Poverty and educational achievement in the U.S.: A less-biased estimate using PISA 2012 data

David Rutkowski[a*], Leslie Rutkowski[a], Justin Wild[b], and Nathan Burroughs [c]

* Corresponding author. Email: david.rutkowski@cemo.uio.no

[a] *Centre for Educational Measurement, University of Oslo, Blindern, Norway;* [b] *School of Education, Indiana University, Bloomington, USA;* [c] *College of Education, Michigan State University, Lansing, USA*

In the current paper, we employ the most recent Programme for International Student Assessment (PISA) data to calculate a less-biased estimate of poverty on U.S. achievement. The PISA was specifically chosen as it is an assessment removed from a specific curriculum and instead focuses on concepts that students should know in order to participate in a global economy. Using a propensity score matching approach, our findings suggest that U.S. students in poverty have notable educational attainment deficiencies compared to a matched group of students who are not in poverty. In other words, when we select two students who have a great deal in common but for the fact that one comes from a poverty background, the student in poverty is expected to perform nearly 28 points, or about a quarter of a standard deviation lower, on the PISA assessment. In real terms, this puts math achievement for children not in poverty on-par with the OECD average, while children in poverty are well below the OECD average.

Keywords: education policy; achievement; propensity score matching; poverty gap; poverty and economics

In 1964, U.S. president Lyndon B. Johnson set forth a plan in his State of the Union address that later became known as the “war on poverty”. Although poverty rates have dropped since Johnson’s speech, the absolute number of Americans living in poverty has risen from approximately 35 million in 1964 to nearly 47 million in 2014, with children under the age of 18 accounting for the largest age group (DeNavas-Walt and Proctor 2015). More troubling statistics can be found in a recent United Nations Children's Emergency Fund (UNICEF) report on the effects of the global recession and child poverty rates around the world. Using fixed reference points for wealth (rather than some percentage of the median income) to assess the absolute change in child poverty from 2008 to 2012, the authors show higher poverty rates in many affluent countries than normally reported by those countries. In the U.S., the number is staggering, with an estimated 32% of children living in poverty. And for families with children that fall below 50% of the poverty threshold (defined as extreme poverty), the data is even more troubling. The UNICEF Office of Research (2014, 3) states that extreme child poverty in the U.S. increased more during the last recession "than it did in the recession of 1982, suggesting that, for the very poorest, the safety net affords less protection now than it did three decades ago". In fact, among the Organisation for Economic Co-Operation and Development (OECD) countries, the U.S. has the 6th highest rate of child poverty–ranking between Bulgaria and Chile (OECD 2014a).

Historically, in the U.S. education has been regarded as one of the most important tools for alleviating poverty at both the individual and societal level (Labaree 1997). Inherent in this idea is that education can act as an equalizer between the rich and poor. For example, the pioneering public educator Horace Mann (1848, 418) stated that education "beyond all other devices of human origin, is a great equalizer of the conditions of men**. . . .** It does better than to disarm the poor of their hostility towards the rich: it prevents being poor". Such sentiments were echoed more recently by then-Secretary of Education Arne Duncan, when he stated, "in America, education is still the great equalizer" (Brenchley 2011). These statements are supported with empirical research that has shown a clear association between higher education levels and greater income and lower unemployment (Bureau of Labor Statistics 2015). A general relationship between poverty and lower educational academic performance is also well-established (Campbell and Ramey 1994; Entwisle and Alexander 1992; Kao and Thompson 2003; van der Klaauw 2008). Yet, as wealth gaps continue to widen in the U.S., more children face the stark reality of living in poverty. As such, more than ever, there is an increased importance in understanding the effects of material deprivation on educational achievement. In the current paper we attempt to answer this question by using the most recent Programme for International Student Assessment (PISA) data to establish less-biased estimates of poverty on achievement. The PISA was specifically chosen as it is an international assessment that is not dependent on a particular curriculum but instead measures knowledge and skills that students toward the end of compulsory education should know and can do in order to fully participate in a modern, global economy (OECD 2013). A unique feature of the PISA is that students are asked to not only reproduce what they have learned but to apply their knowledge and skills in unfamiliar settings, within and beyond the school. Understanding the association of poverty on this particular skill- and knowledge-set offers unique insights into the ways students from disadvantaged backgrounds are equipped to take the next step in their education or the world of work.

Literature Review

Despite the rich literature examining the relationship between socioeconomic status (SES) and student achievement, there remain a number of limitations in the existing body of research. As we discuss below, poverty estimates often rely on flawed measures (frequently relying on definitions of SES that are too narrow), neglect school-based inequalities associated with SES, and fail to account for important confounders like race, immigration status, pre-K education, and educational aspirations. To our knowledge, no other study estimates the effect of poverty on U.S. upper secondary students using a multidimensional indicator of SES and accounts for key covariates such as race.

For several decades, education researchers have recognized an association between child poverty and low student achievement (Campbell and Ramey 1994; Dahl and Lochner 2012; Entwisle and Alexander 1992; Kao and Thompson 2003; van der Klaauw 2008). Children of low SES begin school less prepared than their higher SES peers (Burchinal et al. 2008; Heckman 2005), with income playing an integral factor (Dahl and Lochner). Low SES students also maintain their low achievement (Sirin 2005). In fact, Caro, McDonald, and Willms (2009) showed that in Canada, the gap between lower SES and higher SES students grew over time. Studies indicate that poverty has both a direct effect on students, for example through cognitive deficits (Nelson and Sheridan 2011), as well as a direct effect through weaker instruction (Schmidt et al. 2015) and lower teacher quality (Goldhaber et al. 2015), among other factors (e.g. school resources, class size, etc.). International research suggests that SES achievement gaps are not just a U.S. phenomenon but pervade worldwide (Chudgar and Luschei 2009; Schmidt et al. 2015).

As a proxy for poverty status, U.S.-based studies frequently rely on eligibility for Free and Reduced Lunches (FRL), eligibility for which is set at 185% of the federally defined poverty line. However, the FRL measure has been criticized on a number of grounds, including: a high risk of misclassification, exclusive reliance on income over other forms of SES, large within-category differences, changes in eligibility guidelines, and the fact that FRL wasn't designed to be used as a measure for low SES (Harwell and Lebeau 2010; National Forum on Education

Statistics 2015; Recommendations to the National Center for Education Statistics 2012). Particularly problematic when analyzing student poverty and achievement, a plurality of students now has FRL status according to the National Assessment of Educational Progress (NAEP). However, as FRL is often the best available measure, the indicator continues to be used as a proxy for low SES (e.g. Jennings et al. 2015, Papay et al. 2015; Rondfelt et al. 2013; Kalogrides and Loeb 2013). Other studies have employed alternative measures of student SES, such as family income (Chmielewski and Reardon 2016; Reardon, 2011) and parental education (Bradbury et al. 2015), but, like FRL, these indicators focus on only one dimension of student disadvantage.

However, there is strong reason to believe that SES, and more specifically child poverty, is a multidimensional rather than unitary phenomenon. Family wealth, parental occupational status, and parental education all have distinct effects (Sirin 2005). Focusing on only one dimension could give a misleading picture of a child's true position. For example, two students could each have similar levels of family wealth but very different levels of parental education and, hence, very different relative advantages. One approach for addressing this issue is to employ indices that combine indicators representing multiple dimensions: household possessions, parental education, family income, and occupational status. The most prominent sources for these indices are the U.S. Early Childhood Longitudinal Study (ECLS), focused on younger children, and the OECD PISA, which tests 15-year-olds. The ECLS has been used to examine the Black-White achievement gap in kindergarten using SES as a control (Quinn 2015) and to estimate the difference in student performance between the top and bottom SES quintiles in kindergarten (Garcia 2015). The PISA's index of economic, social, and cultural status (ESCS) was used by Schmidt et al. (2015) to estimate achievement gaps between the top and bottom SES quartiles among all OECD countries.

We present a summary of the aforementioned research estimating SES achievement gaps in mathematics in the U.S. in Table 1. Whichever measure or population is considered, the basic pattern is the same: lower-SES students do substantially worse than their more advantaged peers. Notably, the gaps range from .8 to 1.3 standard deviations, despite very different thresholds (top and bottom quartiles, quintiles, deciles, etc.).

Table 1 Here

These studies all have their limitations, however. For example, except for NAEP, each estimate compares students in the most advantaged categories to those in the least advantaged, rather than comparing the performance of the least advantaged students to the rest of the student population. Only Schmidt et al. (2015) and Garcia (2015) use multidimensional indices of SES, which more accurately capture the complex construct. In many respects, Garcia's study most closely resembles our own approach, but it is explicitly focused on background estimates of SES upon students and is restricted to kindergarteners. It excludes school-based estimates that develop during the course of a student's school attendance. As a consequence, it fails to account for aggregate estimates of poverty on student outcomes, for example through tracking or differences in teacher quality.

Notably, only Garcia (2015) accounts for important correlates of poverty to include race, likely reducing statistical bias of the stated estimates. Especially in the U.S., race is identified as an important predictor of educational achievement (Hemphill and Vanneman 2011; Vanneman et al. 2009). In general, White and Asian students perform above their African American, Hispanic, and Native American peers (Kao and Thompson 2003; Lee 2002). Kao and Thompson (2003) provided an overview of empirical research into the racial and ethnic stratification in educational achievement, reviewing test scores, grades, educational aspirations, educational attainment, high school completion, and college completion. In their review, the authors found that African Americans, Hispanics, and Native Americans consistently perform lower on tests and achieve lower course grades when compared to their White and Asian peers in the U.S. In addition, the former groups have lower educational attainment and lower high school and college completion rates. In a study of NAEP trends over the past 30 years, Lee (2002) noted that the achievement gaps remain large and are a cause for concern among education policy makers.

Although race and SES are correlated, looking beyond SES, society and school environments are viable explanations of race-attributed differences, making it erroneous to conflate race and poverty. Census data indicates that Black (38%) and Hispanic (32%) students are more likely to be in poverty than White students (12%), but according to NAEP, White students account for nearly a third of all free and reduced lunch-eligible students in 8^{th} grade. Taken individually, achievement gaps by FRL and race are about one standard deviation in 8^{th} grade math, but the *conditional* estimate of SES on NAEP scores is quite different: a .7 standard deviation difference between FRL eligible and non-eligible Whites compared with a .5 deviation difference for Black and Hispanic students. Similarly, controlling for race, Garcia's (2015) high/low achievement gap is estimated at .84 standard deviations. Given these differences in conditional SES estimates, there appears to be some portion of achievement variance that is not common between these two variables, making race an important confounder to include in a model that focuses on SES-related achievement gaps.

There are a number of other important correlates of poverty available in the PISA data that can be used to develop a less-biased estimate of poverty gaps. Although race and immigration status are sometimes associated, immigration status has been separated as an area of interest and concern when it comes to educational achievement both in the U.S. and around the world (OECD, 2012b). In a recent study of immigrants from Asia, the Caribbean, and Latin America, Suárez-Orozco, Bang, and Onaga (2010, 6; see also Woessmann 2004) found that immigrant students face “transplant shock”, or culture shock, when placed into an education system that may be vastly different from their home country’s system, resulting in lower achievement.

Early childhood education is one cited contributing factor to better initial educational achievement among less economically advantaged students (Peterson 2014; The White House 2015). Similarly, several studies have demonstrated that early childhood education appears to increase future educational achievement (Campbell and Ramey 1994; Larsen and Robinson 1989; Lee et al. 1990). At the policy level, there is also evidence to suggest that school truancy and classroom tardiness among poor students are both associated with low achievement scores (Atkinson 2005; Baker, Sigmon, and Nugent, 2001; Epstein and Sheldon 2002; Marvul 2012). Unfortunately, issues around truancy and tardiness have often been found to be associated with poverty (Marvul).

Education aspiration has also been linked to academic achievement, especially among poor students (Kao and Thompson 2003; Kao and Tienda 1998; Mau, 1995). In particular, Kao and Tienda showed that over time the educational aspirations of low-income students fall as those students reach the end of compulsory education. The same authors also noted evidence supporting an association between educational aspiration and attainment in their review of racial stratification and achievement.

In sum, we believe that the magnitude of the educational effects of student poverty remains an open question. Although there are many studies that are focused on poverty, use rich measures, or account for important covariates like race and immigrant status, to our knowledge no other study includes all of these features.

Methods

In order for us to calculate an estimate of poverty on educational achievement, we use a propensity score matching (PSM) approach to estimate poverty-based achievement gaps. The basic theory behind PSM is to approximate the counterfactual by identifying a control group (not in poverty) that is as similar as possible to the treatment group (in poverty) through a matching procedure. PSM addresses the likely situation that poverty is not randomly assigned, but rather that demographics and other factors explain "selection" into poverty. Importantly, to get a less-biased estimate of the treatment effect, the most important predictors of poverty should be included in the matching procedure to account for observed differences (and, to a limited extent, unobserved differences) among students in and out of poverty. Given the constraints of the U.S. PISA background questionnaire, we control for differences on observed and latent characteristics measured by the questionnaire, namely race/ethnicity, preschool attendance, truancy/tardiness, educational aspirations, household structure (e.g. two- or single-parent households, parental occupational status), and immigration background. We subsequently elaborate on the methods and justify these variables.

Data

To estimate a less-biased poverty effect on achievement we used U.S. student achievement and background data from the 2012 cycle of the PISA. The PISA is a cross-sectional international large-scale assessment developed by the OECD to assess “the application of knowledge in mathematics, science, and reading literacy to problems within a real-life context” (OECD 2012a, 2) among a nationally representative sample of 15 year-olds in school. The PISA uses a complex two-stage clustered sampling design. In the first stage, approximately 150 schools are chosen based on a probability that is proportional to the school’s size (Rust 2014). Then, a random sample of 35 15-year-olds is selected for testing. In addition to an achievement test of math, science, and reading, students are administered a background questionnaire that asks about their lives and school situations, attitudes, and experiences in and outside of school (OECD 2014b). In total, students are tested for 120 minutes, with a subsequent 30 minutes allotted to filling out the background questionnaire. The unweighted U.S. PISA 2012 sample size was 8,609 students (Kastberg et al. 2014).

Measures

Outcome

The PISA is administered every three years and measures math, science, and reading literacy. In each cycle, one content area is considered the major domain (math in 2012) and is measured more thoroughly than the other two domains, which are considered minor domains (OECD 2014b). For our analysis, we used mathematics achievement as our outcome of interest, which is scaled using item response theory (Embretson and Reise 2000) to a mean of 500 and a standard deviation of 100 (OECD 2014b, 159). Because the PISA features complex booklet designs, the methods used to estimate achievement are similar to multiple imputation (Rubin 1987) and result in several plausible values for each student. These values are intended to be used as population estimates of achievement, rather than as individual estimates (Mislevy 1984; Mislevy et al. 1992). As recommended in the literature, we use all five plausible values of math achievement and combine them according to recommended practice (Rutkowski et al. 2010).

Poverty or treatment condition

The PISA 2012 database (Kastberg et al. 2014) includes the PISA index ESCS, which is a weighted linear combination of three item response theory-derived indices, including the highest occupational status of parents, highest educational level of parents in years of education, and a large number of home possessions (p. H-7).[1] This variable ranged from -3.80 to 3.12, and lower values are associated with poorer socioeconomic conditions. Unfortunately, this measure does not include more direct indicators of SES, such as annual family income or family wealth. Rather, ESCS relies on more indirect measures that can reasonably be asked of a 15-year-old who might not be able to accurately provide more direct information. Nevertheless, the ESCS measure has been used extensively as a proxy for SES (for two recent examples in top journals see Parker et al. 2016; Säälik, Nissinen, and Malin 2015). To create poverty categories, we used the UNICEF (2014) finding that 32.2 percent of U.S. children were living in poverty in 2012. Given the PISA's stringent sampling techniques, which ensure a representative sample of 15-year-olds in the U.S.,[2] we assigned a poverty indicator of *1* to all children at or below the 32.2 percentile on the ESCS index value. All other children were coded *0*, or not in poverty. We refer to this variable as the *treatment condition* throughout this paper. Although the PISA target population is one year older than the UNICEF-defined child, there is little reason to believe that

[1] For full information on what measures comprise this index, see chapter 16 of the *PISA 2012 Technical Report* (OECD, 2014b).
[2] For full information about how PISA samples 15-year-olds in the U.S., see chapter 2 of the *Technical Report and User Guide for the Program for International Student Assessment* (Kastberg et.al. 2014)

the poverty situation of a 15-year-old population in the U.S. would be substantially different from that of those under 15, particularly given that both populations are required to be enrolled in school.[3]

Covariates

To approximate an experimental setting, we used PSM, whereby the treatment and control groups are matched according to the probability of being in the treatment group (*in poverty*). An important feature of PSM is that the probability of being in the treatment group is estimated for all students, regardless of their *actual* poverty status. The treatment and control groups are thus balanced on the set of covariates, providing some assurance that observed differences in the outcome are not due to other factors.

Creating a propensity score involves specifying a logistic regression model from covariates that are associated with the treatment condition; however, there is some debate over model specification. Whereas some authors argue in favor of adding many covariates (Rubin and Thomas 1996), others argue that over-parameterization (Bryson, Dorsett, and Purdon, 2002) can increase the variance of the propensity scores, among other problems. As such, we opted for an approach that sought a wide spectrum of covariates supported by theory.

Due to the extant nature of PISA data and the fact that the 2012 cycle implemented a rotated background questionnaire (meaning that not all students completed all questionnaire items), we were limited by the available variables asked of all U.S. students in the background questionnaire. Furthermore, there are potentially important influences on student achievement

---

[3] The U.S. Census Bureau's more restricted definition of children poverty is currently 21% (Institute of Education Sciences 2016). In future work we hope to extend this analysis to additional countries, and so we have chosen to use UNICEF's metric for purposes of cross-national comparability.

that may also be related to student poverty that were not included in the PISA. We recognize this as one possible limitation to our study. In addition, we excluded school-based covariates such as opportunity to learn (OTL) from our model, in order to develop an estimate of the *aggregate* effect of poverty. The inclusion of OTL, school resources, and other variables could certainly influence the estimate, but is outside the scope of the present work.

Briefly, we included race/ethnicity, preschool and kindergarten attendance, grade repetition, truancy/tardiness frequency, academic aspirations, parental presence at home, full- or part-time work of parents, immigration status, and whether English is spoken at home. We dummy coded each of the covariates in our model. For example, race/ethnicity was re-coded into a series of dummy variables for Black, Hispanic, Asian, multiracial, or other, and White served as the reference category. A full list of variables and associated dummy codes are in Appendix 1A.

Analytic methods

Propensity score matching

We used PSM to isolate the effect of poverty while controlling for observed and, to a limited extent, unobserved differences between children in and out of poverty, such as truancy problems, single-parent home status, and educational aspirations. The method assumes that assignment to the treatment is conditionally independent of outcomes, given the observed covariates. When the assumption holds, we can estimate an unbiased average effect of poverty on achievement (McEwan 2010; Rosenbaum and Rubin, 1983; Steiner and Cook 2013).

Logistic regression models were fit to the treatment variables to create propensity scores on which to match. We included as covariates those student background variables listed in Appendix 1A. In an effort to generate a high-quality match, we explored several matching algorithms, including one-to-one and one-to-two nearest neighbor matching (NN and 2NN, respectively), with and without replacement, and with and without calipers (.010, .025, and .050). We used the package "Matching" (Sekhon 2013) in *R* 3.1.2 (The R Foundation 2015). Based on several criteria, discussed subsequently, we chose the 2NN matching algorithm with replacement and a caliper of .05 as our matching procedure.

To evaluate the quality of the matching procedure, we evaluated the *common support*, or regions where the control sample's propensity scores matched the treatment sample's propensity scores through graphical analysis of the samples' distributions, minima and maxima comparison, and trimming (Caliendo and Kopeinig 2008; see also Caliendo, Hujer, and Thomsen 2005). Further, we estimated the pseudo-R^2 from the propensity score regression with samples from the treatment and control groups before and after matching (Caliendo and Kopeinig; Sianesi 2004); we calculated the t-test for significant differences between the covariates of the control and treatment groups before and after matching; and we found the mean standardized bias (MSB) between covariates before and after matching (Caliendo and Kopeinig 2008; Rosenbaum and Rubin 1985). The MSB is a simple average over covariates of the standardized bias (SB), which is calculated as:

$$SB = 100 * \frac{(\bar{X}_1 - \bar{X}_0)}{\{0.5[\sqrt{(V_1(X) + V_0(X)}]\}}$$

where $\bar{X}_1$ ($\bar{X}_0$) is the mean of the covariate of the treatment (control) group, and $V_1(X)$ [$V_0(X)$] is the variance of covariate X of the treatment (control) group. The mean of all covariates' SBs was found before and after matching. As a critical final check, we conducted a sensitivity analysis to ascertain whether hidden bias due to unobserved heterogeneity might threaten the validity of results. In particular, we used Rosenbaum's approach to estimate the degree of departure from the assumption that no hidden bias exists in the estimate (Aakvik 2001; Caliendo, Hujer, and Thomsen, 2007; Rosenbaum 2002).

Estimate of average treatment effect

As noted, we are interested in achievement outcomes, Y, for a student, i, depending on the condition, D, of living in poverty (1) or not (0). The observed outcome for an individual can be written as:

$$Y_i = Y_i^1 \cdot D_i + Y_i^0 \cdot (1 - D_i)$$

where $D \in \{0, 1\}$, and Y^1 is the outcome of a student living in poverty and Y^0 is the outcome of a student not in poverty. The treatment effect for an individual is the difference between his or her potential outcomes $\Delta_i = Y_i^1 - Y_i^0$. Since we cannot simultaneously estimate outcomes based on both conditions for a single individual, we focus on the average effect between the control and treatment samples, known as the average treatment effect (ATE), expressed by:

$$\text{ATE} = E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$$

If the effect of living in poverty would be the same for all students, that is $E(Y^0 \mid D = 1) = E(Y^0 \mid D = 0)$, then we could use any students not living in poverty as a control group. However, since we do not have experimental data, the process of "selecting" oneself into or out of poverty is non-random, leading to the likely problem of selection bias. If we can reasonably assume that selection bias is due to observable characteristics, we can use those characteristics to find students in the control group (not living in poverty) who are similar to students in the treatment group (living in poverty) and estimate the ATE from these two groups in the sample (Caliendo, Hujer, and Thomsen 2005).

Results

Propensity score model

We fit a logistic regression with all covariates listed in Appendix 1A. Few variables thought to be associated with the latent characteristic were significant in the initial model. Therefore, we collapsed several variables, removed non-significant variables, and fit a second model to the data. Descriptive statistics for variables in the final model are presented in Table 2, and results from the model are presented in Table 3. We present odds ratios and a 95% confidence interval for ease of interpretation. The only non-significant variable is whether the student was born outside of the U.S. However, given its importance in the literature, we chose to leave this variable in the model.

Table 2 Here

Table 3 Here

Average treatment effect and matching quality

Results are presented in Table 4, which includes the ATE for the 2NN matching method with replacement and different calipers (all of which were statistically significant, with $p < .001$). We also include the estimate's standard error and the number of observations used. Based on our selected algorithm and the caliper with the fewest lost observation, we estimate the effect of poverty at 26.745 (*SE* = 1.542). In other words, children in poverty are expected to achieve, on average, 27 points lower on the PISA test than children not in poverty. This is more than one-quarter of a standard deviation on the PISA scale (OECD 2014b).

Table 4 Here

Although we do not include details here, the model used to match treatment and control observations exhibited reasonable common support based on histograms of covariates before and after matching. We also considered the percentage of treated individuals lost after matching from relevant covariate characteristics. The highest losses come from characteristics related to tardiness or truancy from school. A loss of 3% comes from students who aspire to less than an associate's degree. Overall, the treatment group lost only 1.1% of the sample, suggesting that the matched control group is comparable to most students in the treatment sample, and common support was found across the groups and their relevant characteristics. Table 5 presents the comparisons between the pseudo-R^2 values and the MSB for the covariates before and after matching. Systematic differences between the distribution of covariates after matching should be eliminated, so we expect the pseudo-R^2 values to be lower after matching, which is, indeed, the case. Likelihood ratio tests (LRTs) before and after matching drop as expected, although the joint significance of covariates remains after matching. Overall, these indicators point to a balanced quality matching between samples.

Table 5 Here

Results from the t-test for covariate balance between control and treatment groups also suggest improved matching balance (Table 6). Before matching, 13 of 15 covariates displayed significant differences between the control and treatment groups; after matching, four variables– two tied to the student's father's working status, whether the student is Black, and whether the student had been born outside of the U.S.–showed significant differences between the control and treatment groups. However, Imai, King, and Stuart (2008) suggested that the t-test may be a misleading indicator of individual covariate balance if used as evidence by itself. Steiner and Cook (2013) recommended plotting the standardized differences (or Cohen's *d*) of each covariate and its variance ratio before and after matching. They also call for checking for standardized differences different from zero, greater than magnitude .1 and variance ratios different from one, greater than magnitude .2.

Figures 1 and 2 suggest that the balance after matching is well within these parameters, with only one variable (three or four times truant in last two weeks) slightly outside the suggested range for the variance ratios. Please note that figure 2 is on a different scale to clearly show the magnitude of the change after matching. Since this departure generally aligns with the recommended bounds, we conclude that the balance after matching is satisfactory.

Table 6 Here

Figure 1 Here

Figure 2 Here

Sensitivity Analysis

Table 7 presents findings from the Rosenbaum bounds of both the Wilcoxon signed rank test and the Hodges-Lehmann point estimate using the "rbounds" package in *R* (Keele 2010). The Wilcoxon signed rank test suggests that our data is sensitive to an unobserved characteristic with an effect size of 1.7, and the Hodges-Lehmann point estimate suggests similar sensitivity with an effect size of 1.8. Translating these numbers into measures of the degree of departure from the assumption of no hidden bias in the estimate, the critical values are 5.47 for the Wilcoxon signed rank test and 6.05 for the Hodges-Lehmann point estimate. Both of these values are relatively distant from one, or no hidden bias, implying that only large magnitudes of hidden bias would alter our inference from the estimated effect of living in poverty. If, for example, the critical values were 1.05 or 1.10, even small magnitudes of hidden bias would affect any inference from the estimate.

Table 7 Here

In addition to Rosenbaum's bounds technique for sensitivity analysis, we performed a second test, which, while less formal, still supports the findings above. We split the control group before matching into two random halves and then implemented identical matching procedures–finding the propensity score regression, estimating the effect with a 2NN matching algorithm with replacement and a caliper of .05, and checking balance. Each half resulted in propensity score models that were nearly identical to the model in Table 3, with the only differences being in some proxy variables that measure the same latent characteristic as a variable in the original model (for example, a student having skipped *five* or more days in the past two weeks as opposed to *three* or more in the full matched dataset model). Next, a single propensity score model was used for each random half, and the treatment condition effect was estimated for each half, resulting in -28.535 and -26.404 with standard errors of 1.94 and 1.86, respectively. These findings are similar to the original findings of -26.745 with a standard error of 1.54 in Table 4. While the matching quality was not as high as the full matched dataset, estimating the effect with two random sub-samples did not show great departure, and this evidence combined with the sensitivity analysis above, suggests that our findings are relatively robust to unobserved hidden bias.

Discussion/Conclusion

Based on an analysis of PISA 2012 data, our findings suggest that, after matching on a set of covariates identified in the literature as important predictors of poverty status, students in poverty have notable educational attainment deficiencies compared to a matched sample of students who are not in poverty. In other words, when we select two students who have a great deal in common but for the fact that one comes from a poverty background, the poor student is expected to perform about a quarter of a standard deviation lower on the PISA assessment. This estimate is substantially lower than that reported by Garcia's (2015) early childhood study (.76) as well as the NAEP (2015) estimates controlling for race (.70 to .50), but is closer to the effect size (.35) in Sirin's (2005) meta-analysis.

We believe these results have a number of important implications with respect to future research, social equity, and educational and economic policy. First, we contend that the PISA is a powerful tool for discussions around poverty, given that the assessment is focused on the skills students need to participate in a global economy rather than an intended curriculum. As such, our findings suggest that poverty is, in and of itself, a catalyst for lower achievement, suggesting that a third of the U.S. population is less prepared to operate in the global economy. Commensurately, it is reasonable to expect that having fewer skills to participate in the economy would likely lead to lower earnings and a higher chance of living in poverty. This process intensifies the so-called *cycle of poverty*, a phenomenon in which once a family enters poverty, subsequent generations also live in poverty.

Second, these results highlight the distinction between poverty and race on student achievement. Although race and poverty are correlated, they are not substitutes. Black and Hispanic students have double the poverty rates of Whites (DeNavas-Walt and Proctor 2015), statistics that can make it difficult for educational researchers to separate race and poverty in achievement studies. Interestingly, we were able to match students in poverty to students who have similar attributes but who are not in poverty to examine difference in their scores. This statistical matching technique allowed us to estimate less-biased estimates of poverty, demonstrating that, beyond race, ethnicity, or immigrant status, poverty on its own contributes to substantively lower achievement. Although this finding may seem intuitive, our analysis provides a concrete and more accurate estimate of the negative effects of poverty on PISA achievement in the U.S. Importantly, this finding is worth highlighting given growing numbers of children in poverty. Furthermore, although we use race as a matching variable in our study, we do not intend to suggest that biological or ethno-cultural factors are responsible for achievement differences. Rather, race likely serves as a proxy for societal and other conditions experienced by racial and ethnic minorities in the US.

Our findings clearly speak to issues around social justice; however, they are also relevant to economic policy. For some time, researchers and economic organizations such as the OECD have suggested that increases in PISA scores have a positive effect on national economic growth (Hanushek and Woessmann 2010). Although claims about the actual effect of increased PISA scores on the national economy can vary, there is strong evidence that the effect is positive. As such, the U.S. potentially has a great deal to gain by reducing poverty rates, which are much higher than those of other OECD countries. In fact, a number of European countries have poverty rates that are less than half those of the U.S., including Germany, Sweden, Denmark, and Poland. Strikingly, the social safety net in these countries is generally well-developed in comparison to the U.S., pointing to one possible means to alleviate poverty. Additionally, our analysis of the PISA demonstrates that there is a great deal of untapped potential in the U.S., as measured by a meaningful poverty-based achievement gap. This potential is being stifled by poverty, beyond a host of other demographic variables, including race, ethnicity, immigration background, and other factors. In essence, 32% of the U.S. population is underperforming on the PISA test due to being in poverty. And if the PISA can be regarded as one indicator of the stock of skills that the up-and-coming workforce brings to the labor market and to higher education, clearly, this group is at a marked disadvantage. Given that wealth gaps in the U.S. continue to grow and poverty continues to rise (DeNavas-Walt and Proctor 2015; UNICEF Office of Research 2014), underperformance on the PISA will most likely increase over time, leading to lost opportunities and greater untapped potential. In a global knowledge economy, not exploiting this potential is a loss of efficiency in economic terms as well as a social injustice.

Our findings are also relevant to educational policy. Because students in poverty substantially underperform their peers even after accounting for race, immigration status, aspirations, and pre-K education, our results suggest that educational interventions must be focused on the specific effects of student poverty, not generic student underperformance. Schools in the U.K. have experimented with strategies that target support for poorer students within given schools, for example through the use of the “pupil premium.” These sorts of policies might also be promising in a U.S. context.

As analysts of secondary data, we no doubt suffer limitations to our study. First, the cross-sectional nature of the data limits what we can say about our findings and what we are able to measure. Although the procedure used for our analysis creates groups of students matched on important confounders of poverty, our poverty estimate might not be completely unbiased, as unobserved heterogeneity cannot be completely ruled out. In particular, the possibility remains that the PISA database does not include all important matching variables, leaving open the possibility of some persistent omitted variable bias. To that end, we included only covariates from the core block of the background questionnaire due to planned exclusion in the questionnaire design. Nevertheless, we followed best practice in model selection, robustness checks, and sensitivity analysis to avoid introducing unnecessary bias in our models (King and Nielsen 2016). Our robustness analysis showed stable estimates of poverty across a range of conditions, providing some assurance that our estimates of poverty on achievement are reasonably unbiased. Finally, we treated poverty as a dichotomous variable. As such, the poverty estimate compared the average achievement across these two groups. It is reasonable that at the extreme ends of the SES distribution, estimated achievement differences would be much larger. Nevertheless, these findings offer evidence that poverty has a meaningful impact on PISA results, pointing to a real need for addressing the widening SES gap in the U.S.

Notes on contributors

David Rutkowski is Professor of Educational Measurement at the Centre for Educational Measurement at the University of Oslo. His research is in the area of international large-scale assessment.

Leslie Rutkowski is Professor of Educational Measurement at the Centre for Educational Measurement at the University of Oslo. Her research is in the area of international large-scale assessment.

Justin Wild is a doctoral candidate at Indiana University studying Education Policy Studies, with a concentration in International and Comparative Education and Inquiry Methodology.

Nathan Burroughs is a Senior Research Associate at Michigan State University. His research focuses on the role of teachers and institutional structures in mediating educational inequality.

References

Aakvik, A. 2001. Bounding a matching estimator: The case of a Norwegian training program. *Oxford Bulletin of Economics and Statistics*, *63*(1), 115–143.

Atkinson, A. J. 2005. *Improving school attendance: A resource guide for Virginia schools* (p. 130). Richmond, VA: Office of Student Services, Virginia Department of Education.

Baker, M. L., Sigmon, J. N., & Nugent, M. E. 2001. *Truancy reduction: Keeping students in school. Juvenile justice bulletin.* (No. NCJ-188947) (p. 18). Washington, DC: Department of Justice: Office of Juvenile Justice and Delinquency Prevention.

Bradbury, B., Corak, M., Waldfogel, J. & Washbrook , E. (2015). *Too many children left behind*. New York: Russell Sage Foundation.

Brenchley, C. (2011, December 12). In America, education is still the great equalizer. Retrieved from http://blog.ed.gov/2011/12/in-america-education-is-still-the-great-equalizer/

Bryson, A., Dorsett, R., & Purdon, S. 2002. *The use of propensity score matching in the evaluation of labour market policies* (Working Paper No. 4). Department for Work and Pensions.

Burchinal, M. R., Roberts, J. E., Zeisel, S. A., & Rowley, S. J. 2008. Social risk and protective factors for African American children's academic achievement and adjustment during the transition to middle school. *Developmental Psychology*, *44*(1), 286.

Bureau of labor statistics. 2015. Earnings and unemployment rates by educational attainment. Retrieved January 20, 2016, from http://www.bls.gov/emp/ep_chart_001.htm

Caliendo, M., Hujer, R., & Thomsen, S. 2007. The employment effects of job creation schemes in Germany – a microeconometric evaluation. *Advances in Econometrics*, *21*.

Caliendo, M., Hujer, R., & Thomsen, S. L. 2005. Identifying effect heterogeneity to improve the efficiency of job creation schemes in Germany. *Applied Economics*, *40*(9), 1101–1122.

Caliendo, M., & Kopeinig, S. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72.

Campbell, F. A., & Ramey, C. T. 1994. Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, *65*(2), 684–698.

Caro, D. H., McDonald, J. T., & Willms, J. D. 2009. Socio-economic status and academic achievement trajectories from childhood to adolescence. *Canadian Journal of Education/Revue Canadienne de L'éducation*, *32*(3), 558–590.

Chmielewski, A. and Reardon, S. (2016). Patterns of cross-national variation in the association between income and academic achievement. *AERA Open* 2016(2): 1-27.

Chudgar, A., & Luschei, T. F. (2009). National income, income inequality, and the importance of schools: A hierarchical crossnational comparison. *American Educational Research*, *46*, 626–658.

Dahl, G. B., & Lochner, L. 2012. The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review*, *102*(5), 1927–1956.

DeNavas-Walt, C., & Proctor, B. D. 2015. Income and poverty in the United States: 2014, current population reports. *US Department of Commerce, Economics and Statistics Administration, US Census Bureau*.

Embretson, S. E., & Reise, S. P. 2000. *Item Response Theory for Psychologists*. Psychology Press.

Entwisle, D. R., & Alexander, K. L. 1992. Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review*, *57*(1), 72–84.

Epstein, J. L., & Sheldon, S. B. 2002. Present and accounted for: Improving student attendance through family and community involvement. *The Journal of Educational Research*, *95*(5), 308–318. Hanushek, E. A., & Woessmann, L. 2010. *The High Cost of Low Educational Performance: The Long-Run Economic Impact of Improving PISA Outcomes*. OECD Publishing.

Harwell, M. & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in educational research. *Education Researcher* 39(2): 120-131.

Heckman, J. J. 2005. The scientific model of causality. *Sociological Methodology*, *35*(1), 1–97.

Hemphill, F. C., & Vanneman, A. 2011. Achievement gaps: How Hispanic and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress. Statistical Analysis Report. NCES 2011-459. *National Center for Education Statistics*.

Imai, K., King, G., & Stuart, E. A. 2008. Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*(2), 481–502.

Kalogrides, D. & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Education Researcher* 42(6): 304-316.

Kao, G., & Thompson, J. S. 2003. Racial and ethnic stratification in educational achievement and attainment. *Annual Review of Sociology*, *29*, 417–442.

Kao, G., & Tienda, M. 1998. Educational aspirations of minority youth. *American Journal of Education*, *106*(3), 349–384.

Jennings, J.L., Deming, D., Jencks, C., Lopuch, M. & Shueler B.E. (2015). Do differences in school quality matter more than we thought? New evidence on educational opportunity in the twenty-first century. *Sociology of Education* 88(1): 56-82.Kastberg, D., Roey, S., Lemanski, N., Chan, J. Y., & Murray, G. 2014. *Technical report and user guide for the Program for International Student Assessment (PISA): 2012 data files and database with U.S.-specific variables* (No. NCES 2014-025). Washington, DC: NCES, IES, U.S. Department of Education.

Keele, L. 2010. *An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data* (White Paper) (pp. 1–15). Columbus, OH: Penn State University.

King, G., & Nielsen, R. 2016. *Why propensity scores should not be used for matching*. Working paper, Cambridge, MA.

Labaree, D. F. 1997. *How to succeed in school without really learning: The credentials race in American education*. Yale University Press.

Larsen, J. M., & Robinson, C. C. 1989. Later effects of preschool on low-risk children. *Early Childhood Research Quarterly*, *4*(1), 133–144.

Lee, J. 2002. Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher*, *31*(1), 3–12.

Lee, V. E., Brooks-Gunn, J., Schnur, E., & Liaw, F.-R. 1990. Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending head start, no preschool, and other preschool programs. *Child Development*, *61*(2), 495–507.

Mann, H. 1848. Report no. 12 of the Massachusetts school board. *The Republic and the School: Horace Mann on the Education of Free Men*, 415–449.

Marvul, J. N. 2012. If you build it, they will come a successful truancy intervention program in a small high school. *Urban Education*, *47*(1), 144–169.

Mau, W.-C. 1995. Educational planning and academic achievement of middle school students: A racial and cultural comparison. *Journal of Counseling & Development*, *73*(5), 518–526.

McEwan, P. J. 2010. Empirical research methods in the economics of education. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of Education* (pp. 9–14). San Diego: Elsevier Academic Press.

Mislevy, R. J. 1984. Estimating latent distributions. *Psychometrika*, *49*(3), 359–381.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. 1992. Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133–161.

NAEP (2015). *NAEP data explorer* [Computer Software]. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

National Forum on Education Statistics. (2015). *Forum guide to alternative measures of socioeconomic status in education data systems*. (NFES 2015-158). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Nelson, C. A. & Sheridan, M. (2011). Lessons from neuroscience research for understanding causal links between family and neighborhood characteristics and educational outcomes." In *Wither Opportunity?* Eds. G. J. Duncan & R. J. Murnane. New York: Russell Sage Foundation

OECD.(2012a. *PISA 2009 Technical Report*. Paris: OECD Publishing. Retrieved from http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/pisa2009technicalreport.htm

OECD. 2012b. *Untapped skills: Realising the potential of immigrant students.* Paris: OECD Publishing.

OECD. 2013. *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving, and financial literacy*. Paris: OECD Publishing. Retrieved from

OECD. 2014a. Child poverty. OECD Publishing.

OECD. 2014b. *PISA 2012 technical report*. Paris: OECD Publishing.

Papay, J. P., Murnane, R., & Willett, J. (2015). Income-based inequality in educational outcomes: Learning from state longitudinal data systems. *Educational Evaluation and Policy Analysis* 37(1S): 29S-52S

Parker, P. D., Jerrim, J., Schoon, I., & Marsh, H. W. 2016. A multination study of socioeconomic inequality in expectations for progression to higher education: The role of between-school tracking and ability stratification. *American Educational Research Journal*. https://doi.org/10.3102/0002831215621786

Peterson, M. 2014, 02. Gov. Pence visits Mishawaka, discusses Indiana preschool pilot. Retrieved May 30, 2015, from http://www.wndu.com/home/headlines/Local-interest-is-high-in-Indiana-preschool-pilot-253641181.html

Quinn, D. M. (2015). Kindergarten Black–White Test Score Gaps Re-examining the Roles of Socioeconomic Status and School Quality with New Data. *Sociology of Education*, *88*(2), 120–139.

Reardon, S. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations.” In *Wither Opportunity?* Eds. Greg J. Duncan & Richard J. Murnane. New York: Russell Sage Foundation

Recommendations to the National Center for Education Statistics (2012). *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: Theoretical foundations*. National Center for Educational Statistics. Washington, D.C.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal* 50 (1): 4-36.

Rosenbaum, P. R. 2002. *Observational studies*. New York: Springer.

Rosenbaum, P. R., & Rubin, D. B. 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, *45*(2), 212–218.

Rosenbaum, P. R., & Rubin, D. B. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38.

Rubin, D. B. 1987. *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.

Rubin, D. B., & Thomas, N. 1996. Matching using estimated propensity scores: relating theory to practice. *Biometrics*, *52*(1), 249–264.

Rust, K. 2014. Sampling, weighting, and variance estimation in international large-scale assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*. Boca Raton, FL: CRC Press.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. 2010. International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, *39*(2), 142–151.

Säälik, Ü., Nissinen, K., & Malin, A. 2015. Learning strategies explaining differences in reading proficiency. Findings of Nordic and Baltic countries in PISA 2009. *Learning and Individual Differences*, *42*, 36–43.

Schmidt, W.H., Burroughs, N. and Zoido, P. (2015) The role of schooling in perpetuating educational inequality: An international perspective. *Education Researcher* 44(7): 371-386.

Sekhon, J. S. 2013. Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software*.

Sianesi, B. 2004. An evaluation of the Swedish system of active labour market programmes in the 1990s. *Review of Economics and Statistics*, *86*(1), 133–155.

Sirin, S. R. 2005. Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453.

Steiner, P. M., & Cook, D. 2013. Matching and propensity scores. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 1, pp. 237–259). New York: Oxford University Press, Inc.

Suárez-Orozco, C., Bang, H. J., & Onaga, M. 2010. Contributions to variations in academic trajectories amongst recent immigrant youth. *International Journal of Behavioral Development*, *34*(6), 500-510.

The R Foundation. 2015. R: What is R? Retrieved January 17, 2016, from https://www.r-project.org/about.html

The White House. 2015, 18. Early learning. Retrieved May 30, 2015, from https://www.whitehouse.gov/node/174456

UNICEF. 2014. UNICEF Statistics and Monitoring. Retrieved December 8, 2014, from http://www.unicef.org/statistics/index_step1.php

UNICEF Office of Research. 2014. *Children of the Recession: The impact of the economic crisis on child well-being in rich countries* (Innocenti Report Card No. 12). Florence: UNICEF Office of Research.

van der Klaauw, W. 2008. Breaking the link between poverty and low student achievement: An evaluation of Title I. *Journal of Econometrics*, *142*(2), 731–756.

Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. 2009. Achievement gaps: How black and white students in public schools perform in mathematics and reading on the national assessment of educational progress. Statistical Analysis Report. NCES 2009-455. *National Center for Education Statistics*.

Woessmann, L. 2004. *How equal are educational opportunities? Family background and student achievement in Europe and the US* (CESifo Working Paper No. 1162). Munich, Germany: Ifo Institute – Leibniz Institute for Economic Research at the University of Munich.

Table 1. Summary of estimates of SES achievement gaps in mathematics in the U.S.

Study	Population	Measure	Comparison Groups	SES Gaps*
Chmielewski and Reardon 2016	Grade 8	Parental income	Top/bottom 10%	1.30
Reardon 2011	Grade 8	Parental income	Top/bottom 10%	1.02
Schmidt et al. 2015	15-year-olds	ESCS index	Top/bottom 25%	0.92
NAEP 2015 (author calculations)	Grade 8	FRL	Eligible (49%)/ Not eligible (44%)	0.82
Bradbury et al. 2015[5]	Grade 8	Parental education	College Ed (30%)/ High School (37%)	1.08
Garcia 2015	Kindergarten	ECLS index	Top/bottom 20%	0.96

*SES Gaps are in standard deviations

[1] Bradbury et al. estimates can be found in their Technical Appendix: https://www.russellsage.org/sites/all/files/Technical%20Appendix%20to%20Bradbury%20et%20al%202015.pdf

Table 2. Descriptive statistics of covariates in the propensity score model

	Not in Poverty (n=3299)	In Poverty (n=1616)
Black	421	215
Hispanic	471	696
White	2018	528
Attended preschool	2645	972
Was late to school 3-4 times in the past 2 weeks	101	141
Skipped 3 or more days of school in the past 2 weeks	70	82
Aspired level of education is an associate's degree or less	356	494
Aspired level of education is a doctorate degree	869	278
Mother lives at home	3107	1419
Father lives at home	2611	1076
Mother works full-time	2079	806
Mother works part-time	450	232
Father works full-time	2618	1000
Father works part-time	192	124
Student was born outside of the U.S.	180	183
English is the language at home	3069	1183

Table 3. Results from the propensity score model

Variable	Poverty (dummy coded) Odds Ratio	95% CI
(Intercept)	6.09*	[4.19, 8.83]
Black	1.52*	[1.23, 1.87]
Hispanic	3.26*	[2.75, 3.87]
Attended preschool	0.49*	[0.42, 0.57]
Was late to school 3-4 times in the past 2 weeks	1.90*	[1.42, 2.54]
Skipped 3 or more days of school in the past 2 weeks	1.88*	[1.29, 2.75]
Aspired level of education is an associate’s degree or less	2.78*	[2.33, 3.31]
Aspired level of education is a doctorate degree	0.61*	[0.51, 0.73]
Mother lives at home	0.55*	[0.44, 0.70]
Father lives at home	0.71*	[0.61, 0.84]
Mother works full-time	0.59*	[0.50, 0.69]
Mother works part-time	0.71*	[0.57, 0.89]
Father works full-time	0.47*	[0.40, 0.56]
Father works part-time	0.60*	[0.44, 0.80]
Student was born outside of the U.S.	1.00	[0.77, 1.31]
English is the language at home	0.37*	[0.30, 0.46]
Hosmer and Lemeshow pseudo R^2	0.197	
Cox and Snell pseudo R^2	0.221	
Nagelkerke pseudo R^2	0.308	

Note: N = 4915, CI = confidence interval.
*p < .001

Table 4. Poverty estimate with 2NN matching algorithm and various calipers

Matching Algorithm	Effect	S.E.	Total number of Obs.	Number of treatment observations dropped due to caliper
2 NN with replacement	-27.165	1.5447	4848	-
caliper .010	-26.737	1.5786	4392	152
caliper .025	-26.733	1.5431	4740	36
caliper .050	**-26.745**	**1.5422**	**4797**	**17**

Note: Bold text indicates the selected caliper used for reporting.

Table 5. Matching quality indicators

Variable	Before Matching	After Matching
Hosmer and Lemeshow pseudo R^2	0.197	0.009
Cox and Snell pseudo R^2	0.221	0.011
Nagelkerke pseudo R^2	0.308	0.016
LRT[1]	1230.1	54.8
Mean of standardized bias[2]	-3.25	1.40

[1]Both LRTs were significant at the $p < .0001$ level; the degrees of freedom were 4899 for the before-matching sample and 4781 for the after-matching sample.
[2]Mean of standardized bias has been calculated as an unweighted average of all covariates.

Table 6. t-test for mean standardized differences of each covariate

	Before Matching		After Matching	
	Std. Difference	p-value	Std. Difference	p-value
Black	1.598	.597	-4.568	.001
Hispanic	58.128	< .001	2.029	.072
Attended preschool	-40.893	< .001	-0.023	.985
Was late to school 3-4 times in the past 2 weeks	20.063	< .001	1.663	.348
Skipped 3 or more days of school in the past 2 weeks	13.448	< .001	2.292	.209
Aspired level of education is an associate's degree or less	42.917	< .001	-0.567	.629
Aspired level of education is a doctorate degree	-24.206	< .001	-0.266	.817
Mother lives at home	-19.466	< .001	-2.194	.165
Father lives at home	-26.621	< .001	1.897	.155
Mother works full-time	-26.278	< .001	-1.924	.089
Mother works part-time	2.041	.498	2.486	.059
Father works full-time	-35.972	< .001	3.999	.002
Father works part-time	6.961	.017	-5.711	.001
Student was born outside of the U.S.	18.512	< .001	-3.804	.038
English is the language at home	-44.744	< .001	1.222	.291

Table 7. Rosenbaum's bounded sensitivity analyses

	Rosenbaum Sensitivity Test for Wilcoxon Signed Rank P-Value		Rosenbaum Sensitivity Test for Hodges-Lehmann Point Estimate	
Gamma Values	Lower bound	Upper bound	Lower bound	Upper bound
1.0	0.000	0.000	-26.4	-26.4
1.1	0.000	0.000	-31.0	-21.8
1.2	0.000	0.000	-35.1	-17.6
1.3	0.000	0.000	-39.0	-13.8
1.4	0.000	0.000	-42.5	-10.3
1.5	0.000	0.000	-45.8	-7.0
1.6	0.000	0.000	-48.9	-4.0
1.7	0.000	0.168	-51.8	-1.1
1.8	0.000	0.919	-54.5	1.6
1.9	0.000	1.000	-57.0	4.1
2.0	0.000	1.000	-59.4	6.5

Note: Gamma is odds of differential assignment to treatment due to unobserved factors

Appendix

Appendix 1A: Dummy codes of student background questionnaire items, reference category in parentheses

Race variables (White)	
Q5RC_BL	Black
Q4RC_HIS	Hispanic
Q5RC_AS	Asian
Q5RC_MUL	Multicultural
Q5RC_OTH	Other/non-white
Q4_5RC_Miss	Race of student missing
Elementary education variables (did not attend preschool/kindergarten)	
Q6PS_1YR	Attend preschool for one year
Q6PS_OV1	Attend preschool for more than one year
Q6PS_Miss	Preschool variables missing
Q7KIN	Attend kindergarten
Q7KIN_Miss	Kindergarten variable missing
Age at first grade (4)	
Q8AG5	Began first grade at 5
Q8AG6	Began first grade at 6
Q8AG7	Began first grade at 7
Q8AG8	Began first grade at 8
Q8AG9	Began first grade at 9
Q8AG10	Began first grade at 10
Q8AG13	Began first grade at 13
Q8AG14	Began first grade at 14
Q8AG15	Began first grade at 15
Q8AG16	Began first grade at 16
Q8AG_Miss	Age of student missing
Repeat variables (did not repeat any grade)	
Q9RP_K1	Repeat kindergarten once
Q9RP_K2	Repeat kindergarten twice
Q9RP_K_Miss	Repeat kindergarten variable missing
Q9RP_G1TO6_1	Repeat any grades 1-6 once
Q9RP_G1TO6_2	Repeat any grades 1-6 twice
Q9RP_G1TO6_Miss	Repeat any grades 1-6 missing
Q9RP_G7TO9_1	Repeat any grades 7-9 once
Q9RP_G7TO9_2	Repeat any grades 7-9 twice
Q9RP_G7TO9_Miss	Repeat any grades 7-9 missing

Q9RP_G10TO12_1	Repeat any grades 10-12 once
Q9RP_G10TO12_2	Repeat any grades 10-12 twice
Q9RP_G10TO12_Miss	Repeat any grades 10-12 missing
Tardiness or truancy in the last two weeks (not truant or tardy in the last 2 weeks)	
Q10LT1_2	Late 1-2 times
Q10LT3_4	Late 3-4 times
Q10LT5	Late 5 or more times
Q10LT_Miss	Late to class missing
Q11SK_DY1_2	Skipped a day 1-2 times
Q11SK_DY3_4	Skipped a day 3-4 times
Q11SK_DY5	Skipped a day 5 or more times
Q11SK_DY_Miss	Skipped a day missing
Q12SK_CL1_2	Skipped a class 1-2 times
Q12SK_CL3_4	Skipped a class 3-4 times
Q12SK_CL5	Skipped a class 5 or more times
Q12SK_CL_Miss	Skipped a class missing
Academic level aspired to	
Q13LV_HS	High-school
Q13LV_VO	Vocational school
Q13LV_AS	Associate's degree
Q13LV_BA	Bachelor's degree
Q13LV_MA	Master's degree
Q13LV_DOC	Doctorate degree
Q13LV_Miss	Academic level aspired to missing
Who is at home	
Q14HM_MO	Mother
Q14HM_MO_Miss	Mother at home missing
Q14HM_FA	Father
Q14HM_FA_Miss	Father at home missing
Q14HM_BRO	Brother
Q14HM_BRO_Miss	Brother at home missing
Q14HM_SIS	Sister
Q14HM_SIS_Miss	Sister at home missing
Q14HM_GRA	Grandparent(s)
Q14HM_GRA_Miss	Grandparent(s) at home missing
Q14HM_OTH	Other people
Q14HM_OTH_Miss	Other people at home missing
Parent/Guardian type of work	
Q19MWK_FL	Mother has work full-time
Q19MWK_PT	Mother has work part-time
Q19MOTH	Mother has other type of work
Q19MWK_Miss	Mother's work missing
Q24FWK_FL	Father has work full-time

Q24FWK_PT	Father has work part-time
Q24FOTH	Father has other type of work
Q24FWK_Miss	Father's work missing
Born outside of the country	
Q25OR_ST	Student
Q25OR_ST_Miss	Student's birthplace missing
Q25OR_MO	Mother
Q25OR_MO_Miss	Mother's birthplace missing
Q25OR_FA	Father
Q25OR_FA_Miss	Father's birthplace missing
If born outside of US, age of arrival	
Q26AG_AR1	Age of arrival in U.S. is 1
Q26AG_AR2	Age of arrival in U.S. is 2
Q26AG_AR3	Age of arrival in U.S. is 3
Q26AG_AR4	Age of arrival in U.S. is 4
Q26AG_AR5	Age of arrival in U.S. is 5
Q26AG_AR6	Age of arrival in U.S. is 6
Q26AG_AR7	Age of arrival in U.S. is 7
Q26AG_AR8	Age of arrival in U.S. is 8
Q26AG_AR9	Age of arrival in U.S. is 9
Q26AG_AR10	Age of arrival in U.S. is 10
Q26AG_AR11	Age of arrival in U.S. is 11
Q26AG_AR12	Age of arrival in U.S. is 12
Q26AG_AR13	Age of arrival in U.S. is 13
Q26AG_AR14	Age of arrival in U.S. is 14
Q26AG_AR15	Age of arrival in U.S. is 15
Q26AG_AR_Miss	Age of arrival in U.S. is missing
Language at home (English)	
Q27LNG_SP	Language at home Spanish
Q27LNG_OTH	Language at home not English or Spanish
Q27LNG_Miss	Language at home missing