OCCURRED” (pp. 737-38, fn 20; upper and lower case as in original).

environment on speech. Our analysis sample includes 86 households with AAVE data available for more than one youth. Comparing AAVE rates for one pair of siblings from each of these households, the average correlation is 0.39 (*P*<0.001), repeating the selection 100 times to account for the random selection of two youth in households with three youth in our sample. We do not have comparable sibling correlations from other AAVE studies, however, Mazumder (53) reports sibling correlations for a variety of outcomes using a United States national sample. He finds sibling correlations of about 0.50 on economic outcomes and lower correlations for noneconomic outcomes such as illegal drug use (0.27) and self-esteem (0.25). As an additional check of the internal consistency of our measure, we used preliminary timing data that was available for about 75% of the 14,191 tokens analyzed for this paper to calculate the correlation between AAVE use in the first and second halves of each youth's tokens (split by the time at which the token occurred in the speech sample). The correlation between halves was 0.42 (*P*<0.0001, n=588) based on an average of about 9 tokens per youth per half. Some support for the *validity* of our AAVE measure comes from its relationship to socioeconomic characteristics and outcomes as discussed in the Supplementary Results below.

To measure neighborhood socio-demographic composition, HUD has tracked MTO respondents from baseline through the time of our long-term survey. In addition to HUD's own administrative records, other sources of address information available for MTO families include the U.S. Postal Service's National Change of Address system, local housing authorities, and interviews conducted with the youth and their parent (or other adult in the family) as part of the interim (2001) and long-term (2008-10) MTO evaluations. As part of the long-term survey, we collected the youth's current address and a detailed address history for the adults. We assign to MTO youth certain portions of the household adult's address history depending on whether the youth was still living with the adult and whether they had ever lived apart. We geo-coded the address histories of MTO youth over the 10- to 15-year study period, linked them to tract-level data from the 1990 and 2000 decennial censuses and the 2005-09 American Community Survey, and interpolated tract attributes for the years that fall between these Census Bureau data collections to measure tract characteristics at the time the youth was living at the given address.

We calculated census tract characteristics for the addresses at which participants were living at baseline and at various points after random assignment (RA): 1 year post-RA represents the address where LPV families moved using their MTO program vouchers; 5 years post-RA is close to address as of the interim evaluation; and 10 to 15 years post-RA is the address where the youth was living as of May 2008, just prior to the start of the fielding period for the long-term survey. We also calculated duration-weighted average tract characteristics for each participant's post-baseline address history, where the tract characteristics for each address are weighted by the share of the follow-up study period the family spent at each address. We examine census tract poverty rates, the neighborhood measure that MTO was explicitly designed to change for program participants, as well as a variety of other neighborhood characteristics, including the shares of tract residents who are members of racial or ethnic minority groups; the shares of adults who have completed high school, some college, and a college degree; the share of the civilian population who are employed overall; the share of workers in managerial or professional occupations; the share of tract residents receiving public assistance; median household income (in 2009 dollars); the share of families headed by single females; and the share of housing that is owner-occupied.

We also constructed a full history of schools attended for each youth by combining parent reports on the youths' schooling through the time of the interim follow-up survey (or kindergarten for youth who were not of school age when the family volunteered for the MTO program) with youth self-reports through the time of the long-term survey (or the highest grade attended for youth who were no longer in a primary or secondary school). We then matched the school histories to a variety of school-level socioeconomic and demographic characteristics data from two National Center for Education Statistics databases (the Common Core of Data for public schools and the Private School Universe Survey for private schools) and a school-level test score database (the National Longitudinal School-Level State Assessment Score Database). School characteristics were averaged across all schools/academic years for grades K-12 from random assignment through the youth's current school/grade as of the long-term survey interview (or through the most recent primary or secondary school grade for youth who were no longer enrolled).

3.3. Analytic strategy

In this section we first describe how we estimate the intention-to-treat (ITT) effects presented in Table 2 and Figure 1 in the main text and in the Supplementary Results section below. We then discuss how we can move beyond the pure design of the randomized MTO experiment to identify the effects of actually moving through MTO, known as the effects of treatment on the treated (TOT), which are presented below and which readers can infer from the ITT estimates that we do present by roughly doubling the ITT estimates (details below). Finally, we describe our instrumental variables (IV) approach for estimating the relationship between AAVE and specific neighborhood characteristics.

A. Intention-to-treat (ITT) effects

We begin with simple comparisons of the average AAVE use of youth assigned to the LPV and control groups, known as the ITT effect, which identifies the causal effect of offering families the services made available through the LPV treatment (as a reminder we focus in our analysis only on the LPV and control groups, since we did not collect speech samples from the TRV group). One advantage of the ITT estimate is that it fully capitalizes on the strength of MTO's randomized experimental design. The disadvantage of the ITT estimation is that it does not provide any information about the size of the effect on those who actually change neighborhoods, or about the relationship between specific neighborhood characteristics and people's life outcomes.

Our dataset is at the token-person level, so that we have multiple tokens per person in the analysis sample. Let Y represent our AAVE measure of interest, which is a 1/0 indicator for whether a given token is characterized by some AAVE feature. Let Z be an indicator for assignment to the LPV group. We calculate the ITT effect as π_{11} in Equation 1 using ordinary least squares (OLS), conditioning on a set of (pre-random assignment) baseline characteristics (X). These include indicators for each person's MTO demonstration site and survey measures of the socio-demographic characteristics of household members. Because the distribution of pre-program characteristics should be balanced across treatment groups due to random assignment, conditioning on these variables serves mainly to improve the precision of the treatment effect estimates. All estimates in this paper are computed using the sample weights described above. In practice, the coefficients from applying OLS to dichotomous dependent variables tend to be quite similar to the average marginal effects that come from probit or logit models (54), and indeed we find that average marginal effects calculated from probit and logit models are quite similar to the OLS estimates reported here (see below).

(Eq. S1) $Y = Z\pi_{11} + X\pi_{12} + e_1$

Unbiased estimation of the ITT effect requires several assumptions that we believe are likely to be met in the MTO application. The first assumption is that random assignment was carried out correctly, which we believe is the case based on a review of the randomization procedures employed by Abt Associates, which carried out random assignment on behalf of HUD, and given evidence presented in Tables 1 and S3 that the distribution of baseline characteristics is generally similar across randomly assigned MTO groups. A second assumption is that there is no selective attrition in our measurement of follow-up outcomes across randomized groups. We believe this assumption is likely to be met because the effective response rate is generally similar across randomly assigned groups (see Table S1). A third assumption for the standard interpretation of the ITT estimate is that the effect of MTO random assignment on a given family is independent of the treatment-assignment status of other families in the study sample, which Rubin called the "stable unit treatment value assumption" (SUTVA) and what has also been called the "no-interference" assumption (55). Sobel has raised concerns that SUTVA may not be met in the MTO application, as could occur if for example families assigned to the treatment group share information with controls that lead some control group families to also move to lower-poverty neighborhoods (56).

As we have indicated elsewhere (57), we think major violations of SUTVA are unlikely. Only around one-quarter of eligible public housing families applied to participate in MTO. Since around two-thirds of families that signed up for MTO were assigned to treatment, and fewer than three of five assigned to treatment moved with a MTO voucher, the share of public housing families who moved out of public housing through MTO is not more than 10% (that is, 25%*66%*60%). The actual share will be lower still given that not all public housing families were eligible for MTO (for example because they did not include children). Moreover the families that signed up for MTO seem to have been fairly socially isolated at baseline: among the household heads of the families of youth in our analysis sample, nearly 40% percent indicated on the baseline surveys when applying to MTO that they had no friends in the baseline neighborhood, and nearly 60% percent reported that they had no family in the neighborhood (see Table S3). For this reason we suspect that social interactions among MTO families were probably limited. Moreover the MTO program administrators tried to limit the clustering of LPV families in the same low-poverty neighborhoods, which maps of MTO relocation outcomes suggest was successful; see for example (1). Understanding more about the degree to which MTO families both within and across randomly assigned groups had important social interactions with one another remains a useful topic for future research.

B. Effects of treatment on the treated (TOT)

It is also possible to use data from the MTO experiment to estimate the effects of MTO moves on those who actually move through MTO, known as the effect of treatment on the treated (TOT) and calculated as the ITT effect divided by the treatment take-up rate (58). The standard error for the TOT effect is calculated the same way, by dividing the ITT standard error by the treatment take-up rate, such that the p-value for the ITT and TOT estimates will be the same under this method. The TOT estimates derived using this approach are very similar to those generated using two-stage least squares to estimate the effects of relocating through the LPV group, using the indicator for random assignment to the LPV group as an instrumental variable (59). Since 52% of the LPV group families in the sample of the non-Hispanic African-American youth for whom we have speech data relocated with a MTO voucher, the TOT effect will be (1/0.52) = 1.92 times as large as the ITT effect.

C. Estimating the relationship between AAVE and specific neighborhood conditions

Also of interest is understanding the relationship between specific neighborhood attributes and AAVE use. Let *W* represent candidate mediating mechanisms through which MTO might influence AAVE use, such as the poverty rate for the census tracts in which MTO families are residing. Let *X* represents the baseline control variables discussed above. The relationship between the candidate mediator(s) and AAVE (*Y*) is summarized by the parameter(s) π_{21} in Equation S2.

(Eq. S2) $$Y = W\pi_{21} + X\pi_{22} + e_2$$

For purposes of estimation of Eq. S2, we view any single variable used as an element of *W* to be a summary measure of neighborhood economic disadvantage. For example, when *W* is a scalar equal to the census tract poverty rate, we interpret π_{21} as the effect of moving to a neighborhood with a lower poverty rate and other aspects of neighborhood economic disadvantage that co-vary with tract poverty rates. We provide a similar interpretation for our single mediator models that examine two other key mediating measures in place of tract poverty—the share of the census tract population that is black and the share that is minority.

OLS estimation of Eq. S2 may be biased by endogenous residential choices. Families that wind up living in lower-poverty tracts may be systematically different from those who live in high-poverty areas in ways that are difficult to measure in a social science dataset and that may directly affect people's outcomes. This type of selection bias (or omitted variables bias) manifests itself as a correlation between *W* and e_2 in Eq. S2 and leads OLS estimates to mistakenly attribute to *W* the effects of unobserved measures in e_2.

Rather than use OLS to estimate Eq. S2, we use the random assignment of families to treatment and control conditions in MTO as an instrumental variable (IV) for *W* and estimate Eq. S2 using two-stage least squares (2SLS) and related IV estimators. One possible way to do this would be to use an indicator for assignment to the LPV treatment condition (*Z*) as an instrument for a single candidate mediating measure, *W*. The first stage equation (Eq. 2) is used to generate a predicted value of the mediating measure that is then substituted for the actual measure in the second stage. The second-stage equation (Eq. S3) estimates the relationship between neighborhood conditions and the outcomes isolating the experimentally-induced variation in the mediator.

(Eq. S3) $$W = Z\pi_{31} + X\pi_{32} + e_3$$

In general one potential drawback of this type of one-instrument, one-mediator ("just-identified") model is precision. But if the effect of the LPV treatment on the candidate mediator of interest varies across subgroups, then interacting treatment assignment with subgroup indicators can improve the precision of the second-stage 2SLS estimates by increasing the explanatory power of the first-stage equation (60). Previous MTO research has shown that there is substantial variation across the five MTO sites in the degree to which treatment assignment affects neighborhood poverty and other candidate mediators (30, 61). We follow Kling, Liebman, and Katz (30) and interact an indicator for assignment to the LPV group (Z) with five MTO demonstration site indicators (S), controlling for the main demonstration-site effects in the baseline covariates (X). The model essentially estimates a "dose-response" relationship, asking whether those groups that experience relatively larger changes in some candidate mediator as a result of treatment also experience larger changes in AAVE use. With multiple instruments we can then also try to test for the effects of multiple mediators (endogenous explanatory variables) included simultaneously in the same model, although statistical power can become a challenge.

We first present instrumental variables estimates calculated using 2SLS. To address concerns about weak instruments (low explanatory power in the first stage equation), we also estimate via limited information maximum likelihood (LIML) (62) as well as modified versions of LIML suggested by Fuller and others (63–65).

4. Supplementary Results

4.1. Descriptive characteristics of the MTO sample

Table S3 is an expanded version of Table 1 in the main text that displays descriptive characteristics for our study sample of MTO youth separately for the control group and the treatment (LPV) group. Almost all of the households that signed up for MTO were female-headed (not shown in the table), while over three-quarters of household heads were on welfare at baseline and less than 40% had completed high school. Although the overall MTO sample is nearly two-thirds African-American and nearly one-third Hispanic, as noted above our analysis sample is restricted to non-Hispanic African-Americans given our focus on use of AAVE.

More than 40% of households that applied had a household member victimized by a crime during the previous six months. Three-quarters of MTO families reported that getting away from gangs and drugs—that is, crime—was the first or second most important reason for enrolling in the program. More than half of the households said the first or second most important reason for signing up for MTO was so that their children could attend a better school.

Table S3 also confirms that random assignment appears to have been correctly carried out in MTO, given the balance across randomized MTO groups in the distributions of the observed baseline characteristics. An omnibus F-test fails to reject the null hypothesis that the set of baseline characteristics shown in the table are similar for the LPV versus the control group at the usual 5% cutoff (P=0.088).[††]

[††] We conduct an omnibus F-test of the differences between the treatment and control groups by estimating a seemingly unrelated regression where all of the characteristics listed in Table S3 are stacked as Y (outcome) variables and the only X variable is an indicator for treatment group status and a constant. This approach follows Jacob, Ludwig and Miller (81).

Table S4 shows that there is also similarity across randomized MTO groups with respect to the socio-demographic characteristics of the ISR interviewers who administered the MTO survey. The average age of ISR's survey interviewers was about 45, and over three-quarters of the interviewers were female. Some three-quarters had attended at least some college, about one-third had a college degree, and about 20% had an advanced degree. About 50% were African-American and another 35% were white, while nearly 20% were Hispanic (of any race).

In Table S5 we present the correlations between these interviewer characteristics and AAVE use (using a person-level share of AAVE tokens measure described in further detail below). While there is not a large volume of research in sociolinguistics about interviewer effects on speech production – the "observer's paradox" – the research that is available suggests the effects of the race of the interviewer on respondent language patterns can be significant,[‡‡] and the results in Table S5 appear consistent with those findings. AAVE use was positively correlated with having an African-American interviewer and negatively correlated with having a white or Hispanic interviewer or an interviewer who was neither African-American nor white (P<0.01 in all cases). However *because on average the characteristics of the ISR interviewers who surveyed youth in the LPV group were similar to those who surveyed youth in the control group*, the fact that interviewer characteristics are correlated with AAVE use should not bias our estimates for the difference in AAVE use rates between youth randomly assigned to the LPV vs. control groups.

As mentioned above, about half the families assigned to the MTO LPV group in our analysis sample for this paper were able to lease-up and relocate using an MTO voucher (the MTO "compliance rate"). The MTO compliance rate is less than 100% for a variety of reasons including that families were given only a limited amount of time to search for a new unit, many housing units were not affordable under voucher program rules, and some landlords may have discriminated against voucher holders. The voucher use rate in MTO is in line with what other studies have found—equal to 65% in the Experimental Housing Allowance Program (ref. 66, p. 146), and around 20% in the Gautreaux mobility program in Chicago (ref. 67, p. 67). Families within the MTO treatment groups who comply (use an MTO voucher to move) are younger, more dissatisfied with their original neighborhoods, and have fewer children than the noncompliers (for details see refs. 68and 69).

4.2. MTO effects on neighborhood conditions, mobility, and school environment

Tables S6 and S7 display estimates of MTO effects on neighborhood conditions, mobility, and the school environment that could be candidate mediators for the effect of MTO moves on AAVE use. Table S6 shows how the characteristics of the neighborhoods of LPV and control group youth changed over time. Each set of columns shows the control mean and ITT effect of the LPV treatment at a given address. Panel A shows that at baseline the average control group youth in our sample was living in a census tract that was about 59% poor, or about 3.6 standard deviations above the national average as calculated from the national tract-poverty distribution

‡‡ For example, several studies have found significant effects of interviewer race on use of one or more AAVE features by African American interviewees (40, 44–47). However, Cukor-Avila and Bailey (82) find no significant effects of interviewer race on language use among a sample of African-American respondents, albeit with a white interviewer who had considerably closer familiarity with the community, over several years, than most outside interviewers do, and certainly more so than the interviewers in our MTO study, who were strangers.

from the 2000 decennial census. The first set of columns in the table shows that there are no differences across randomly assigned groups in the characteristics of the baseline neighborhoods, as we would expect with random assignment.

The second set of columns of Table S6 (labeled 1 Year Post-RA) shows that MTO moves accomplished their goal of helping families move into lower-poverty neighborhoods. One year after random assignment, the average control group youth lived in a census tract with a 56% poverty rate, while the ITT effect on the poverty rate for youth in the LPV group was 21 percentage points, yielding a poverty rate for LPV youth of about 35% ($P<0.05$). As a way to think about the magnitude of these estimates, this change is about 1.7 standard deviations within the national tract-poverty distribution, or about 1.3 standard deviations within the MTO control group's tract-poverty distribution. While Table S6 focuses on presenting ITT effects for parsimony, as discussed above the TOT effects will be about 1.9 times as large as the ITT effects for the LPV group. Or put differently, those youth whose families relocated through MTO with an LPV voucher experience a decline in tract poverty rates measured 1 year after baseline equal to (1.9 * 21) = 41 percentage points.

As Table S6 and Figure S2 show, the difference across MTO groups in census tract poverty rates narrowed over time, due largely to declines in the tract poverty rates experienced by the control group – which went from 59% at baseline to 32% at the time of the MTO long-term follow-up (10 to 15 years later in May 2008). The control group trend is due more to control families moving into lower-poverty neighborhoods over time on their own, as opposed to control families living in neighborhoods that are gentrifying around them.[§§] Ten to fifteen years after baseline, the ITT effect of the LPV group on tract poverty rates equaled about 4 percentage points ($P<0.05$). This impact equals about 0.35 standard deviations in the national census-tract poverty distribution in the 2000 census (or about 0.26 standard deviations in the MTO control group distribution).

The duration-weighted average tract poverty rate for all addresses between random assignment and the start of the long-term survey fielding period (10-15 years) was around 43% for the control group, with an ITT effect of 11.3 percentage points for the LPV group ($P<0.05$ in both cases). This means that the TOT effect on those youth in the LPV group whose families actually relocated with an MTO voucher was about (1.9*11.3)=22 percentage points, or put differently, the average census tract in which these youth lived over the study period had a tract poverty rate of about (43-22)=21 percent. This figure is not so different from the average poverty rate overall in the five MTO cities (70). The standard "dissimilarity index" used in studies of segregation is defined as the share of people within a group that would need to relocate across neighborhoods in order for each neighborhood to have the same share poor as the overall city. So the youth whose families move as part of the LPV treatment in MTO essentially wind up in neighborhoods that correspond to the dissimilarity index definition of perfect poverty integration. We also find that using a duration-weighted measure of a broader set of indicators for concentrated

§§ We test this assertion by reproducing the estimates shown Table S6 in these supplementary materials measuring the share poor in each tract using only data from the 2000 decennial census, rather than interpolating each census tract's poverty rate at the time the MTO family was actually living in the tract. The estimates using 2000 tract poverty rates are fairly similar to those shown in Table S6, suggesting that most of the change in the control group's tract poverty rate over time occurs because control families are moving into lower-poverty areas, rather than because the control group is living in census tracts that are becoming less poor around them.

disadvantage based on Sampson, Sharkey, and Raudenbush (71) yields results that are qualitatively similar to the results for duration-weighted tract poverty – see panel B.

MTO moves also led LPV group families to live in census tracts that had slightly lower minority shares compared with controls, even if the tracts were still mostly minority (for further discussion of this and related issues, see refs. 57 and 72–74). We focus on the share of the census tract that is African-American given our focus in this paper on AAVE use and our target study sample of non-Hispanic African-American youth. Panel C shows that one year after random assignment the average tract share black was 71 percent for control families, with ITT effects of 5 percentage points for the LPV group ($P<0.05$). The LPV ITT effect is equal to about 0.21 standard deviations in the national tract distribution, or about 0.16 standard deviations in the control group's tract distribution. The ITT effect on duration-weighted tract share black was 4 percentage points ($P<0.05$, equal to 0.18 standard deviations in the national tract distribution, or 0.16 standard deviations of the MTO control group's distribution). Results for tract share minority (panel D) are similar to those for share black.

Additionally, MTO moves exposed the LPV group to neighborhoods with more educated residents. Youth assigned to the LPV group lived in neighborhoods with greater shares of residents who completed high school, who attended some college, and who have a college degree than youth in the control group. The duration-weighted ITTs on share with at least some college and share college graduates are about 25% and 40% of their respective control means (see panel E). Individual and household educational attainment shows a strong negative correlation with AAVE use (results presented later in this SI Appendix), which suggests that MTO may have had a notable effect on the AAVE environment in the neighborhoods of the LPV group.

The pattern of results for other census tract characteristics presented in panel F is similar to those for poverty and minority composition—large initial effects that faded over time but are significant when averaged over the entire follow-up period. LPV group youth lived in tracts with a higher proportion of employed residents and lower proportions of welfare recipients and families headed by single mothers. The ITT effect on median household income (over $10,000) is nearly 40% of the control mean value of about $26,000.

Panel A of Table S7 shows that MTO families assigned to the LPV group wound up making about 0.7 extra moves over the course of the study period compared to the control group, which moved on average 2.6 times over the 10- to 15-year period. The panel also shows that, at the start of the fielding period for the MTO long-term survey, LPV families were more likely than control group families to live a greater distance away from their baseline address.

Additionally, panel B shows that the MTO LPV treatment had a statistically significant effect on at least some measures of perceived neighborhood safety but no statistically significant effect on how satisfied youth were with their neighborhoods. Because each of the individual measures in panel B is likely to be a noisy measure of the underlying general concept of neighborhood safety and satisfaction, we also combine the measures by first standardizing each measure (based on the MTO control group's mean and standard deviation) and then averaging them (and re-standardizing the average) to create an index. The table shows that LPV youth experienced a 0.22 SD improvement on this combined neighborhood safety and satisfaction index.

In addition to changing the neighborhood context of youth, panel C also shows that MTO produced modest changes on their school environments. Youth in the LPV group attended schools with lower shares of free lunch eligible, African-American, and Hispanic students than the schools attended by control group youth. The LPV group's schools rank slightly higher than those of controls in terms of student performance on statewide exams (21^{st} versus 17^{th} percentile rank) but do not differ on size or pupil-teacher ratios. As above, we combine the measures into a standardized index and show that LPV group youth attended schools that were 0.40 standard deviations less disadvantaged than those attended by control group youth.

In panels D and E of Table S7, we generally do not detect significant effects on most of our measures of youth or adult social networks (even when we aggregate the measures to standardized indices), although we do observe a marginally significant increase in the likelihood that adults report having a close friend who graduated from college (P=0.051). Panel F suggests that the LPV treatment may have led to less engagement in church activities and to youth spending more time at shops.

In summary, our findings are generally consistent with the sociolinguistic literature that suggests a link between neighborhood dimensions and dialect acquisition. First, the duration-weighted neighborhood characteristic measures covering the entire post-random assignment period (Table S6) show that the neighborhoods in which LPV group youth lived are significantly less disadvantaged than the neighborhoods of control group youth. Second, although the differences between the LPV and control groups in our "social network" indices are not significant (Table S7, panel D), those in the school-related mediators are significant and consistent with our expectations (Table S7, panel C). For example, schools attended by LPV youth had significantly fewer black and minority students (i.e., AAVE speakers) than those attended by control group youth. Thus the former group likely experienced less peer group pressures to "diss" the standard (75) and used the standard varieties more often.

The fact that MTO changed so many aspects of the social and physical environment of families at once presents a challenge for isolating which specific mediators are most important for affecting the youth's AAVE use. We have carried out some exploratory analyses using random assignment interacted with MTO demonstration site as instruments for specific measures of tract attributes, following Kling, Liebman, and Katz (30). However with the sample size we have here we are unable to statistically isolate the effects of different tract attributes such as tract poverty versus tract share black – see below for further discussion.

4.3. MTO effects on AAVE use

The remaining tables in this SI Appendix explore how MTO affected the use of AAVE among the youth in our analysis sample. Tables S8 through S11 present additional ITT effects to test the sensitivity of the main findings in Table 2 and to further explore how the effects of MTO vary by subgroup and language feature (each of these tables begins with a panel A that replicates our main finding from Table 2). Tables S12 and S13 present TOT and IV estimates, respectively, and Tables S14 and S15 present control-group only analysis to test the validity of our AAVE measure and help frame the size of the MTO effect on AAVE use.

Table S8 presents sensitivity analyses for our main findings about how MTO changes AAVE use among non-Hispanic African-American youth. As mentioned above, panel A replicates our main findings from Table 2. In Panel B, instead of using each of the 14,191 tokens in our sample as the unit of observation, we collapse the data to the person level (n=629) and calculate the average share of tokens spoken by each person that were AAVE. The estimate is very similar to our main result that was analyzed at the person-token level (3.3 vs. 2.8 percentage points). In panel C we use probit and logit regressions to determine whether the ordinary least squares estimates of the binary AAVE vs. SAE outcome that we utilize in the token-level analysis are similar to those from models designed specifically for binary outcomes, and we find that the main result is nearly identical to the marginal effects from the probit and logit models. Furthermore, panels D and E show that our results are qualitatively similar if we restrict our analysis sample to youth whose audio-recorded transcripts included at least 50 words or 5 of 10 tokens (which excludes about 10% to 20% of the sample) or if we exclude all baseline covariates (aside from indicators for MTO demonstration site).

The results are also generally similar when we use different variations of race and ethnicity when defining our analysis sample (panel F). Given our focus on AAVE use, our main analyses focus on non-Hispanic African-American youth using the youth's own self-report about their race/ethnicity. The first row of panel F shows that the results are nearly identical if we expand the analysis sample slightly to include all African-American youth, including the additional 30 or so youth who self-report as Hispanic African-Americans. Defining our youth analysis sample using the MTO household head's race/ethnicity leads to point estimates that are slightly smaller in absolute value compared to our main results that use the race/ethnicity of the MTO youth.

In Table S9 we further explore the differential effects by age and gender that are presented in Table 2 and also take into account duration of exposure. Table 2 shows that LPV vs. control differences in AAVE use are more pronounced for female youth and for youth who are relatively older adolescents at the time of our in-person data collection. In Table S9, panels B and C replicate the respective panels from Table 2, but here we add rows that present a test of the difference of the effects of MTO by subgroup. We cannot reject the null hypothesis that the effects for male and female youth are the same, but we do reject that null hypothesis when comparing effects by age group (as of December 2008, or roughly the average age of youth when they were interviewed). Panel D of Table S9 splits the sample into age groups based on age at the time the youth's family joined the MTO program (at baseline) and, similar to panel C, shows that MTO effects on AAVE use are concentrated among relatively older youth. Panels E and F explore the relationship between gender and age to determine whether MTO effects are strongest among any combinations thereof, where age groups are based on contemporary age in panel E and on baseline age in panel F. While older females (whether we use contemporary or baseline age) in the LPV group used less AAVE than their control group counterparts, the ITT effects are only statistically distinguishable from the effects on younger males. The MTO effect on AAVE use among older males is close to marginal significance for contemporary age (P=0.113) and is significant for baseline age (P=0.016), and the effects on older males are statistically distinguishable from those on younger males regardless of age definition. But otherwise MTO effects by gender and age are neither statistically significant on their own nor statistically distinguishable from one another.

Panel G splits the sample into baseline and contemporary age groups. LPV group youth who were older at baseline (ages 6-11) and older at interview (ages 17-21) used less AAVE than their control group counterparts, while the ITT effect for youth who were under age 17 at interview, regardless of their baseline age, was not statistically significant. The ITT effect for the older youth at baseline and at interview is statistically distinguishable from the effect for the younger youth at baseline and at interview. And given the research evidence cited in the main section of the paper—that African-American youth seem to reduce their AAVE use in late adolescence/young adulthood due to workplace and other pressures (10, 50, 76–78)—we might argue that LPV neighborhoods offer more supportive conditions than control neighborhoods do for the late adolescence decline in terms of schools, networks and opportunities for jobs in which SAE is required or preferred.[***] But we cannot reject the null hypothesis that the effects by age at interview holding baseline age constant and the effects by baseline age holding age at interview constant are the same.

Panels H and I explore the relationship of age at the time of data collection and at baseline with the duration of exposure to neighborhoods of varying poverty levels. Because randomization in the MTO study occurred over a five-year period (1994-98), we are able to divide our language sample into youth whose families entered the MTO program earlier vs. later to determine whether a longer period of exposure to lower-poverty neighborhoods has a more pronounced effect on LPV group youth. The pattern of results suggests that the combination of age and duration of exposure yields more pronounced MTO effects in some cases, but in most cases we cannot reject the null hypotheses that the results across subgroups are the same. Our inability to disentangle the effects of age at baseline, age at the time of data collection, and duration of exposure stems from how they are all confounded (in the same way that in demographic applications, age, period and cohort effects are intrinsically confounded). For example youth who were relatively younger at the time of MTO random assignment experienced a relatively larger "dose" of reduced neighborhood poverty during early childhood. But they were also relatively younger at the time of the MTO long-term follow-up survey that assessed AAVE use such that any effect of age of exposure to lower-poverty neighborhoods is confounded with any effect of age at time of measurement on AAVE use (which is relevant because sociolinguistic research suggests that AAVE use rates overall tend to change over the course of adolescence). Similarly, if we compare two youth of the same age at the time of random assignment to determine how duration of exposure to lower-poverty neighborhoods affects AAVE use, the youth with the relatively longer exposure will be relatively older at the time of the long-term follow-up.

Table S10 shows that our results on the relatively more pronounced effects of MTO on AAVE use for those whose parents cared most about drugs, gangs or school quality at baseline appear to be qualitatively similar regardless of how we define this subgroup. Panel B of Table S10 replicates the subgroup analysis shown in the bottom panel of Table 2 that compares youth whose parents say that *either* drugs/gangs *or* school quality was the single most important reason for signing up for MTO to youth whose parents provided another reason. Panels C through F show how the results change if we instead focus on youth whose parents listed drugs/gangs as the first *or* second most important reason for signing up for MTO (versus all other reasons) or listed schools as the first *or* most important reason for signing up for MTO (versus all others).

[***] See Sankoff and Laberge (13) for rankings of Montreal speakers according to "the relative importance of the legitimized language in the socioeconomic life of the speaker" (p. 241). And note that Macaulay (8) found that the differences between 10-year-olds, 15–year-olds, and adults in the non-standard pronunciation of [t] with a glottal stop [ʔ] in Glasgow were minimal among trades-workers (a range of about 10%), but much bigger among working class clerks (about 70%) and middle class managers (about 47%, with the 15-year-olds significantly closer to the adult norm) (p. 47). Both studies are discussed by Chambers (10).

Adding credibility to these subgroup analyses is the fact that, within the LPV group, MTO voucher use rates, duration-weighted poverty rates, and the number of months youth lived in low-poverty areas are all similar across the reason for moving subgroups analyzed in Tables 2 and S10.

Table S11 explores how the results change if we limit our analysis to certain subsets of the ten language features for which we coded tokens. Panel B splits the tokens into broad feature type (grammatical vs. phonological) and shows that the size of the MTO effect on grammatical tokens as a group is very similar to what we see for phonological tokens. However the effect size for the latter is marginally significant, while the effect size for the former is not. This is to be expected given that phonological tokens account for nearly nine of every 10 tokens in our analysis sample.

Interestingly enough, however, when we look at the individual features in panels C and D, we find that the ITT effect is stronger for the grammatical rather than phonological features for which the combined feature ITT effect is stronger. To begin with, four of the five grammatical features show negative coefficients, that is, reduced vernacular usage, as a move to a lower poverty neighborhood would predict, while only three of the phonological features do so. More importantly, two of the negative coefficients for the grammatical features are significant (multiple negation solidly so, at P=0.008, and *was*-leveling marginally so, at P=0.066), while none of the ITT effects on the phonological features reaches significance (the closest candidate is r-deletion, the feature with the greatest number of tokens, at P=0.138). These findings presumably stem from the fact that grammatical vernacular features are, in general, much more sharply stratified by class than phonological features are (7) and, with the exception of shibboleths like *aks* for "ask", more pointedly the focus of overt social stereotyping, comment, and correction.[†††] We also separately analyze grammatical and phonological tokens by gender (panel E) and age (panel G), with results paralleling the corresponding subgroup effects for all tokens in Tables 2 and S9. Small sample sizes prevent us from analyzing the individual grammatical features by subgroup, while analysis of individual phonological features by gender (panel F) and age (panel H) produces largely insignificant effects (in line with the results in panel D). However, the low number of grammatical feature tokens and youth in Table S11 should not lead us to under-estimate the social significance of those features—even a single occurrence of one of these features, standard or non-standard, is subject to evaluation by one's peers, parents, teachers or employers (ref. 21, pp. 190-91).

Panel I of Table S11 estimates the effect of MTO on AAVE use controlling for the linguistic environment. The tradition within sociolinguistics, a field that usually involves analysis of observational or non-experimental data, is to control for the features of the language that speakers produce. For example, people may be more likely to use the AAVE variant *ain't* in place of "isn't" than they are to substitute *ain't* for either "hasn't" or "didn't", such that controlling for whether a given token was an opportunity to use either the AAVE or SAE variant of "isn't" versus an opportunity to use either the AAVE or SAE variant of "hasn't" helps better understand the variation in AAVE use within the non-experimental study sample. We do not control for these language features in our main analysis because the choice to produce a token that provides an opportunity to say *ain't* for "isn't" rather than a token that provides an

[†††] On the other hand, quantitative or gradiently stratified features like consonant cluster reduction (e.g. *fas*' for "fast" and *han'* for "hand") would probably have to rise about the norms for middle class use (about 50%, or above 39% for r-deletion) for a listener to "notice" that a speaker is "talking black" (7).

opportunity to say *ain't* for "hasn't" is an outcome—a decision that occurs after randomization and may itself be affected by treatment assignment.

However, understanding the degree to which MTO impacts on people's choices more broadly mediate MTO impacts on AAVE use is itself of interest. Panel I presents the results of models that include a set of controls that vary at the token-person level and control for the ten very broad language features in Table S2 (for example an indicator for whether the token provides them with the opportunity to say *ain't* instead of SAE variants "isn't", "hasn't", or "didn't", or a separate indicator for whether the token is a chance to use a double negation). Controlling for these broad language features, the MTO effect is quite similar to our main estimate.

Table S12 replicates the ITT effects from Table 2 in the main text alongside the TOT effects to show the effects of MTO on youth whose families actually moved using their MTO voucher. As a benchmark for judging the size of the TOT effect, we present our estimate for the control complier mean (CCM), the average outcome of youth in the control group whose families would have used a voucher if assigned to the treatment group, which can take on negative values because of sampling variability. The CCM is calculated as the average value of the outcome for the treatment-group compliers minus the TOT estimate (79). As mentioned above, the size of the TOT effect is about 1.92 times the size of the ITT effect, so we see that, among all non-Hispanic African-American youth, moving with an MTO voucher reduced AAVE use by 5.7 percentage points, or 12% of the CCM. The TOT effect is even larger as a percentage of the CCM (roughly 20%) for female youth, older youth, and youth from families for whom concerns about safety or schools were the motivation for wanting to move via MTO.

Table S13 presents the results of our analyses that use interactions of random assignment and MTO demonstration site as instruments for specific measures of neighborhood attributes. The models essentially estimate a "dose-response" relationship to determine whether the groups that experience relatively larger changes in some candidate mediator as a result of the MTO treatment also experience larger changes in AAVE. First we examine the relationship between AAVE and census tract share poor, share black, or share minority using two-stage least squares (2SLS), and then we test the sensitivity of our estimates to using limited information maximum likelihood (LIML) and using modified versions of LIML suggested by Fuller designed to decrease the variability of LIML estimators in small samples and perform better with weak instruments. Panel A of Table S13 estimates the relationship between AAVE and each neighborhood characteristic separately. Our first stage statistics suggest that our instruments are strongest for share poor, the neighborhood characteristics targeted by the MTO program. The relationship between poverty and AAVE use is positive (i.e. higher neighborhood poverty is associated with more AAVE) and marginally statistically significant (P=0.096). The estimated relationships for share black and share minority are also positive but not statistically significant. The models in panel B instrument for neighborhood poverty and racial composition simultaneously, but because of the size of our sample, we are unable to statistically isolate the effects of the different tract attributes.

Table S14 offers some evidence of the validity of our AAVE measure. Panel A shows that control group youth in our main sample of non-Hispanic African-Americans use more AAVE than the Hispanic (non-African American) youth from whom we also collected language samples (but who are not included in the main analysis). Likewise panel B shows that AAVE use is generally higher among control group youth from households that were more disadvantaged at baseline, e.g. with household income below the median value or where the sample adult had a lower level of education achievement.

Finally, Table S15 shows the correlation between AAVE use and different behavioral outcomes, which we use to estimate the potential implications of these MTO effects on AAVE use for other life outcomes for these youth. Since we have no source of exogenous identifying variation, these correlations should be interpreted as upper bounds rather than causal relationships; we carry out this exercise mainly to help illustrate the potential order-of-magnitude of the implied effects on long-term life chances from the MTO effects on AAVE use.

The first set of columns of Table S15 show the results of a simple bivariate regression with the person-level share of AAVE tokens measure for each person as the key explanatory variable and different outcome variables as described in the row labels, the upper panel examines correlations between youth outcomes and youth AAVE use, and the lower panel examines correlations between adult outcomes and adult AAVE use. The data on adult AAVE use was collected in a manner similar to that described above for youth. The second set of columns adds in age, MTO demonstration site, and gender as additional explanatory variables. Because in principle the relationship between AAVE could be non-linear, we also tried different functional forms for our regression, but we do not have enough data to be able to detect any non-linearities.

The result for adult earnings implies that going from 0% AAVE tokens to 100% AAVE tokens is correlated with a difference in annual earnings (in 2009 dollars) of about $12,300 (second set of columns).[‡‡‡] This implies that the MTO ITT effect on AAVE use of about 3 percentage points would be correlated with a difference in annual earnings of roughly $350 (0.03 x $12,300 = $369). Since the TOT effect is about twice as large, this implies that an upper bound for the change in annual earnings that might arise from the MTO effect on AAVE for those who move through MTO might be as large as about $700 per year.[§§§]

We can then also calculate a rough approximation of the net present value of this annual earnings gain by assuming that each youth works every year between the ages of 18 and 65. Using a 3%

‡‡‡ This regression of earnings against AAVE use does not include any adults who were incarcerated at the time of our in-person data collection. However we suspect the share of MTO adults who are incarcerated would be low for two reasons. First, the vast majority of MTO households are headed by women, for whom incarceration rates are in general much, much lower than for males. (About 7% of all people in prison are female—see Table 4 of www.bjs.gov/content/pub/pdf/p13.pdf). Second, the average age of the MTO adults at the time our data collection period began was about 45. One of the most consistent findings in criminology is that rates of criminal activity start increasing during adolescence, peak around age 18 or so (depending on the exact type of crime being examined), and then steadily decline as people age – known as the "age-crime curve." The rates of criminal behavior and hence incarceration for women in their mid-40s should be fairly low.

§§§ We use adult earnings instead of the youth's own earnings for this calculation because the youth were generally too young to evaluate their adult earnings at the time of our follow-up study. Sanbonmatsu et al. (2) estimated the LPV ITT effect on individual annual earnings for youth as +$327, with a standard error of $583, which together imply a 95% confidence interval that ranges from about -$800 to +$1,500 (table ES-8). This estimate is not precise enough to be able to detect a change in annual earnings of roughly $700 (as calculated using the results in Table S15). However, more recent data show that MTO had positive impacts on the adult earnings of MTO youth who were under 13 years old at randomization (11).

discount rate, the net present value of the implied change in annual earnings is equal to $18,200. If we use a 5% discount rate instead, the net present value for the implied change in lifetime earnings is around $13,300. By way of comparison, the average annual income for MTO households in the control group at the time of our long-term follow-up survey was about $20,000, and thus neighborhoods could increase lifetime earnings by as much as about 3-4% of lifetime income. Although the sociolinguistics literature to date has not established a threshold for AAVE use at which a speaker would be identified as "talking black," the above-referenced association of AAVE use with MTO adult earnings at least suggests that the AAVE use within the range of variation we see in our data is potentially noticeable to listeners.

Supporting References

1. Orr L, et al. (2003) *Moving to Opportunity for Fair Housing Demonstration Program: Interim Impacts Evaluation* (U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC).

2. Sanbonmatsu L, et al. (2011) *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation* (Washington, DC) Available at: www.huduser.org/publications/pdf/MTOFHD_fullreport_v2.pdf.

3. Division USCBG (2000) *Census Tracts and Block Numbering Areas* Available at: http://www.census.gov/geo/www/cen_tract.html (accessed December 1, 2011).

4. Goering J, et al. (1999) *Moving to Opportunity for fair housing demonstration program: Current status and initial findings* (U.S. Department of Housing and Urban Development, Washington, DC).

5. Goering J, Feins JD, Richardson T (2003) What have we learned about housing mobility and poverty deconcentration? *What Have We Learned about Housing Mobility and Poverty Deconcentration? In Choosing a Better Life? Evaluating the Moving to Opportunity Social Experiment*, eds Goering J, Feins J (The Urban Institute Press), pp 3–36.

6. Labov W (1966) *The Social Stratification of English in New York City* (Center for Applied Linguistics, Washington, DC).

7. Wolfram WA (1969) *A sociolinguistic description of Detroit Negro Speech* (Center for Applied Linguistics, Washington, DC).

8. Macaulay RKS (1977) *Language, Social Class and Education: A Glasgow Study* (Edinburgh University Press, Edinburgh).

9. Milroy L (1987) *Language and Social Networks* (Blackwell, Oxford and New York). 2nd Ed.

10. Chambers JK (2003) *Sociolinguistic Theory: Linguistic Variation and its Social Significance* (Blackwell, Oxford). 2nd Ed.

11. Chetty R, Hendren N, Katz LF (2015) *The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.* (National Bureau of Economic Research Working Paper 21156. Cambridge, MA).

12. Bourdieu P, Boltanski L (1975) Le fétichisme de la langue. *Actes Rech Sci Soc* 1(4):2–32. Available at: http://www.persee.fr/web/revues/home/prescript/article/arss_0335-5322_1975_num_1_4_3417.

13. Sankoff D, Laberge S (1978) The Linguistic Market and the Statistical Explanation of Variability. *Linguistic Variation: Models and Methods*, ed Sankoff D (Academic Press, New York), pp 239–250.

14. Lippi-Green R (2012) *English with an accent: Language, Ideology, and Discrimination in the United States* (Routledge, London and New York). 2nd Ed.

15. Rickford JR (1991) Sociolinguistic variation in Cane Walk: A quantitative case study. *English Around The World: Sociolinguistic Perspectives*, ed Chesire J (Cambridge University Press, Cambridge, UK), pp 609–616.

16. Labov W (1973) The linguistic consequences of being a lame. *Lang Soc* 2(1):81–115.

17. Ash S, Myhill J (1986) Linguistic correlates of inter-ethnic contact. *Diversity and Diachrony*, ed Sankoff D (John Benjamins, Amsterdam), pp 33–44.

18. Edwards WF (1992) Sociolinguistic behavior in a Detroit inner-city black neighborhood. *Lang Soc* 21(01):93.

19. Milroy L, Milroy J (1992) Social network and social class: Toward an integrated sociolinguistic model. *Lang Soc* 21(1):1–26.

20. Fordham S, Ogbu JU (1986) Black students' school success: Coping with the "burden of 'acting white.'" *Urban Rev* 18(3):176–206.

21. Rickford JR (1992) Grammatical variation and divergence in Vernacular Black English. In Internal and External Factors. *Syntactic Change*, eds Gerritsen M, Stein D (Mouton, The Hague), pp 175–200.

22. Woolard KA (1985) Language Variation and Cultural Hegemony: Toward an Integration of Sociolinguistic and Social Theory. *Am Ethnol* 12(4):738–748. Available at: http://www.jstor.org/stable/644180.

23. Rickford JR (1986) The need for new approaches to social class analysis in sociolinguistics. *Lang Commun* 6(3):215–221. Available at: http://www.stanford.edu/~eckert/PDF/rickford1986.pdf.

24. Eckert P (2000) *Linguistic Variation as Social Practice* (Blackwell, Oxford).

25. Ryan EB (1975) Why do low-prestige varieties persist? *Language and Social Psychology*, eds Giles H, St. Clair RN (University Park Press, Baltimore), pp 145–157.

26. Carmichael K (2014) "I never thought I had an accent until the hurricane." Sociolinguistic Variation in Post-Katrina Greater New Orleans. PhD Dissertation, Linguistics, Ohio State University.

27. Rickford JR, Rickford RJ (2000) *Spoken Soul: The Story of Black English* (Wiley, New York).

28. Fortson JG, Sanbonmatsu L (2010) Child Health and Neighborhood Conditions: Results from a Randomized Housing Voucher Experiment. *J Hum Resour* 45(4):840–864.

29. Kling JR, Ludwig J, Katz LF (2005) Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment. *Q J Econ* 120(1):87–130.

30. Kling JR, Liebman JB, Katz LF (2007) Experimental Analysis of Neighborhood Effects. *Econometrica* 75(1):83–119.

31. Sanbonmatsu L, Kling JR, Duncan GJ, Brooks-Gunn J (2006) Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment. *J Hum Resour* 41(4):649–691.

32. Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E (2004) *Survey Methodology* (John Wiley & Sons, New York). 7th Ed.

33. The American Association of Public Opinion Research (2011) *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 7th Edition*.

34. Lenneberg EH (1967) *Biological Foundations of Language* (Wiley, New York).

35. Sankoff G (2004) Adolescents, young adults and the critical period: Two case studies from “Seven Up.” *Sociolinguistic Variation: Critical Reflections*, ed Fought C (Oxford University Press, New York), pp 121–139.

36. Siegel J (2012) *Second Dialect Acquisition* (Cambridge University Press, Cambridge, UK).

37. Long MH (2007) *Problems in SLA* (Earlbaum, London and New York).

38. Labov W (1972) *Language in the Inner City: Studies in the Black English Vernacular* (University of Pennsylvania Press, Philadelphia).

39. Baugh J (1983) *Black Street Speech: Its History, Structure, and Survival* (University of Texas Press, Austin).

40. Rickford JR, McNair-Knox F (1994) Addressee- and Topic-Influenced Style Shift: A Quantitative Sociolinguitic Study. *Sociolinguistic Perspectives on Register*, eds Biber D, Finegan E (Oxford University Press, New York), pp 235–276.

41. Labov W (1972) Some principles of linguistic methodology. *Lang Soc* 1(1):97.

42. MacWhinney B (2014) The CHILDES Project: Tools for Analyzing Talk, Part 2: The CLAN programs. (Carnegie Mellon University. Pittsburgh). Available at http://childes.psy.cmu.edu.

43. Craig HK, Grogger JT (2012) Influences of Social and Style Variables on Adult Usage of African American English Features. *J Speech, Lang Hear Res* 55(5):1274–1288.

44. Anshen F (1969) *Speech Variation among Negroes in a Small Southern Community* (New York University, New York).

45. Fasold R (1972) *Tense marking in Black English: A linguistic and social analysis* (Center for Applied Linguistics, Washington, DC).

46. Baugh JG (1979) *Linguistic Style-Shifting in Black English* (Dissertations available from ProQuest. Paper AAI7928106. http://repository.upenn.edu/dissertations/AAI7928106, Philadelphia).

47. Alim HS (2004) *You Know My Steez: An Ethnographic and Sociolinguistic Study of Styleshifting in a Black American Speech Community* (Duke University Press, Durham, NC).

48. Craig HK, Washington JA (2006) *Malik goes to school: Examining the language skills of African American students from preschool-5th grade* (Lawrence Erlbaum Associates, Mahwah, New Jersey).

49. Renn J, Terry JM (2009) Operationalizing style: Quantifying style shift in the speech of African American adolescents. *Am Speech* 8(4):367–390.

50. Van Hofwegen J, Wolfram W (2010) Coming of age in African American English: A longitudinal study. *J Socioling* 14(4):427–455.

51. Purnell T, Idsardi W, Baugh J (1999) Perceptual and Phonetic Experiments on American English Dialect Identification. *J Lang Soc Psychol* 18(1):10–30.

52. Grogger J (2011) Speech Patterns and Racial Wage Inequality. *J Hum Resour* 46(1):1–25.

53. Mazumder B (2008) Sibling similarities and economic inequality in the US. *J Popul Econ* 21(3):685–701.

54. Angrist JD (2001) Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors. *J Bus Econ Stat* 19(1):2–28.

55. Rubin DB (1980) Comment on “Randomization Analysis of Experimental Data: The Fisher Randomization Test,” by D. Basu. *J Am Stat Assoc* 75(371):591–593. Available at: http://www.jstor.org/stable/2287653.

56. Sobel ME (2006) What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference. *J Am Stat Assoc* 101(476):1398–1407.

57. Ludwig J, et al. (2008) What Can We Learn about Neighborhood Effects from the Moving to Opportunity Experiment? *Am J Sociol* 114(1):144–188.

58. Bloom HS (1984) Accounting for No-Shows in Experimental Evaluation Designs. *Eval Rev* 8(2):225–246.

59. Angrist JD, Imbens GW, Rubin DB (1996) Identification of Causal Effects Using Instrumental Variables. *J Am Stat Assoc* 91(434):444–455.

60. Angrist JD, Pischke J-S (2009) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

61. Ludwig J, Kling JR (2007) Is Crime Contagious? *J Law Econ* 50(3):491–518.

62. Murray MP (2006) Avoiding Invalid Instruments and Coping with Weak Instruments. *J Econ Perspect* 20(4):111–132.

63. Fuller WA (1977) Some properties of a modification of the limited information estimator. *Econometrica* 45(4):939–953.

64. Staiger D, Stock J (1997) Instrumental variables regression with weak instruments. *Econometrica* 65(3):557–586.

65. Hahn J, Hausman JA, Kuersteiner G (2004) Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations. *Econom J* 7:272–306.

66. Kennedy SD, Leger ML (1990) *Final Comprehensive Report of the Freestanding Housing Voucher Demonstration* (U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, DC).

67. Rubinowitz LS, Rosenbaum JE (2000) *Crossing the class and color lines: From public housing to white suburbia* (University of Chicago Press, Chicago).

68. Feins JD, Shroder M (2005) Moving to opportunity: The demonstration's design and its effects on mobility. *Urban Stud* 42(8):1275–1299.

69. Shroder M (2002) Locational constraint, housing counseling, and successful lease-up in a randomized housing voucher experiment. *J Urban Econ* 51(2):315–338.

70. Ludwig J (2012) The Long-Term Results From the Moving to Opportunity Residential Mobility Demonstration. *Cityscape* 14(2):1–28.

71. Sampson RJ, Sharkey P, Raudenbush SW (2008) Durable effects of concentrated disadvantage on verbal ability among African-American children. *Proc Natl Acad Sci* 105(3):845–852.

72. Clark WAV (2005) Intervening in the residential mobility process: Neighborhood outcomes for low-income populations. *Proc Natl Acad Sci* 102(43):15307–15312.

73. Sampson RJ (2008) Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure. *Am J Sociol* 114(1):189–231.

74. Clampet-Lundquist S, Massey DS (2008) Neighborhood Effects on Economic Self-Sufficiency: A Reconsideration of the Moving to Opportunity Experiment. *Am J Sociol* 114(1):107–143.

75. Fordham S (1999) Dissin' "the Standard": Ebonics as Guerrilla Warfare at Capital High. *Anthropol Educ Q* 30(3):272–293.

76. Rickford JR, Price M (2013) Girlz II women: Age-grading, language change and stylistic variation. *J Socioling* 17(2):143–179.

77. Baugh J (1996) Dimensions of a theory of econolinguistics. *Towards a Social Science of Language: Papers in Honor of William Labov*, eds Guy GR, Feagin C, Schiffrin D, Baugh J (John Benjamins, Amsterdam), pp 397–419.

78. Wolfram W, Van Hofwegen J (2016) On the utility of composite indices in longitudinal language study. *Panel Studies of Language Variation and Change*, eds Buchstaller I, Wagner S (Routledge, London and New York).

79. Katz LF, Kling JR, Liebman JB (2001) Moving to Opportunity in Boston : Early Results of a Randomized Mobility Experiment. *Q J Econ* 116(2):607–654.

80. Labov W (1969) Contraction, deletion and inherent variation of the English copula. *Language (Baltim)* 45(4):715–762.

81. Jacob BA, Ludwig J, Miller DL (2013) The effects of housing and neighborhood conditions on child mortality. *J Health Econ* 32(1):195–206.

82. Cukor-Avila P, Bailey G (2001) The effects of the race of the interviewer on sociolinguistic fieldwork. *J Socioling* 5(2):252–270.

Supporting Exhibits
(click any line below to skip to the relevant exhibit)

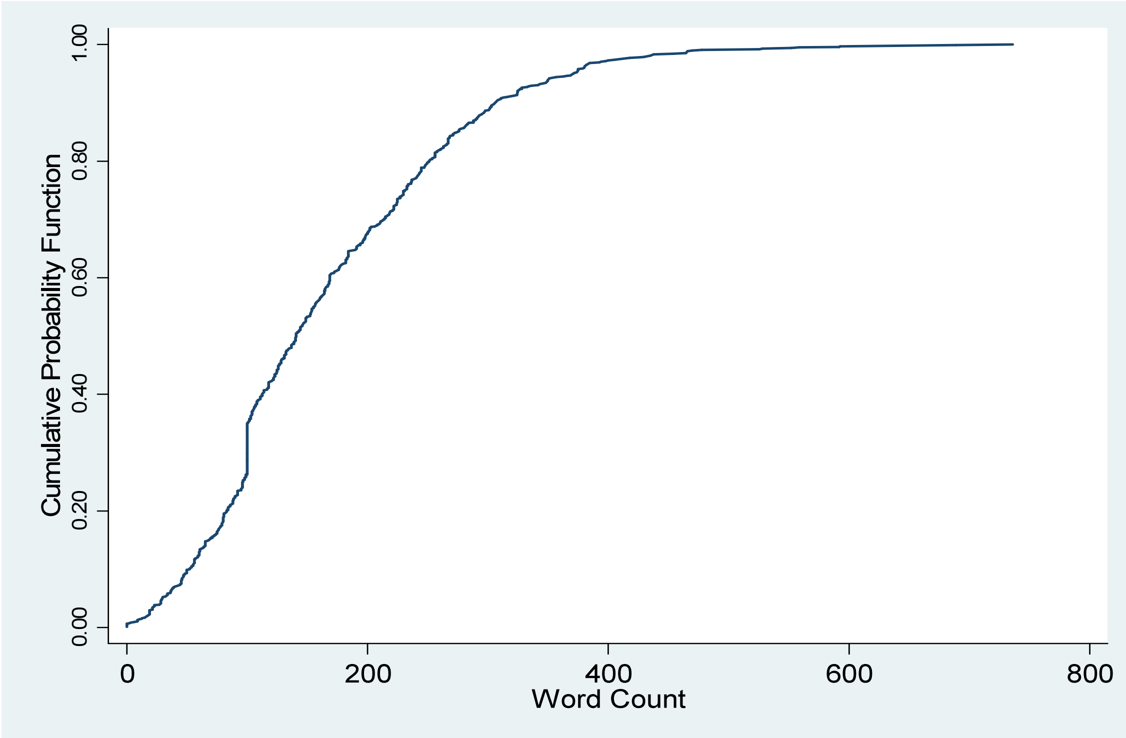

Figure S1. Cumulative Distribution Function for Speech Sample Word Count. The word counts represent the number of words in each speaker's speech sample analyzed for this study. The speech samples include responses to a question asking respondents to describe the happiest moment of their lives, but for some speakers the counts also included a description of their scariest moment (in addition to or instead of their happiest moment) as well as their general comments on MTO, their neighborhood, and housing programs. The sample is all non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the low-poverty voucher and control groups only (n=629).

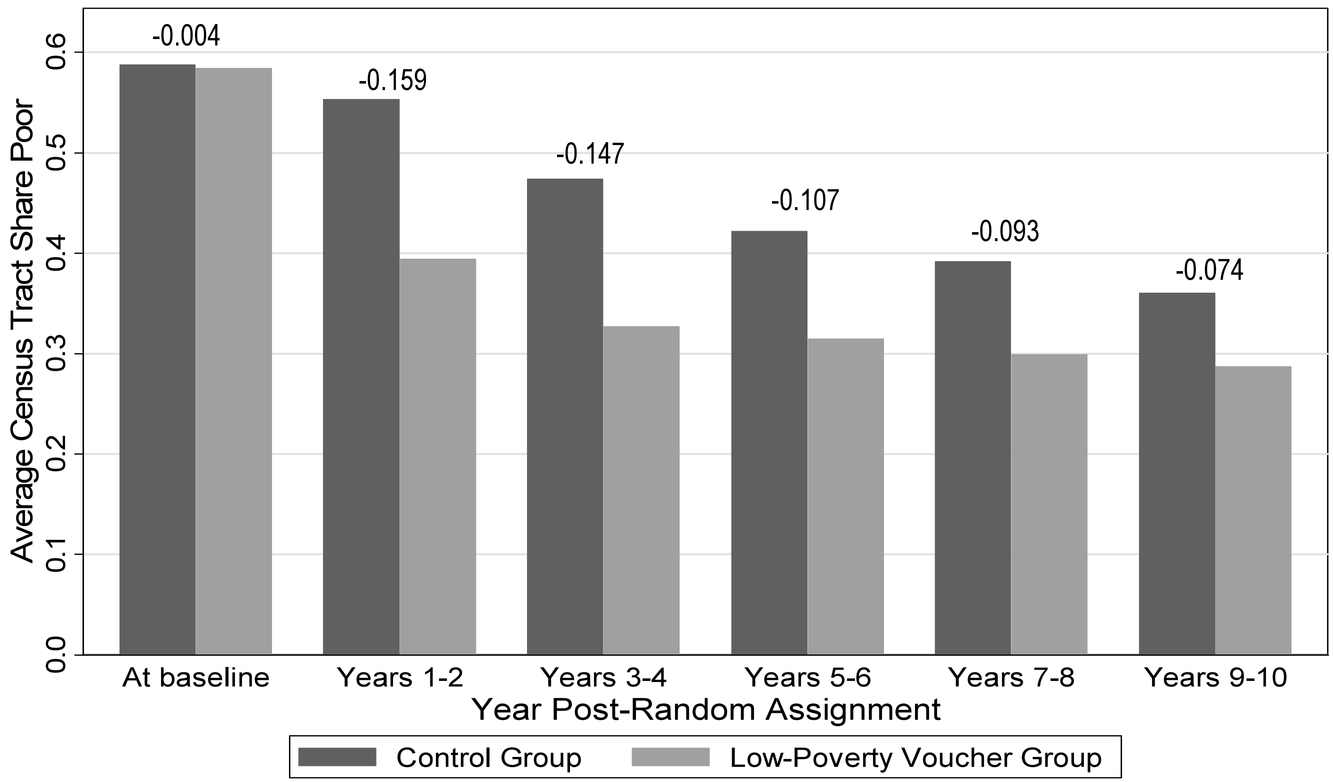

Figure S2. Neighborhood Poverty Rates Over Time by Treatment Status. Mean census tract poverty rates for low-poverty voucher (LPV) and control group youth for the addresses where the youth was living when the family enrolled in the MTO program (baseline) and average poverty rates for all the addresses where the youth lived for the subsequent two-year periods through 10 years after randomization, weighted by the amount of time the youth lived at each address. The value labels are the raw difference in weighted, unadjusted means between groups (LPV minus control). Census tract characteristics are interpolated data from the 1990 and 2000 decennial censuses as well as the 2005-09 American Community Survey. The sample is non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=629 youth).

Table S1. Effective Response Rates and Sample Counts by Treatment Group

	All Youth			Non-Hispanic African-American Youth		
	Control Group	Low-Poverty Voucher Group	Overall	Control Group	Low-Poverty Voucher Group	Overall
Response rates						
Overall sample						
Interviewed	0.892	0.901	0.897	0.914	0.908	0.911
Selected for language component						
Interviewed	0.869	0.915 ~	0.894	0.894	0.923	0.911
Consented to audio recording	0.749	0.764	0.757	0.789	0.765	0.775
Audio recorded	0.734	0.736	0.735	0.776	0.736	0.753
Language analyzed	0.628	0.607	0.616	0.741	0.681	0.707
Counts of youth						
Overall sample						
Selected for interview	2018	2417	4435	1122	1455	2577
Interviewed	1629	2013	3642	919	1227	2146
Selected for language component						
Selected for interview	736	948	1684	398	582	980
Interviewed	577	782	1359	316	487	803
Consented to audio recording	496	666	1162	278	411	689
Audio recorded	486	639	1125	272	395	667
Language analyzed	409	535	944	260	369	629

Notes : * = $P < 0.05$, ~ = $P < 0.10$ on an independent group t-test of the difference in effective response rates between the low-poverty voucher group and the control group. Effective response rates are calculated using sample weights to account for changes in random assignment ratios across randomization cohorts, survey sample selection, two-phase interviewing, and language component selection.

Table S2. Distribution of Language Feature Tokens by Treatment Group

	Control Mean	Low-Poverty Voucher Mean	Overall Mean
Grammatical features			
Ain't	0.026	0.022	0.023
Multiple negation	0.025	0.014 *	0.019
Third-person singular s-absence	0.025	0.016	0.020
Copula absence	0.023	0.022	0.022
Was-leveling	0.020	0.021	0.021
Phonological features			
Consonant cluster reduction	0.160	0.162	0.161
R-deletion (omitted category)	0.318	0.339	0.329
DH-stopping	0.226	0.207	0.216
TH-stopping	0.038	0.035	0.036
Ai monophthongization	0.140	0.163 ~	0.153

Notes : * = P <0.05, ~ = P <0.10 on an independent group t-test of the difference between the low-poverty voucher (LPV) group and the control group. Values represent the share of tokens that were analyzed for each feature regardless of whether the token represents the African-American Vernacular English (AAVE) or Standard American English (SAE) variant of the feature. The sample is all tokens from the speech samples of non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only (n=14,191 tokens from n=629 youth).

Table S3. Full List of Baseline Characteristics Controlled for in the Analysis

	Control Mean	Low-Poverty Voucher Mean	Overall Mean
Youth Characteristics			
Male	0.554	0.494	0.521
Age (as of December 31, 2007)			
13	0.104	0.094	0.099
14	0.091	0.099	0.095
15 (omitted category)	0.126	0.127	0.126
16	0.160	0.135	0.147
17	0.152	0.172	0.163
18	0.127	0.162	0.146
19	0.132	0.088	0.108
20	0.108	0.123	0.116
Age 6 or over at baseline	0.429	0.410	0.419
Health problems that limited activity	0.047	0.053	0.050
Health problems that required special medicine or equipment	0.070	0.076	0.074
Characteristics of youth age 6+ at baseline			
Gifted student or did advanced coursework	0.064	0.042	0.052
Suspended or expelled from school in past two years	0.016	0.007	0.011
School called about behavior in past two years	0.094	0.084	0.088
Behavioral or emotional problems	0.035	0.011 ~	0.022
Learning problems	0.062	0.016 *	0.037
Characteristics of youth age <6 at baseline			
In hospital before first birthday	0.127	0.074 *	0.098
Weighed less than 6 pounds at birth	0.092	0.080	0.086
Adult read to youth more than once per day	0.193	0.168	0.179
Household Adult Characteristics			
Age range (as of December 31, 2007)			
≤35	0.197	0.226	0.213
36-40	0.305	0.317	0.311
41-45	0.257	0.208	0.230
46-50	0.111	0.139	0.126
>50 (omitted category)	0.131	0.110	0.119
Education level			
Certificate of General Educational Development (GED)	0.209	0.119 *	0.160
High school diploma	0.354	0.406	0.382
Currently in school	0.246	0.171	0.205
Currently employed	0.245	0.258	0.252
Never married	0.751	0.791	0.772
Had first child before age 18	0.352	0.311	0.330
Household Characteristics			
Receiving Aid to Families with Dependent Children (AFDC)	0.818	0.853	0.837
Own car	0.126	0.172	0.151
Disabled household member	0.130	0.114	0.122
No teens in household	0.773	0.814	0.796
Household size			
Two	0.106	0.134	0.121
Three	0.227	0.254	0.242
Four	0.238	0.259	0.250
Five or more (omitted category)	0.429	0.353	0.387

Table S3. (continued)

	Control Mean	Low-Poverty Voucher Mean	Overall Mean
Neighborhood Characteristics			
Household member was crime victim in last 6 months	0.438	0.416	0.426
Streets unsafe at night	0.440	0.456	0.449
Very dissatisfied with neighborhood	0.485	0.485	0.485
Lived in neighborhood 5+ years	0.636	0.605	0.619
Moved more than 3 times in past 5 years	0.070	0.095	0.083
No family in neighborhood	0.603	0.563	0.581
No friends in neighborhood	0.447	0.334 *	0.385
Chatted with neighbors at least once per week	0.540	0.588	0.566
Very likely to tell neighbor about child getting into trouble	0.583	0.594	0.589
Confident about finding a new apartment	0.485	0.478	0.482
Had Section 8 voucher before	0.405	0.359	0.380
Head of household's primary or secondary reason for wanting to move			
To get away from gangs and drugs	0.779	0.756	0.767
Better schools for children	0.503	0.540	0.523
Randomization Site			
Baltimore	0.202	0.167	0.183
Boston	0.128	0.087	0.106
Chicago	0.374	0.338	0.355
Los Angeles	0.148	0.239 *	0.197
New York (omitted category)	0.148	0.169	0.159

Notes : * = P <0.05, ~ = P <0.10 on an independent group t-test of the difference between the low-poverty voucher (LPV) group and the control group. All values represent shares. Values are calculated using sample weights to account for changes in random assignment ratios across randomization cohorts, survey sample selection, two-phase interviewing, and language sample selection. Missing values were imputed based on randomization site and whether randomized through 1997 or in 1998. An omnibus F-test fails to reject the null hypothesis that the full set of baseline characteristics controlled for in the regression models and presented here is the same for the LPV group and the control group (P =0.088). The only analysis control variables not listed above are flags for missing data for the following characteristics: gifted student, suspended or expelled, behavioral problems, learning problems, hospitalization, low birth weight, read to by household member, activity-limiting health problems. Race and ethnicity were not controlled for in the analysis because they do not vary in the main sample. The sample is non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=629 youth).

Table S4. Characteristics of the Survey Interviewers by Treatment Group

	Control Mean	Low-Poverty Voucher (LPV) Mean	P-value of LPV-Control Difference
Age in whole years as of December 2007	48.4	45.7 ~	0.058
Gender			
Male	0.161	0.205	0.357
Female	0.785	0.765	0.681
Education Level			
High school diploma/GED or less	0.062	0.047	0.491
Some college	0.378	0.399	0.685
College graduate	0.298	0.379	0.127
Advanced degree	0.262	0.175 *	0.046
College graduate or advanced degree	0.560	0.554	0.911
Race/Ethnicity			
African-American	0.532	0.468	0.234
White	0.314	0.386	0.143
Other race	0.153	0.146	0.854
Hispanic (any race)	0.165	0.209	0.278

Notes : * = P <0.05, ~ = P <0.10 on an independent group t-test of the difference between the low-poverty voucher (LPV) group and the control group. All values (except age) represent shares. Values are calculated using sample weights to account for changes in random assignment ratios across randomization cohorts, survey sample selection, two-phase interviewing, and language sample selection. The characteristics are for the interviewers who administered the MTO long-term survey to non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (the unit of observation is the n=629 youth).

Table S5. Person-Level Regression Analysis Predicting Share AAVE Tokens by Interviewer Characteristics

	Bivariate Regression			***Controlling for Age, Site and Gender***			
Interviewer Characteristic	**Coefficient**		**P-value**	**Coefficient**		**P-value**	**N**
Age in whole years as of December 2007	0.002	*	0.006	0.001		0.474	603
Gender (male flag)	-0.051	*	0.041	-0.017		0.412	603
Education level							
High school diploma/GED or less	-0.023		0.525	-0.022		0.524	603
Some college	0.027		0.242	-0.001		0.964	603
College graduate	-0.060	*	0.009	-0.003		0.896	603
Advanced degree	0.048	~	0.052	0.012		0.600	603
College graduate or advanced degree	-0.022		0.339	0.006		0.771	603
Race/ethnicity							
African-American	0.108	*	0.000	0.036	~	0.085	596
White	-0.080	*	0.000	-0.026		0.194	596
Other race	-0.069	*	0.038	-0.015		0.585	596
Hispanic (any race)	-0.069	*	0.017	-0.022		0.362	597

Notes: * = P <0.05, ~ = P <0.10. Coefficients are from separate person-level, family-clustered, and weighted regression models that use the interviewer characteristic listed to predict the share of the speaker's tokens that reflect an African-American Vernacular English (AAVE) variant of the feature rather than Standard American English (SAE). The bivariate regression models include only the interviewer characteristic measure and are contrasted with models that also included controls for age (continuous), site, and gender. The characteristics are for the interviewers who administered the MTO long-term survey to non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the low-poverty voucher and control groups only and whose speech samples included at least one analyzable language token (the unit of observation is the n=629 youth; the actual sample sizes are lower due to missing interviewer characteristics data).

Table S6. MTO Effects on Neighborhood Conditions Over Time

	Baseline [n=618]		1 Year Post-RA [n=616]		5 Years Post-RA [n=608]		10-15 Years Post-RA [n=612]		Overall Post-RA Period (Duration-Weighted) [n=628]	
	CM	ITT	CM	ITT	CM	ITT	CM	ITT	CM	ITT
A. Tract share poor										
Share poor	0.588	-0.004	0.563	-0.211 *	0.422	-0.116 *	0.320	-0.044 *	0.427	-0.113 *
		(0.010)		(0.017)		(0.017)		(0.015)		(0.013)
Share poor, z-score on U.S. tracts	3.635	-0.032	3.438	-1.713 *	2.289	-0.942 *	1.466	-0.354 *	2.335	-0.917 *
		(0.085)		(0.135)		(0.139)		(0.119)		(0.103)
Share poor, z-score on MTO controls	0.000	-0.024	0.000	-1.272 *	0.000	-0.664 *	0.000	-0.257 *	0.000	-0.831 *
		(0.063)		(0.100)		(0.098)		(0.087)		(0.093)
B. Concentrated disadvantage index										
Concentrated disadvantage index	2.621	0.018	2.508	-0.595 *	2.037	-0.292 *	1.670	-0.131 *	2.075	-0.333 *
		(0.044)		(0.063)		(0.063)		(0.053)		(0.045)
Concentrated disadvantage index, z-score on MTO controls	0.000	0.023	0.000	-0.783 *	0.000	-0.447 *	0.000	-0.224 *	0.000	-0.645 *
		(0.056)		(0.083)		(0.096)		(0.090)		(0.086)
Concentrated disadvantage index (excluding percent black)	2.019	0.007	1.915	-0.553 *	1.476	-0.282 *	1.153	-0.110 *	1.505	-0.297 *
		(0.033)		(0.051)		(0.050)		(0.038)		(0.036)
Concentrated disadvantage index (excluding percent black), z-score on MTO controls	0.000	0.012	0.000	-0.999 *	0.000	-0.552 *	0.000	-0.261 *	0.000	-0.754 *
		(0.058)		(0.091)		(0.097)		(0.090)		(0.092)
C. Tract share black										
Share black	0.719	0.013	0.708	-0.050 *	0.670	-0.012	0.618	-0.025	0.682	-0.043 *
		(0.018)		(0.023)		(0.025)		(0.027)		(0.017)
Share black, z-score on U.S. tracts	2.474	0.057	2.428	-0.213 *	2.264	-0.052	2.044	-0.107	2.316	-0.184 *
		(0.075)		(0.099)		(0.105)		(0.113)		(0.071)
Share black, z-score on MTO controls	0.000	0.044	0.000	-0.161 *	0.000	-0.039	0.000	-0.075	0.000	-0.164 *
		(0.058)		(0.075)		(0.079)		(0.079)		(0.064)
D. Tract share minority										
Share minority	0.915	0.012	0.914	-0.107 *	0.907	-0.069 *	0.838	-0.029	0.905	-0.073 *
		(0.015)		(0.019)		(0.021)		(0.024)		(0.015)
Share minority, z-score on U.S. tracts	1.911	0.037	1.906	-0.345 *	1.885	-0.222 *	1.663	-0.092	1.876	-0.234 *
		(0.047)		(0.061)		(0.066)		(0.079)		(0.047)
Share minority, z-score on MTO controls	0.000	0.056	0.000	-0.544 *	0.000	-0.372 *	0.000	-0.115	0.000	-0.526 *
		(0.071)		(0.096)		(0.111)		(0.099)		(0.105)

Table S6. (continued)

	Baseline [n=618]		1 Year Post-RA [n=616]		5 Years Post-RA [n=608]		10-15 Years Post-RA [n=612]		Overall Post-RA Period (Duration-Weighted) [n=628]	
	CM	ITT	CM	ITT	CM	ITT	CM	ITT	CM	ITT
E. Tract share high school and college graduates										
Share with at least high school education	0.460	0.005 (0.010)	0.478	0.161 * (0.015)	0.574	0.096 * (0.015)	0.714	0.039 * (0.013)	0.584	0.089 * (0.011)
Share with at least some college	0.207	0.009 (0.009)	0.223	0.158 * (0.015)	0.308	0.092 * (0.016)	0.410	0.035 * (0.014)	0.311	0.081 * (0.011)
Share college graduates	0.089	0.008 (0.006)	0.098	0.108 * (0.012)	0.144	0.063 * (0.012)	0.210	0.032 * (0.012)	0.145	0.058 * (0.009)
F. Other tract characteristics										
Share Hispanic	0.191	0.000 (0.009)	0.195	-0.082 * (0.013)	0.220	-0.076 * (0.017)	0.187	-0.008 (0.016)	0.204	-0.047 * (0.012)
Share employed	0.692	-0.002 (0.008)	0.709	0.087 * (0.010)	0.782	0.045 * (0.010)	0.833	0.011 (0.008)	0.775	0.048 * (0.007)
Share workers in managerial or professional occupations	0.294	0.010 (0.009)	0.269	0.088 * (0.011)	0.213	0.045 * (0.011)	0.250	0.017 (0.011)	0.229	0.043 * (0.008)
Share receiving public assistance	0.377	0.002 (0.009)	0.330	-0.121 * (0.013)	0.187	-0.057 * (0.010)	0.078	-0.015 * (0.007)	0.193	-0.060 * (0.008)
Median household income (2009 dollars)	$17,297	-361 (689)	$18,501	19725 * (1511)	$26,940	10770 * (1594)	$32,445	3576 * (1583)	$26,310	10066 * (1112)
Share of families with single female-heads	0.685	0.014 (0.011)	0.665	-0.153 * (0.016)	0.564	-0.073 * (0.017)	0.510	-0.039 * (0.019)	0.579	-0.085 * (0.012)
Share under age 18	0.407	-0.007 (0.006)	0.397	-0.067 * (0.008)	0.351	-0.035 * (0.008)	0.291	-0.018 * (0.008)	0.348	-0.038 * (0.006)
Share owner-occupied housing	0.089	-0.007 (0.011)	0.099	0.211 * (0.019)	0.234	0.124 * (0.023)	0.372	0.032 (0.021)	0.244	0.115 * (0.017)

Table S6. (continued)

Notes : * = *P* <0.05, ~ = *P* <0.10 on two-tailed t-test. CM = control mean. ITT = low-poverty voucher (LPV) vs. control intention-to-treat (ITT) effect. Robust standard errors shown in parentheses. ITT effects were estimated using an ordinary least squares (OLS) regression model controlling for the baseline covariates in Table S3, using person-level survey weights, and clustering by family ID. The concentrated disadvantage index is a weighted combination of census tract percent [i] poverty, [ii] black (excluded where indicated), [iii] on welfare, [iv] unemployed, [v] female-headed family households, and [vi] under age 18, with loading factors developed using 2000 Census tracts in Chicago by Sampson, Sharkey, and Raudenbush (71). Census tract characteristics are presented for the addresses where the youth was living when the family enrolled in the MTO program (baseline) and where the youth was living 1, 5 and 10-15 years after random assignment (RA) as well as averaged across all addresses from random assignment through the address 10-15 years post-RA (as of May 31, 2008 and just prior to the start of the long-term survey fielding period), weighted by the amount of time the youth lived at each address. Census tract characteristics are interpolated data from the 1990 and 2000 decennial censuses as well as the 2005-09 American Community Survey. The sample is non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=629 youth). The sample sizes reflected in the column headers are for the tract share poor measure, but the sample varies only minimally by tract characteristic.

Table S7. MTO Effects on Mobility, Neighborhood Safety, School Characteristics and Social Networks

	Control mean	LPV vs. Control ITT Effect	P-value	N
A. Residential mobility				
Number of moves after random assignment	2.586	0.741 *	0.000	629
		(0.189)		
Distance between baseline and long-term survey addresses is at least…				
1 mile	0.745	0.113 *	0.002	600
		(0.036)		
2 miles	0.635	0.133 *	0.001	600
		(0.041)		
5 miles	0.413	0.111 *	0.029	600
		(0.051)		
10 miles	0.214	0.069	0.129	600
		(0.045)		
15 miles	0.171	0.032	0.437	600
		(0.041)		
B. Neighborhood safety and satisfaction				
Z-score index of neighborhood safety and satisfaction measures (higher values indicate less safety/satisfaction)	0.000	-0.216 *	0.023	629
		(0.094)		
Feel unsafe during day	0.144	-0.019	0.535	629
		(0.031)		
Feel unsafe at night	0.437	-0.079 ~	0.088	629
		(0.046)		
Saw drugs sold or used in past 30 days	0.417	-0.071 ~	0.098	627
		(0.043)		
Household member was victimized in past six months	0.265	-0.054	0.167	627
		(0.039)		
Gangs present in neighborhood	0.682	-0.065	0.128	627
		(0.042)		
Neither somewhat nor very satisfied with neighborhood	0.488	-0.059	0.192	628
		(0.045)		
C. Average school characteristics				
Z-score index of average school characteristics (higher values indicate more advantaged characteristics)	0.000	-0.406 *	0.000	629
		(0.095)		
Share students eligible for free lunch	0.710	-0.059 *	0.000	623
		(0.014)		
Share schools eligible for Title I	0.718	-0.090 *	0.000	627
		(0.022)		
Share black students	0.692	-0.031 *	0.037	624
		(0.015)		
Share minority students (not included in index)	0.917	-0.054 *	0.000	627
		(0.015)		

Table S7. (continued)

	Control mean	LPV vs. Control ITT Effect	P-value	N
C. Average school characteristics (continued)				
Share Hispanic students	0.201	-0.036 * (0.012)	0.002	627
Number of students (not included in index)	883.1	14.8 (29.2)	0.614	627
Pupil-teacher ratio	17.722	-0.015 (0.198)	0.941	626
School-level percentile ranking on state exam (not included in index)	16.852	4.132 * (1.326)	0.002	607
School-level ranking on state exam above 50th percentile	0.093	0.033 ~ (0.019)	0.081	607
D. Youth social networks				
Z-score index of youth social networks measures (higher values indicate less risky social networks)	0.000	0.063 (0.085)	0.457	629
Visits with friends from baseline neighborhood at least once per week	0.314	-0.032 (0.043)	0.454	619
Close friends think that studying is very important	0.391	0.000 (0.046)	0.996	627
Close friends think that continuing their education past high school is very important	0.697	0.021 (0.041)	0.605	628
Close friends involved in school activities	0.802	0.019 (0.036)	0.593	623
Close friends have used marijuana or other drugs	0.421	-0.041 (0.044)	0.346	611
Close friends have dropped out of school	0.219	0.020 (0.036)	0.575	628
E. Adult social network measures applied to youth sample				
Z-score index of adult social networks measures (higher values indicate more diverse social networks)	0.000	0.136 (0.097)	0.159	601
Household adult has a close friend who graduated from college	0.615	0.095 ~ (0.049)	0.051	591
Household adult has a close friend who works full-time	0.786	0.044 (0.038)	0.257	601
Household adult has a close friend of a different race or ethnicity	0.405	0.008 (0.050)	0.867	601

Table S7. (continued)

	Control mean	LPV vs. Control ITT Effect	P-value	N
F. Youth time use				
Attended youth activities at church at least once a month in past year	0.444	-0.091 *	0.048	629
		(0.046)		
Hangs out in neighborhood or at basketball court once or more per week	0.484	0.023	0.580	629
		(0.041)		
Shops at mall or store once or more per week	0.442	0.125 *	0.006	629
		(0.045)		
Hangs out at someone else's house once or more per week	0.631	-0.052	0.226	629
		(0.043)		
Never hangs out at home	0.074	0.002	0.918	629
		(0.021)		

Notes : * = *P* <0.05, ~ = *P* <0.10 on two-tailed t-test. Robust standard errors shown in parentheses. Low-poverty voucher (LPV) vs. control intention-to-treat (ITT) effects were estimated using an ordinary least squares (OLS) regression model controlling for the baseline covariates in Table S3, using person-level survey weights, and clustering by family ID. Long-term survey address is the address where the youth was living as of May 31, 2008, just prior to the start of the long-term survey fielding period and the end point for our duration-weighted address measures. The safety measures reflect whether the respondent felt unsafe or very unsafe (vs. safe or very safe) in the neighborhood during the day or at night. School characteristics are for grades K-12 and are averaged across all schools/academic years from random assignment through the youth's current school as of the survey interview (or most recent primary or secondary school for youth who were no longer enrolled). The index measures for selected panels combined z-scored versions of the individual measures within each panel (standardized using the control group mean and standard deviation). The index is a restandardized average of the z-scored components. Except where indicated, index components are all individual measures within each panel (for the school characteristics index, the indicator for whether the average school percentile ranking was above the 50th percentile was reversed for consistency with other measures). Neighborhood, social network, and time use measures come from the youth long-term survey (except for panel E, where measures come from adult long-term survey data applied to youth). School characteristics come from the Common Core of Data, the Private School Universe Survey, and the National Longitudinal School-Level State Assessment Score Database. The sample is non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=629 youth).

Table S8. Sensitivity of MTO Effects on African-American Vernacular English Use Among Youth

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	P-value	N: Tokens	N: Youth
A. Main estimate from Table 2	0.485	-0.028 ~ (0.015)	0.056	14191	629
B. Person-level share of AAVE tokens	0.488	-0.033 ~ (0.019)	0.078	N/A	629
C. Alternative regression approaches					
Probit model (marginal effects)	0.485	-0.028 ~ (0.015)	0.056	14191	629
Logit model (marginal effects)	0.485	-0.029 ~ (0.015)	0.052	14191	629
D. Models restricted to longer audio samples					
Audio sample contains 50+ words	0.485	-0.028 ~ (0.015)	0.065	13933	571
Audio sample contains 5+ tokens	0.485	-0.029 ~ (0.015)	0.054	14065	578
Audio sample contains 10+ tokens	0.486	-0.026 ~ (0.015)	0.082	13595	510
E. Model includes only site covariates	0.485	-0.026 (0.017)	0.137	14191	629
F. By race/ethnicity definition					
All African-Americans (including Hispanics), youth self-report	0.488	-0.028 ~ (0.015)	0.055	14523	656
All African-Americans (including Hispanics), household adult's race	0.488	-0.024 (0.015)	0.106	14250	646
Non-Hispanic African-Americans, household adult's race/ethnicity	0.484	-0.023 (0.015)	0.119	13971	622

Notes : * = P <0.05, ~ = P <0.10 on two-tailed t-test. Robust standard errors shown in parentheses. Low-poverty voucher (LPV) vs. control intention-to-treat (ITT) effects were estimated using an ordinary least squares (OLS) regression model controlling for the baseline covariates in Table S3 (except panel E), using person-level survey weights, and clustering by family ID. The analysis was primarily based on a question asking speakers to describe the happiest moment of their lives, but for some respondents it also included a description of their scariest moment (in addition to or instead of their happiest moment) as well as their general comments on MTO, their neighborhood, and housing programs. Each token was analyzed for whether the speaker used the African-American Vernacular English or the Standard American English variant. The tokens analyzed include 10 individual language features: ain't use, multiple negation, third-person singular s-absence, copula absence, was-leveling, consonant cluster reduction, R-deletion, DH-stopping, TH-stopping, and ai monophthongization. Except where indicated, the sample is all tokens from the speech samples of non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=14,191 tokens from n=629 youth; the person-level analysis in panel B uses the n=629 youth sample).

Table S9. MTO Effects on African-American Vernacular English Use by Gender, Age and Duration of Exposure

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	P-value	N: Tokens	N: Youth
A. Main estimate from Table 2	0.485	-0.028 ~ (0.015)	0.056	14191	629
B. By gender					
Female	0.482	-0.045 * (0.020)	0.030	7347	307
Male	0.489	-0.011 (0.021)	0.589	6844	322
Male minus female difference		0.033 (0.029)	0.245		
C. By age in 2008					
Age <17 in 2008 (baseline age: mean=2.4, range: 0-5; mean exposure=12.5 years)	0.477	0.014 (0.026)	0.572	4459	210
Age 17+ in 2008 (baseline age: mean=6.1, range: 2-11; mean exposure=12.5 years)	0.489	-0.047 * (0.017)	0.007	9732	419
Age 17+ minus age <17 difference		-0.062 * (0.030)	0.041		
D. By baseline age group					
Age <6 at baseline (2008 age: mean=16.3, range: 14-20; mean exposure=12.8 years)	0.487	0.003 (0.020)	0.882	8050	377
Age 6+ at baseline (2008 age: mean=19.5, range: 17-21; mean exposure=12.1 years)	0.484	-0.068 * (0.021)	0.001	6141	252
Age 6+ minus age <6 difference		-0.071 * (0.029)	0.013		

Table S9. (continued)

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)		P-value	N: Tokens	N: Youth
E. By gender and 2008 age group						
Female/age <17 in 2008	0.463	-0.020		0.571	2275	106
		(0.035)				
Female/age 17+ in 2008	0.489	-0.057	*	0.022	5072	201
		(0.025)				
Male/age <17 in 2008	0.490	0.048		0.191	2184	104
		(0.037)				
Male/age 17+ in 2008	0.488	-0.039		0.113	4660	218
		(0.024)				
Female, difference between age <17 and 17+ in 2008		0.037		0.397	7347	307
		(0.043)				
Age <17 in 2008, difference between female and male		-0.068		0.184	4459	210
		(0.051)				
Female/age <17 in 2008 - Male/age 17+ in 2008		0.019		0.656	6935	324
		(0.042)				
Female/age 17+ in 2008 - Male/age <17 in 2008		-0.105	*	0.017	7256	305
		(0.044)				
Age 17+ in 2008, difference between female and male		-0.018		0.606	9732	419
		(0.035)				
Male, difference between age <17 and 17+ in 2008		0.087	*	0.048	6844	322
		(0.044)				
F. By gender and baseline age group						
Female/age <6 at baseline	0.497	-0.030		0.239	4178	190
		(0.026)				
Female/age 6+ at baseline	0.461	-0.067	*	0.038	3169	117
		(0.032)				
Male/age <6 at baseline	0.475	0.036		0.204	3872	187
		(0.029)				
Male/age 6+ at baseline	0.502	-0.069	*	0.016	2972	135
		(0.028)				
Female, difference between age <6 and 6+ at baseline		0.037		0.368	7347	307
		(0.041)				
Age <6 at baseline, difference between female and male		-0.067	~	0.078	8050	377
		(0.038)				
Female/age <6 at baseline - Male/age 6+ at baseline		0.039		0.300	7150	325
		(0.037)				
Female/age 6+ at baseline - Male/age <6 at baseline		-0.104	*	0.018	7041	304
		(0.044)				
Age 6+ at baseline, difference between female and male		0.001		0.974	6141	252
		(0.044)				
Male, difference between age <6 and 6+ at baseline		0.105	*	0.009	6844	322
		(0.040)				

Table S9. (continued)

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	P-value	N: Tokens	N: Youth
G. By age group at baseline and in 2008					
Age <6 at baseline/age <17 in 2008 (mean exposure=12.5 years)	0.477	0.013 (0.025)	0.604	4459	210
Age <6 at baseline/age 17+ in 2008 (mean exposure=13.2 years)	0.498	-0.009 (0.030)	0.752	3591	167
Age 6+ at baseline/age 17+ in 2008 (mean exposure=12.1 years)	0.484	-0.068 * (0.021)	0.001	6141	252
Age <6 at baseline, difference between age 17+ and <17 in 2008		-0.023 (0.039)	0.562	8050	377
Age 17+ in 2008, difference between age 6+ and <6 at baseline		-0.059 (0.037)	0.112	10600	462
Age 6+ at baseline/age 17+ in 2008 - Age <6 at baseline/age <17 in 2008 difference		-0.082 * (0.032)	0.012	9732	419
H. By 2008 age group and randomization cohort					
Age <17 in 2008/early cohort (mean baseline age=1.4)	0.555	-0.031 (0.039)	0.415	2384	110
Age <17 in 2008/late cohort (mean baseline age=3.3)	0.403	0.064 ~ (0.037)	0.088	2075	100
Age 17+ in 2008/early cohort (mean baseline age=5)	0.507	-0.049 ~ (0.026)	0.056	5346	237
Age 17+ in 2008/late cohort (mean baseline age=7.2)	0.473	-0.045 ~ (0.024)	0.060	4386	182
Age <17 in 2008, difference between late and early cohort		0.095 ~ (0.057)	0.092	4459	210
Early cohort, difference between age 17+ and <17 in 2008		-0.018 (0.044)	0.683	7730	347
Age 17+ in 2008/late cohort - Age <17 in 2008/early cohort difference		-0.013 (0.045)	0.770	6770	292
Age 17+ in 2008/early cohort - Age <17 in 2008/late cohort difference		-0.113 * (0.046)	0.014	7421	337
Late cohort, difference between age 17+ and <17 in 2008		-0.109 * (0.044)	0.014	6461	282
Age 17+ in 2008, difference between late and early cohort		0.005 (0.035)	0.890	9732	419

Table S9. (continued)

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	P-value	N: Tokens	N: Youth
I. By baseline age group and randomization cohort					
Age <6 at baseline/early cohort (mean age in 2008=17.1)	0.523	-0.021 (0.030)	0.471	4346	199
Age <6 at baseline/late cohort (mean age in 2008=15.5)	0.452	0.028 (0.027)	0.309	3704	178
Age 6+ at baseline/early cohort (mean age in 2008=20.3)	0.496	-0.075 * (0.032)	0.018	3631	141
Age 6+ at baseline/late cohort (mean age in 2008=18.7)	0.472	-0.060 * (0.030)	0.047	2510	111
Age <6 at baseline, difference between late and early cohort		0.049 (0.041)	0.233	8050	377
Early cohort, difference between age 6+ and <6 at baseline		-0.054 (0.041)	0.193	7977	340
Age 6+ at baseline/late cohort - Age <6 at baseline/early cohort difference		-0.039 (0.043)	0.368	6856	310
Age 6+ at baseline/early cohort - Age <6 at baseline/late cohort difference		-0.103 * (0.042)	0.015	7335	319
Late cohort, difference between age 6+ and <6 at baseline		-0.088 * (0.040)	0.028	6214	289
Age 6+ at baseline, difference between late and early cohort		0.015 (0.044)	0.735	6141	252

Notes : * = *P* <0.05, ~ = *P* <0.10 on two-tailed t-test. Robust standard errors shown in parentheses. Low-poverty voucher (LPV) vs. control intention-to-treat (ITT) effects were estimated using an ordinary least squares (OLS) regression model controlling for the baseline covariates in Table S3, using person-level survey weights, and clustering by family ID. Subgroup analyses were run as an interaction with the treatment group indicator. The analysis was primarily based on a question asking speakers to describe the happiest moment of their lives, but for some respondents it also included a description of their scariest moment (in addition to or instead of their happiest moment) as well as their general comments on MTO, their neighborhood, and housing programs. Each token was analyzed for whether the speaker used the African-American Vernacular English or the Standard American English variant. The tokens analyzed include 10 individual language features: ain't use, multiple negation, third-person singular s-absence, copula absence, was-leveling, consonant cluster reduction, R-deletion, DH-stopping, TH-stopping, and ai monophthongization. The sample is all tokens from the speech samples of non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=14,191 tokens from n=629 youth).

Table S10. MTO Effects on African-American Vernacular English Use by Reason for Wanting to Move

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	P-value	N: Tokens	N: Youth
A. Main estimate from Table 2	0.485	-0.028 ~ (0.015)	0.056	14191	629
By reason for wanting to move					
B. Primary reason: drugs and gangs or better schools					
To get away from drugs and gangs or for better schools for the children	0.494	-0.052 * (0.018)	0.003	9754	440
Another reason	0.465	0.030 (0.025)	0.224	4437	189
Drugs/gangs or schools minus another reason difference		-0.083 * (0.030)	0.007		
C. Primary or secondary reason: drugs and gangs					
To get away from drugs and gangs	0.498	-0.050 * (0.017)	0.003	10947	487
Another reason	0.441	0.050 ~ (0.030)	0.091	3244	142
Drugs/gangs minus another reason difference		-0.101 * (0.035)	0.004		
D. Primary reason only: drugs and gangs					
To get away from drugs and gangs	0.500	-0.040 ~ (0.020)	0.050	7509	440
Another reason	0.469	-0.013 (0.022)	0.551	6682	189
Drugs/gangs minus another reason difference		-0.027 (0.030)	0.373		
E. Primary or secondary reason: better schools					
Better schools for the children	0.498	-0.039 ~ (0.020)	0.054	7399	322
Another reason	0.472	-0.016 (0.022)	0.461	6792	307
Better schools minus another reason difference		-0.023 (0.031)	0.453		
F. Primary reason only: better schools					
Better schools for the children	0.475	-0.085 * (0.038)	0.024	2245	104
Another reason	0.487	-0.016 (0.016)	0.309	11946	525
Better schools minus another reason difference		-0.069 ~ (0.041)	0.094		

Table S10. (continued)

Notes : * = *P* <0.05, ~ = *P* <0.10 on two-tailed t-test. Robust standard errors shown in parentheses. Low-poverty voucher (LPV) vs. control intention-to-treat (ITT) effects were estimated using an ordinary least squares (OLS) regression model controlling for the baseline covariates in Table S3, using person-level survey weights, and clustering by family ID. Subgroup analyses were run as an interaction with the treatment group indicator. The analysis was primarily based on a question asking speakers to describe the happiest moment of their lives, but for some respondents it also included a description of their scariest moment (in addition to or instead of their happiest moment) as well as their general comments on MTO, their neighborhood, and housing programs. Each token was analyzed for whether the speaker used the African-American Vernacular English or the Standard American English variant. The tokens analyzed include 10 individual language features: ain't use, multiple negation, third-person singular s-absence, copula absence, was-leveling, consonant cluster reduction, R-deletion, DH-stopping, TH-stopping, and ai monophthongization. The sample is all tokens from the speech samples of non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=14,191 tokens from n=629 youth).

Table S11. MTO Effects on African-American Vernacular English Use by Language Feature, Gender and Age

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: P-value	N: Tokens	N: Youth
A. Main estimate from Table 2	0.485	-0.028 ~ (0.015)	0.056	14191	629
B. By language feature category					
Grammatical token	0.490	-0.032 (0.038)	0.405	1492	422
Phonological token	0.485	-0.028 ~ (0.015)	0.068	12699	627
Phonological - grammatical token difference		0.004 (0.039)	0.926		
C. By grammatical language feature					
Ain't token	0.290	-0.015 (0.063)	0.812	361	202
Multiple negation token	0.646	-0.211 * (0.078)	0.008	253	155
Third-person singular s-absence token	0.359	0.165 * (0.076)	0.032	258	114
Copula absence token	0.412	-0.116 (0.071)	0.105	330	161
Was-leveling token	0.805	-0.138 ~ (0.075)	0.066	290	164
D. By phonological language feature					
Consonant cluster reduction token	0.708	-0.027 (0.024)	0.267	2278	503
R-deletion token	0.388	-0.035 (0.024)	0.138	4684	598
DH-stopping token	0.418	0.005 (0.027)	0.839	3027	462
TH-stopping token	0.694	0.014 (0.051)	0.787	541	267
Ai monophthongization token	0.500	-0.010 (0.034)	0.773	2169	488
E. Language feature category analysis by gender					
Grammatical token					
Female	0.430	-0.074 ~ (0.043)	0.086	851	211
Male	0.572	0.027 (0.051)	0.592	641	211
Male - female difference		0.101 (0.066)	0.129	1492	422
Phonological token					
Female	0.490	-0.045 * (0.021)	0.035	6496	306
Male	0.480	-0.009 (0.021)	0.655	6203	321
Male - female difference		0.036 (0.029)	0.226	12699	627

Table S11. (continued)

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	P-value	N: Tokens	N: Youth
F. Phonological language feature analysis by gender					
Consonant cluster reduction token					
Female	0.662	-0.039 (0.037)	0.296	1169	245
Male	0.758	-0.015 (0.030)	0.611	1109	258
Male - female difference		0.024 (0.046)	0.607	2278	503
R-deletion token					
Female	0.413	-0.054 ~ (0.033)	0.096	2304	294
Male	0.368	-0.017 (0.032)	0.601	2380	304
Male - female difference		0.038 (0.044)	0.393	4684	598
DH-stopping token					
Female	0.395	-0.020 (0.036)	0.582	1587	233
Male	0.438	0.034 (0.038)	0.374	1440	229
Male - female difference		0.054 (0.052)	0.301	3027	462
TH-stopping token					
Female	0.665	0.008 (0.071)	0.908	314	141
Male	0.729	0.021 (0.071)	0.768	227	126
Male - female difference		0.013 (0.101)	0.898	541	267
Ai monophthongization token					
Female	0.538	-0.021 (0.048)	0.653	1122	236
Male	0.465	0.002 (0.053)	0.977	1047	252
Male - female difference		0.023 (0.073)	0.754	2169	488

Table S11. (continued)

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	P-value	N: Tokens	N: Youth
G. Language feature category analysis by age in 2008					
Grammatical token					
Age <17 in 2008	0.489	0.064 (0.061)	0.298	436	134
Age 17+ in 2008	0.490	-0.072 ~ (0.039)	0.070	1056	288
Age 17+ minus age <17 difference		-0.135 ~ (0.073)	0.063	1492	422
Phonological token					
Age <17 in 2008	0.475	0.011 (0.025)	0.665	4023	209
Age 17+ in 2008	0.489	-0.045 * (0.018)	0.016	8676	418
Age 17+ minus age <17 difference		-0.055 ~ (0.030)	0.067	12699	627
H. Phonological language feature analysis by age in 2008					
Consonant cluster reduction token					
Age <17 in 2008	0.731	-0.039 (0.043)	0.363	755	169
Age 17+ in 2008	0.699	-0.022 (0.031)	0.479	1523	334
Age 17+ minus age <17 difference		0.017 (0.054)	0.746	2278	503
R-deletion token					
Age <17 in 2008	0.362	-0.001 (0.038)	0.972	1521	203
Age 17+ in 2008	0.400	-0.051 ~ (0.029)	0.082	3163	395
Age 17+ minus age <17 difference		-0.050 (0.048)	0.298	4684	598
DH-stopping token					
Age <17 in 2008	0.418	0.076 (0.049)	0.119	918	155
Age 17+ in 2008	0.418	-0.026 (0.031)	0.404	2109	307
Age 17+ minus age <17 difference		-0.102 ~ (0.057)	0.075	3027	462
TH-stopping token					
Age <17 in 2008	0.678	0.025 (0.086)	0.772	195	89
Age 17+ in 2008	0.702	0.008 (0.066)	0.904	346	178
Age 17+ minus age <17 difference		-0.017 (0.110)	0.878	541	267

Table S11. (continued)

Outcome	Control Mean	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: Coefficient (SE)	Low-Poverty Voucher vs. Control Intention-to-Treat Effect: P-value	N: Tokens	N: Youth
H. Phonological language feature analysis by age in 2008 (continued)					
Ai monophthongization token					
Age <17 in 2008	0.508	0.003 (0.057)	0.960	634	162
Age 17+ in 2008	0.498	-0.015 (0.043)	0.730	1535	326
Age 17+ minus age <17 difference		-0.018 (0.072)	0.807	2169	488
I. Model controls for language feature flags	0.485	-0.026 ~ (0.015)	0.081	14191	629

Notes : * = *P* <0.05, ~ = *P* <0.10 on two-tailed t-test. Robust standard errors shown in parentheses. Low-poverty voucher (LPV) vs. control intention-to-treat (ITT) effects were estimated using an ordinary least squares (OLS) regression model controlling for the baseline covariates in Table S3, using person-level survey weights, and clustering by family ID. The model in panel I also controlled for the ten language feature flags in panels C and D. Subgroup analyses were run as an interaction with the treatment group indicator. The analysis was primarily based on a question asking speakers to describe the happiest moment of their lives, but for some respondents it also included a description of their scariest moment (in addition to or instead of their happiest moment) as well as their general comments on MTO, their neighborhood, and housing programs. Each token was analyzed for whether the speaker used the African-American Vernacular English or the Standard American English variant. The tokens analyzed include 10 individual language features: ain't use, multiple negation, third-person singular s-absence, copula absence, was-leveling, consonant cluster reduction, R-deletion, DH-stopping, TH-stopping, and ai monophthongization. The sample is all tokens from the speech samples of non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=14,191 tokens from n=629 youth).

Table S12. MTO Intention-to-Treat (ITT) and Treatment-on-the-Treated (TOT) Effects on African-American Vernacular English Use

Outcome	Control Mean	ITT (SE)	Control Complier Mean	TOT (SE)	P-value	N: Tokens	N: Youth
AAVE variant used in token							
A. Overall	0.485	-0.028 ~	0.472	-0.057 ~	0.056	14191	629
		(0.015)		(0.030)			
B. By gender							
Female	0.482	-0.045 *	0.476	-0.097 *	0.030	7347	307
		(0.020)		(0.044)			
Male	0.489	-0.011	0.471	-0.021	0.589	6844	322
		(0.021)		(0.038)			
C. By age in 2008							
Age <17 in 2008	0.477	0.014	0.429	0.028	0.572	4459	210
		(0.026)		(0.050)			
Age 17+ in 2008	0.489	-0.047 *	0.491	-0.095 *	0.007	9732	419
		(0.017)		(0.035)			
D. By household head's primary reason for wanting to move							
To get away from drugs and gangs or for better schools for the children	0.494	-0.052 *	0.499	-0.102 *	0.003	9754	440
		(0.018)		(0.034)			
Another reason	0.465	0.030	0.396	0.065	0.224	4437	189
		(0.025)		(0.053)			

Notes : * = P <0.05, ~ = P <0.10 on two-tailed t-test. Robust standard errors shown in parentheses. Low-poverty voucher (LPV) vs. control intention-to-treat (ITT) effects were estimated using an ordinary least squares (OLS) regression model controlling for the baseline covariates in Table S3, using person-level survey weights, and clustering by family ID. Treatment-on-the-treated (TOT) effects were calculated by inflating the ITT effects by the LPV group compliance (or MTO voucher use) rate. Subgroup analyses were run as an interaction with the treatment group indicator. The analysis was primarily based on a question asking speakers to describe the happiest moment of their lives, but for some respondents it also included a description of their scariest moment (in addition to or instead of their happiest moment) as well as their general comments on MTO, their neighborhood, and housing programs. Each token was analyzed for whether the speaker used the African-American Vernacular English (AAVE) or the Standard American English variant. The tokens analyzed include 10 individual language features: ain't use, multiple negation, third-person singular s-absence, copula absence, was-leveling, consonant cluster reduction, R-deletion, DH-stopping, TH-stopping, and ai monophthongization. The sample is all tokens from the speech samples of non-Hispanic African-American youth speakers (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the LPV and control groups only and whose speech samples included at least one analyzable language token (n=14,191 tokens from n=629 youth).

Table S13. Instrumental Variables Estimates of the Relationship between Person-Level Share AAVE Tokens and Duration-Weighted Tract Share Poor, Share Black, and Share Minority (Z-Score Measures)

	Model										First Stage Statistics		
	2SLS		LIML		Fuller (c=1)		Fuller (c=2)		Fuller (c=4)				
Mediator(s) Included in Model	**Coeff. (SE)**	**P-value**	**Coeff. (SE)**	**P-value**	**Coeff. (SE)**	**P-value**	**Coeff. (SE)**	**P-value**	**Coeff. (SE)**	**P-value**	**Partial R-Sq.**	**Angrist-Pischke F-stat**	**Cragg-Donald F-stat**
A. Single mediator in model													
Share poor (z-score)	0.155 ~ (0.093)	0.096	0.156 ~ (0.094)	0.098	0.155 ~ (0.093)	0.097	0.154 ~ (0.093)	0.096	0.153 ~ (0.091)	0.095	0.190	20.85	
Share black (z-score)	0.235 (0.317)	0.459	0.261 (0.390)	0.504	0.254 (0.370)	0.493	0.247 (0.352)	0.482	0.236 (0.321)	0.462	0.039	4.25	
Share minority (z-score)	0.204 (0.141)	0.148	0.213 (0.150)	0.158	0.209 (0.147)	0.154	0.206 (0.143)	0.150	0.200 (0.136)	0.142	0.068	7.16	
B. Both mediators in model													
Share poor (z-score) vs. share black (z-score)													
Share poor controlling for share black	0.175 (0.142)	0.219	0.181 (0.151)	0.231	0.177 (0.145)	0.223	0.174 (0.141)	0.217	0.168 (0.133)	0.206	0.139	16.86	3.27
Share black controlling for share poor	-0.110 (0.482)	0.819	-0.133 (0.536)	0.804	-0.118 (0.501)	0.813	-0.106 (0.470)	0.823	-0.084 (0.419)	0.842	0.029	3.01	
P-value of test that coefficients are equal		0.633		0.635		0.633		0.633		0.631			
Share poor (z-score) vs. share minority (z-score)													
Share poor controlling for share minority	0.404 (0.415)	0.331	0.503 (0.580)	0.386	0.390 (0.395)	0.323	0.326 (0.305)	0.286	0.255 (0.217)	0.241	0.017	1.88	0.69
Share minority controlling for share poor	-0.399 (0.634)	0.529	-0.558 (0.903)	0.537	-0.378 (0.601)	0.529	-0.277 (0.453)	0.542	-0.166 (0.306)	0.588	0.006	0.53	
P-value of test that coefficients are equal		0.440		0.472		0.437		0.421		0.410			

Table S13. (continued)

Notes : * = *P* <0.05, ~ = *P* <0.10. Coefficient estimates for the various instrumental variable regressions shown use site/treatment group interactions as instruments. Each regression also controlled for the baseline covariates in Table S3 and was weighted and clustered by family ID. Columns labels are as follows: 2SLS columns report results for two-stage least squares, LIML is an unmodified limited information maximum likelihood (LIML) model, and columns labeled Fuller present Fuller-modified LIML models with constants 1, 2 and 4, respectively. Corresponding p-values are shown next to each model. Robust standard errors are shown in parentheses. The analysis was primarily based on a question asking speakers to describe the happiest moment of their lives, but for some respondents it also included a description of their scariest moment (in addition to or instead of their happiest moment) as well as their general comments on MTO, their neighborhood, and housing programs. The share of AAVE tokens measure is the sum of AAVE tokens for all 10 individual features (ain't use, multiple negation, third-person singular s-absence, copula absence, was-leveling, consonant cluster reduction, R-deletion, DH-stopping, TH-stopping, and ai monophthongization) divided by the sum of total tokens for all 10 features. Share poor is the fraction of census tract residents living below the poverty threshold, and share black and share minority are the fractions of census tract residents who are black and either non-white or Hispanic, respectively. Share poor, share black, and share minority come from interpolated data from the 1990 and 2000 decennial census as well as the 2005-09 American Community Survey and are average measures weighted by the amount of time respondents lived at each of their addresses between random assignment and May 31, 2008 (just prior to the start of the long-term survey fielding period). The AAVE and neighborhood characteristic measures are all in z-score form, standardized by the control group mean and standard deviation. The sample is non-Hispanic African-American youth (ages 13-20 as of December 2007) who were randomly selected for the linguistic component of the long-term survey from the low-poverty voucher and control groups only and whose speech samples included at least one analyzable language token (n=629).

Table S14. Control Group AAVE Use by Race/Ethnicity and Baseline Household Demographics

Outcome/Sample	Mean	N	Difference in Means (SE)	P-value Difference
A. Youth AAVE by Race/Ethnicity				
Hispanics (non-African-American)	0.414	135	-0.074*	0.016
African-Americans (non-Hispanic)	0.488	260	(0.030)	
B. Youth AAVE by Baseline Household Demographics				
Economic measures				
Baseline household income				
Median value or higher	0.457	131	-0.067*	0.034
Below median value	0.524	124	(0.031)	
Adult employment at baseline				
Adult is currently employed	0.464	60	-0.031	0.404
Adult is not currently employed	0.495	200	(0.037)	
Education measures				
Adult had high school diploma/GED at baseline				
Adult has diploma or GED	0.465	143	-0.052~	0.090
Adult has neither diploma nor GED	0.517	117	(0.030)	
Adult had high school diploma at baseline				
Adult has diploma	0.453	91	-0.054~	0.095
Adult does not have diploma	0.507	169	(0.032)	
Adult had college degree at long-term survey				
Adult has degree	0.387	31	-0.116*	0.004
Adult does not have degree	0.503	211	(0.040)	

Notes : * = $P<0.05$, ~ = $P<0.10$. The means presented are for the share of the speaker's tokens that used the African-American Vernacular English (AAVE) vs. Standard American English (SAE) variant of the language feature being analyzed, split by the race/ethnicity groups or adult/household outcome measure listed. The differences in means were calculated via bivariate regression of share AAVE tokens against an indicator for the baseline characteristic in the first row of each pair (weighted analysis with robust standard errors in parentheses). The sample is non-Hispanic African-American youth speakers (ages 13-20 as of December 2007) in the control group who were randomly selected for the linguistic component of the long-term survey and whose speech samples included at least one analyzable language token (N=260), except for the sample in top row of panel A, which includes control group youth who are Hispanic and not African-American (N=135), and the sample for the college degree measure, which is limited to the subset of control group youth in households where the sample adult was interviewed as part of the long-term survey with valid college degree data (N=242).

Table S15. Control Group Person-Level Regression Analysis Predicting Long-Term Survey Outcomes by Share AAVE Tokens

	Bivariate Regression		*Controlling for Age, Site and Gender*	
	Coefficient	**P-value**	**Coefficient**	**P-value**
Predicting Youth Outcomes Using Youth AAVE Rates				
Achievement test scores (control group z-scores)				
Math	-0.597 *	0.041	-0.689 *	0.034
Reading	-0.147	0.642	-0.393	0.258
Math/reading	-0.410	0.174	-0.601 ~	0.076
Grades earned				
Mostly Bs or better	-0.131	0.407	-0.177	0.323
Worse than mostly Cs	-0.002	0.981	0.118	0.252
4.0 scale	-0.113	0.639	-0.236	0.403
Experienced racial discrimination in past 6 months				
At school or work	-0.128	0.309	-0.233 ~	0.085
In a shop or restaurant	-0.360 *	0.012	-0.544 *	0.000
When meeting someone for the first time	-0.038	0.772	-0.087	0.588
When dealing with the police	-0.166	0.192	-0.287 ~	0.075
Experienced class discrimination in past 6 months				
At school or work	-0.110	0.395	-0.226	0.130
In a shop or restaurant	0.005	0.964	-0.059	0.643
When meeting someone for the first time	-0.085	0.444	-0.088	0.498
When dealing with the police	-0.141	0.131	-0.213 ~	0.071
Predicting Adult Outcomes Using Adult AAVE Rates				
Employment and earnings				
Adult currently employed	-0.122	0.418	-0.139	0.377
Adult's annual earnings (2009 dollars)	-12783 *	0.004	-12332 *	0.010
Education level as of long-term survey				
Adult has high school diploma	-0.204	0.185	-0.162	0.343
Adult has certificate of General Educational Development (GED)	-0.277 *	0.017	-0.268 *	0.037
Adult has high school diploma or GED	-0.456 *	0.002	-0.389 *	0.013
Adult has college degree	-0.225 *	0.004	-0.188 *	0.020

Notes : * = $P<0.05$, ~ = $P<0.10$. Coefficients are from a person-level, family-clustered, and weighted regression model using the share of the speaker's tokens that used the African-American Vernacular English (AAVE) vs. Standard American English (SAE) variant of the language feature being analyzed to predict the survey outcome measures listed. The bivariate regression models included only the AAVE variant measure and are contrasted with models that also included controls for age (continuous), site, and gender (youth only). The sample includes non-Hispanic African-American speakers in the control group who were randomly selected for the linguistic component of the long-term survey and whose speech samples included at least one analyzable language token (N=260 youth ages 13-20 as of December 2007 and N=337 adults).