**Exposure to Classroom Poverty and Test Score Achievement:
Contextual Effects or Selection?**

Douglas Lee Lauen
Assistant Professor
Department of Public Policy
University of North Carolina at Chapel Hill

S. Michael Gaddis
Department of Sociology
University of North Carolina at Chapel Hill

Keywords: poverty, contextual effects, academic achievement, growth model, fixed effects, marginal structural model

Corresponding Author:
Douglas Lee Lauen
Department of Public Policy
University of North Carolina at Chapel Hill
Abernethy Hall, CB #3435
Chapel Hill, NC 27599-3435
919-843-5010
dlauen@unc.edu

Acknowledgements: The authors gratefully acknowledge the support of the Spencer Foundation and the North Carolina Education Research Data Center for providing the data. We thank Kyle Crowder, Patrick Curran, Mike Foster, Eric Grodsky, Ashu Handa, Roz Mickelson, Mike Shanahan, Chris Wiesen, and especially Stephen R. Cole for their assistance and many useful comments.

Abstract
Social scientists and policymakers generally share the widely held belief that impoverished contexts have harmful effects on children. Disentangling the influence of the effects of individual and family background from the effects of context, however, is conceptually and methodologically complex, making causal claims about contextual effects suspect. This study examines the effect of exposure to classroom poverty on student math and reading test achievement using data on a complete cohort of North Carolina children who entered third grade in 2001 and were followed up through grade eight. Using cross-sectional methods, we observe a substantial negative association between exposure to high poverty classrooms and math test scores that grows with grade level and becomes especially large for middle school students. Evidence from growth models, however, produces much smaller effects of classroom poverty exposure on math and reading test score achievement. Even smaller effects emerge from student fixed effects models, which control for time-invariant unobservables, and marginal structural models, which properly adjust for observable time-dependent confounding. These findings suggest that causal claims about the effects of classroom poverty exposure on cognitive achievement may be unwarranted.

Scholars have spent decades researching and debating the influence of school and neighborhood context on academic achievement, aspirations and attitudes (Alexander and Eckland 1975; Crosnoe 2009; Felmlee and Eder 1983; Rumberger and Willms 1992; Wilson 1959). The scholarly consensus is that high SES schools and neighborhoods positively affect individual academic outcomes (Brooks-Gunn et al. 1993; Entwisle, Alexander and Olson 1994; Willms 1986), whereas high poverty schools and neighborhoods negatively affect academic outcomes (Crane 1991; Harding 2003; South, Baumer and Lutz 2003). For example, Coleman and colleagues, in their seminal Equality of Educational Opportunity report, argued that peer effects were strong predictors of academic achievement: “the social composition of the student body is more highly related to achievement, independent of the student’s own social background, than is any other school factor” (Coleman et al. 1966: 325). Social science evidence on contextual effects has informed social science theory and educational policy in the United States, which for the past four decades has sought to mix students by racial background and, more recently, by poverty status (Bazelon 2008; Grant 2009; Kahlenberg 2001). The relevance of contextual effects research is demonstrated by the prominent role such research played in the recent social science statement submitted as an amicus curiae brief in a 2007 school assignment Supreme Court case.[1]

The scholarly consensus on contextual effects, however, rests largely upon cross-sectional studies, which do not provide a strong basis for causal inference. Selection bias, perhaps the most important threat to the validity of point-in-time studies, can give rise to what Hauser (1970) termed the "contextual fallacy": "…the contextual method rests on the arbitrary identification of residual group differences in the dependent variable with correlated aspects of group composition on an independent variable…The only way to eliminate such correlations is to assign individuals randomly to groups, and this is impossible with observational data" (p. 660). Recent work in sociology (Crosnoe 2009; Harding 2003) and in economics (Hanushek, Kain and Rivkin 2009; Hoxby and Weingarth 2005; Solon, Page and Duncan 2000) attempts to reduce bias in contextual effects through propensity score matching and weighting, comparison of sibling and neighbor correlations, fixed effects, instrumental variables, and natural experiments. Experimental evidence on the effect of changes in school and neighborhood context and academic achievement has emerged from the Moving to Opportunity program (Kling, Liebman and Katz 2007; Orr et al. 2003; Sanbonmatsu et al. 2006). Some of this recent work raises important questions about whether causal inferences about contextual effects are warranted (Mouw 2006). Finally, very few longitudinal contextual effects studies account for time-dependent confounding. Time-dependent confounders, which predict both future treatment and future outcome, conditional on past treatment, present a challenge to estimating unbiased treatment effects. For example, in estimating the effect of poverty context on child outcomes,

[1] Statement of American Social Scientists of Research on School Desegregation to the U.S. Supreme Court in "Parents v. Seattle School District" and "Meredith v. Jefferson County" Urban Review. V. 40. No 1.

one may wish to control for intermediate outcomes such as educational experiences while in school (such as assignment to gifted and remedial programs or being retained in grade). If these intermediate outcomes then predict both future treatment and future outcome, standard methods – controlling for these factors, omitting them, or controlling for baseline values – can produce biased estimates (Hong and Raudenbush 2008; Robins, Hernan and Brumback 2000). Methods for addressing treatment effect bias from time-dependent confounding have been developed in epidemiology by Robins and colleagues (Cole and Hernán 2008; Hernan, Brumback and Robins 2000; Robins 1999; Robins, Hernan and Brumback 2000). Recent work using these methods has demonstrated negative effects of exposure to neighborhood concentrated disadvantage on verbal ability (Sampson, Sharkey and Raudenbush 2008).

This study uses longitudinal data to estimate the effect of exposure to a high poverty classroom on elementary and middle school students' test scores. These data include interval metric and vertically equated mathematics and reading test scores and variation across time in classroom-level poverty from a complete cohort of public school children in grades three through eight in the state of North Carolina from 2001 to 2006 (N of more than 500,000 student-year observations). The study contributes to contextual effects research by carefully specifying and accounting for bias from omitted and mismeasured time-invariant student and family background characteristics. We report effects of classroom poverty based on three measures: attending a high poverty classroom (i.e., one in the top quartile of the classroom poverty distribution), cumulative exposure to a high poverty classroom, and continuous classroom poverty. We first present cross-sectional multilevel estimates of the association between classroom poverty and math test score. These estimates reproduce the negative effects reported in previous research with cross-sectional designs. The strength of the cross-sectional association increases with grade level. By eighth grade, these estimates are particularly large, which suggests that the cognitive disadvantage of classroom poverty exposure appears to accumulate over time. Growth models produce very small negative effects on two of the three measures (high poverty classroom and continuous classroom poverty) and larger negative effects on the other (cumulative exposure to a high poverty classroom). To address endogenous self-selection based on fixed unobservables, we present student fixed effects estimates, which remove between-student confounding (Allison 2009). This approach controls for time-invariant unmeasured and mismeasured aspects of student and family background that may predict both family choice of neighborhood and school and test score achievement. These models produce estimates distinguishable from zero, but of negligible size. We also estimate marginal structural models with inverse probability of treatment weighting to address time-dependent confounding (Hong and Raudenbush 2008; Robins, Hernan and Brumback 2000). These models produce non-significant effects on math and very small effects of classroom poverty on reading, which suggests that our estimates are robust to two different threats to validity.

The effects reported do not suggest that all children's life course outcomes are insensitive to classroom poverty, but they raise important doubts about the causal status of the effect of classroom poverty on student test scores among children and young adolescents, an implication which we discuss in our conclusion.

THEORY AND EVIDENCE ABOUT CONTEXTUAL EFFECTS

Drawing upon the theory and evidence from the contextual effects literature on school and neighborhood effects, we suggest four explanations specific of the effect of classroom poverty on student achievement growth for children and young adolescents (Harris 2010; Jencks and Mayer 1990; Willms 2010). First, classroom poverty may have a *negative effect* on student achievement growth due to *institutional* mechanisms: low parental involvement in schooling, lower quality teachers, lower expectations and slower pacing, and less rigorous curriculum (Barr and Dreeben 1983; Lee, Bryk and Smith 1993; Sedlak et al. 1986). Second, classroom poverty may have a *negative* effect due to *contagion* mechanisms: the downward leveling norms of predominantly low achieving peers (Crane 1991; Harding 2003; South, Baumer and Lutz 2003). Third, classroom poverty may have a *positive effect* due to *relative deprivation* mechanisms: the lack of competitive pressure and a lower average comparison group (Attewell 2001; Crosnoe 2009; Davis 1966). Fourth, classroom poverty may have *no effect* on student achievement growth once student background is properly controlled, which could point to a *selection* mechanism, i.e., that the apparent effect of context is due to the selection of families into schools and classrooms based on factors that are also correlated with test score growth and classroom poverty level (Hauser 1970; Mouw 2006).

In the next section, we summarize the cross-sectional contextual effects literature, organizing studies by the type of effects reported (i.e., positive effect of affluent context, negative effect of affluent context, no significant effect). We then discuss findings from alternative designs (longitudinal and experimental). To conclude our review we critique existing literature and outline the contributions of our study.

Cross-Sectional Evidence

Cross-sectional contextual effects research generally finds a positive association between socially desirable youth outcomes and average school and neighborhood socioeconomic status (SES). For example, studies find positive effects of school mean parental education on standardized test scores (Entwisle, Alexander and Olson 1994) and 4-year college enrollment (Choi et al. 2008), positive effects of school mean SES on grades and attainment (Willms 1986), and negative effects of the school mean poverty rate on academic self-esteem, educational aspirations and expectations, and standardized test scores (Battistich et al. 1995). Neighborhood effects research finds positive effects of high poverty neighborhoods on teenage pregnancy and high school drop-out rates (Crane 1991; Harding 2003), negative effects of early childhood neighborhood poverty on educational attainment measured in adulthood (Entwisle, Alexander and Olson 2005), and negative effects of neighborhood deprivation on educational attainment in Scotland (Garner and Raudenbush 1991). Similarly, low levels of neighborhood poverty have been associated with positive effects on educational attainment (Duncan 1994), positive effects on standardized test scores (Entwisle, Alexander and Olson 1994), positive effects on IQ, and negative effects on high school dropout rates (Brooks-Gunn et al. 1993). Finally, there is some evidence of positive additive effects of both high SES neighborhoods and high SES schools on earning a bachelor's degree (Owens 2010).

There is also evidence to support the hypothesis that affluent peers and neighbors can have *negative* effects on youth outcomes. Scholars posit that relative deprivation, sometimes referred to as the "frog pond effect," discourages and depresses the aspirations, achievement, and attainment of students in more affluent schools (Attewell 2001; Bachman and O'Malley 1986; Crosnoe 2009; Davis 1966; Jencks and Mayer 1990; Marsh 1987; Marsh and Parker 1984). Though it may be advantageous to associate with affluent neighbors and peers, high achieving peers may harm aspirations, grades, curricular placement, and other academic outcomes, especially when students must compete for scarce resources. For example, Davis (1966) investigated whether the theory of relative deprivation explained college student career and graduate school application decisions. His results indicate that school mean achievement may have a negative effect on career aspirations, suggesting that students in more competitive environments may remove themselves from contention for high status careers and graduate schools. Another study finds that students in elite public high schools suffer a competitive disadvantage in entering elite colleges due to the importance of class rank in the college admissions process (Attewell 2001). This disadvantage may produce an organizational adaptation to triage resources in favor of the top students. Therefore, students in high, but not the highest quantiles of class rank, may receive worse grades and take less advanced courses than they would if they had attended a less elite public high school (ibid).

On the other hand, peers may have little or no influence on individual outcomes. Contextual effects of classroom poverty and affluence may simply reflect self-selection (Evans, Oates and Schwab 1992; Hauser 1970; Leventhal and Brooks-Gunn 2000; Quigley and Raphael 2008). Important omitted and mismeasured family and student background characteristics may be causal determinants of both test score achievement and how individuals sort into neighborhoods and schools. Controlling for these factors may greatly reduce the unadjusted difference in outcomes between students from high and low poverty contexts. For instance, Alexander and colleagues investigate the nature of school effects and find that controlling for individual SES reduces the effect of school mean SES on college plans to near zero (Alexander et al. 1979). Their conclusion is that "the school SES influences are shown to result to a considerable degree simply from SES differences in the kinds of students attending various schools" (235). Cross-sectional research that controls for prior test scores or grades has reported relatively small and statistically insignificant contextual effects. In a study of high school students, Gamoran (1987) finds very minimal and mostly non-significant effects of school mean SES on test score outcomes in six subjects while controlling for prior achievement. The author incorporates mediators of the contextual effect, such as types of coursework and tracking variables, and concludes that within-school differences in opportunity to learn are more important than, and perhaps explanations for, contextual effects.

Alternative Designs of Contextual Effects

Much of the research discussed thus far employs cross-sectional designs, which ignore the cumulative nature of students' educational development and do not adequately control for self-selection bias. This section summarizes research from two strands of literature: studies with longitudinal designs and neighborhood relocation experiments.

A point-in-time study captures the effect of schooling in a focal year as well as the effects of prior educational experiences and student and family background. Reviews of the literature note the importance of controlling for exogenous factors (i.e., those that do not depend on type of neighborhood/school) and call for more longitudinal designs (Duncan and Raudenbush 1999; Galster et al. 2007; Harris 2010; Jencks and Mayer 1990; Saporito and Sohoni 2007). Rumberger and Palardy (2005) examine the effect of school SES composition on test score growth in high school with NELS, a nationally representative database. They use a three-level growth model (time within student within school), finding that the predictive power of school SES on composite test score growth is as strong as family SES (.12σ effect size for individual SES and a .11σ effect size for school SES). As the authors note, these effects on a standardized composite test score mask important differences across different subjects. Effects of school SES on test score growth in math and reading are relatively small (.05σ and .06σ, respectively), while effects in science and history, perhaps because of differential opportunity to learn these subjects in low SES high schools, are larger (.21σ and .14σ, respectively). Another contribution of this study is showing that the effect of school SES is explained by teacher expectations, the amount of homework students do, course taking, and student perceptions of school safety. Although this study uses an impressive array of control variables to adjust for *observable* differences in student populations that could confound the school SES effect, its design does not permit ruling out bias from the sorting of students into schools based on *unobservables*. It also does not account for the problem of time-dependent confounding, which could arise if a student's school SES is a function of lagged values of school SES and lagged values of the outcome.

The gold-standard for addressing unobservables in contextual effects research is an experimental design (Kling, Liebman and Katz 2007; Sampson 2008). Although no experiment conducted to date allows for direct examination of school contextual effects, evaluations of a housing relocation program, Moving to Opportunity (MTO), provide suggestive evidence about the impact of changes in both neighborhood and school context (see DeLuca and Dayton 2009 for a review of this research). The MTO experiment randomly assigned participants to three groups: a treatment group that was provided a Section 8 voucher and allowed to move without restrictions, another treatment group provided a rental assistance voucher but allowed to move only to a census tract with less than 10% poverty, and a control group offered no voucher to move. While early results indicated a number of positive academic outcomes for the treatment groups, later follow-up studies found that these positive results dissipated. Children in the treatment group showed no academic improvement, except for black children's reading test scores, and were in only marginally better schools than before the switch (Kling, Liebman and Katz 2007; Orr et al. 2003; Sanbonmatsu et al. 2006).

To our knowledge only two studies of poverty context account for time-dependent confounding in modeling effects on children's cognitive outcomes. In the first, Sampson, Sharkey, and Raudenbush (2008) examine the effect of changes in neighborhood concentrated disadvantage on children's verbal ability across three waves of African American families in the

Project on Human Development in Chicago Neighborhoods study. To address the problem of time-varying confounding, this authors estimate a marginal structural model (MSM) with inverse probability of treatment weighting (IPTW) and report that the effect of neighborhood concentrated disadvantage on children’s verbal ability is large and negative, equivalent to missing a year of school (ibid). In the second, Sharkey and Elwert (2011) also estimate a MSM with IPTW and find that multigenerational neighborhood poverty has a negative effect on children's cognitive ability.

In summary, existing research on school contextual effects rests primarily on a base of cross-sectional designs of correlational evidence. One study of school contextual effects employs a longitudinal design, but ignores the problem of unobserved heterogeneity and time-dependent confounding. Housing relocation studies provide evidence about changes in neighborhood, which also involve changes in school context, but suffer from limitations of generalizability to non-poor and non-minority populations and leave unexamined the effects of increases in classroom poverty. Two studies from the neighborhood effects literature use appropriate techniques to address time-dependent confounding and report negative effects of neighborhood poverty on children’s cognitive ability.

The present study makes a contribution to existing research on school contextual effects by employing a rigorous longitudinal research design. First, we estimate a quadratic growth curve models over six years (grades three through eight) that relate the effect of changes in classroom poverty to changes in students’ test score achievement. Second, we address selection bias by including student fixed effects into our growth model. A large literature in economics and a growing literature in sociology (e.g. England, Allison and Wu 2007; Jacobs and Carmichael 2001; Jacobs and Tope 2007; Kocak and Carroll 2008; Mouw 2003; Schneiberg, King and Smith 2008) uses fixed effects methods to control for time-invariant unobserved heterogeneity.[2] These models, which require treatment variation within units over time, remove confounding bias that can emerge from omitted observable, mismeasured, or unobservable time-invariant student or group characteristics (Allison 2009; Halaby 2004; Mouw 2006; Wooldridge 2003). In the present context, this technique accounts for important student-level confounders such as low birth weight, early childhood education, and genetic factors, as well as family-level confounders such as parental IQ and class background. Third, following Sampson, Sharkey, and Raudenbush (2008) and Sharkey and Elwert (2011), we account for time-dependent confounding by producing growth model estimates with inverse probability of treatment weighting.

Unlike most prior school contextual research, we measure classroom poverty at the classroom level rather than the school level, which, due to the non-random sorting of students to classrooms and middle school tracking based on achievement level, may produce less valid estimates of classroom poverty effects. We measure classroom poverty three ways: attending a high poverty classroom (i.e., in the top quartile of the classroom poverty distribution), cumulative exposure to a high poverty classroom, which more accurately reflects the time-varying exposure to context over a youth's life course, and continuous classroom poverty (defined as percent receiving free or reduced priced lunch). We examine the effects of both increases and decreases in classroom poverty among a diverse population of students enrolled in the North Carolina public school system (a population that includes in large numbers whites, blacks, Hispanics, non-poor and poor students in urban, suburban, and rural locales).[3] Finally, we

---

[2] Here fixed effects refers to the panel data technique of using differencing or including indicator variables to control for unit-specific effects, not fixing random effects to zero or fixed (versus random) coefficients in random effects model.

[3] Relative to national statistics, blacks are overrepresented in North Carolina public elementary and secondary schools (30.4% vs. 17.2% nationally) and Hispanics are underrepresented (5.0% vs. 19.8% nationally).

focus on elementary and middle school aged student test score growth for two reasons: 1) the effects of classroom poverty on younger students is relatively understudied, and 2) the effect of classroom poverty has been shown to be stronger for cognitive and achievement outcomes than for behavioral and health outcomes (Duncan and Brooks-Gunn 1997).

DATA

This project uses test score and related data for one cohort of public school students in North Carolina beginning in grade three in 2001 through grade eight in 2006. North Carolina is a particularly appropriate setting for this analysis because it is one of the few states to consistently administer comparable tests over this time period, with scores produced from a three-parameter logistic item response theory (IRT) model[4] and scored on a developmental scale to allow computation of growth across grade levels.[5] The sample includes more than 500,000 student-year observations, beginning with about 100,000 third graders in 2001. By 2006, we observe about 75% of the original sample as being enrolled in a public school in North Carolina.[6] We analyze both reading and math, but to conserve space, we will present descriptive analysis of only math results. Math scores for students in grades three through eight range from 303 to 388, with an average of 350.8 and a standard deviation of 11.8 (table 1). By the end of third grade, the average

The percentage of whites in the North Carolina system closely mirrors the national percentage (59.5% vs. 57.1%). National data from the 2007 Digest of Education Statistics, table 40; North Carolina data from our sample.

[4] Item response theory is the technique pioneered by the Educational Testing Service and used to create the National Assessment of Educational Progress (NAEP). It involves measuring skill by simultaneously taking into account the ability of the test taker and various aspects of item characteristics (difficulty, ability to discriminate high and low skill, and guessability).

[5] The test is anchored at grade three and re-centered on grade five. The math test was rescaled in 2006. To compute growth scores for the state's accountability system the North Carolina Department of Public Instruction (NCDPI) conducted equating studies to permit conversion of scores across time. These studies, which used equipercentile equating, produced concordance tables to convert old scores to the new metric for the purposes of the state's accountability calculations. This study used these concordance tables to convert scores to a consistent metric.

[6] We retain all students in the cohort regardless of grade retention or promotion status. Students become censored from the sample due to leaving the public school system for in-state private schools and schools in different states. Due to the age of the sample (third through eighth graders), we suspect that very few are school drop outs, but we have no way of verifying drop out status with the data available for this study.

student math score is 339; by the end of 8th grade it is 360, suggesting a linear growth rate of about 4.2. This average masks the relatively large increases in the elementary grades (6-7 points per grade) and relatively small increases in middle school grades (3 points per grade). To define high poverty classroom, we begin by standardizing the mean level of a student's classroom peers' free/reduced lunch status by grade. Consistent with prior research (Sampson, Sharkey and Raudenbush 2008) we dichotomize this variable into a variable coded 1 if a student is in the top quartile of classroom poverty and 0 if a student is in the bottom three quartiles of classroom poverty.[7] Classroom is defined as the group of students with whom the student took their math test in each year.[8] Similar to recent research on neighborhood effects (Crowder and South 2011; Jackson and Mare 2007; Wodtke, Harding and Elwert 2011), we also derive an alternate measure of classroom poverty designed to better capture cumulative effects, which we call *cumulative exposure* to a high poverty classroom. This time-varying variable measures the proportion of years up to and including the current year a student has attended a high poverty classroom: $\sum_{t=1}^{T} \frac{HPC_{ti}}{T}$. Thus, a student can be coded only 0 or 1 during third grade but can be coded 0, 0.5, or 1 during fourth grade. An eighth grader coded as 0 was never exposed to a high poverty classroom, while one coded as 1 was always exposed to high poverty classrooms. Descriptive statistics in table 1 indicate that, on average, students in our cohort spend 24% of their third through eighth grade years in high poverty classrooms. About half of 8th graders were never exposed to a high poverty classroom; only 5% of 8th graders were always exposed. Since we would expect the effect of the contrast between never exposed and always exposed to be larger

[7] Using an absolute definition of peer poverty, coded 1 if 75% or more of a student's classroom peers are classified as free or reduced lunch and 0 otherwise, rather than a relative one, does not affect our conclusions about the effect of peer poverty exposure on test score growth (results from authors upon request).

[8] Classrooms with five or fewer students (less than two percent of the student-year observations) were dropped from the analysis.

than the contrast between high and low poverty classroom at one point in time, the cumulative exposure measure provides perhaps the strongest possible test of the contextual effects hypothesis with longitudinal data. To be consistent with research using a continuous measure, we also report results from classroom percent free or reduced price lunch eligible. For ease of exposition, below we will refer generically to the construct of classroom poverty to encompass all three measures, distinguishing among them when needed.

Classroom poverty is time-varying rather than fixed because 1) students can be assigned to classrooms with varying poverty composition over time, 2) students change schools due to residential changes and school choice, and 3) students make structural school enrollment changes (i.e., those arising from policy-induced school mobility due to how grade configurations are structured, chiefly changing from an elementary to a middle school, rather than family choices). Measuring classroom poverty at the classroom level rather than the school level permits within-school variation in classroom poverty to contribute to estimates. There is considerable variation in classroom poverty both within and between schools. School average classroom poverty rates range from 0% to 100%, with an average of 50% and a standard deviation of 24%. About 75% of total variation in classroom poverty rates lies between elementary schools, while 25% of variation is between classrooms within schools. Perhaps due to early tracking, the portion of variation that lies between classrooms in middle schools is larger, at 40%, leaving 60% between schools.

Control variables available for this study are race/ethnicity, gender, family poverty status, and parental education; educational designations as gifted, special education, or limited English proficient; whether the student was ever retained in grade; and structural and non-structural school transitions. Family poverty (free/reduced lunch eligibility) is a time-varying covariate because student free and reduced lunch eligibility changes from year to year due to changes in family income. For the population used in this study, the family poverty level of the student changes at least once for about 15% of the students. School mobility is separated into structural and non-structural measures based on whether a school switch was mandated by school district policy (a structural move like moving from elementary to middle school) or was a result of family choice or residential mobility (a non-structural move). We impute missing values for covariates at time *t* by assigning the subject-specific panel average. For example, if a student has a missing value in their panel of the family poverty indicator, we impute the average of that student's family poverty indicators across their other panels. For the dependent variable, math test score, we drop subjects whose panels contain less than half non-missing scores and then impute with the grade level average of students who were ever missing, since students who were ever missing had lower test scores than kids who were never missing. A table of means before and after imputation for analysis variables is shown in appendix table A1.

METHODS

Cross-Sectional Model

To reproduce cross-sectional estimates commonly reported in previous research, we begin by presenting point-in-time estimates of the association of classroom poverty on student achievement from a multilevel model (students nested within classrooms). We model math achievement, *A*, for student *i* in classroom *j* as a function of classroom poverty, *Z*, and **X**, a vector of student covariates which includes student's own family poverty status:

$$A_{ij} = \beta_0 + \beta_1 Z_i + \boldsymbol{\gamma X_{ij}} + \mathrm{u_{0j}} + \varepsilon_{ij} \tag{1}$$

In (1), we include a random intercept for each classroom, u_{0j}, and estimate (1) by grade level to examine whether the effect of classroom poverty varies by grade. The classroom poverty estimate from (1) could be considered causal if **X** contains all confounders of the effect of Z on A, if these confounders are measured without error, and if the random effects are uncorrelated with each other and the covariates in the model. These conditions would apply if $E(u_{oi}|Z_i) = 0$ and $E(\varepsilon_{ij}|Z, \boldsymbol{X}_{ij}, u_{oi}) = 0$. For example, many contextual effects studies, including the present one, have no or poorly measured information about the quality of students' early childhood education. If students with high quality early childhood education experiences are less likely to enroll in high poverty classrooms, we would expect β_1 to be downwardly biased; that is, if we controlled for the quality of early childhood education, the hypothesized negative effect of attending a high poverty classroom would be closer to zero than the unadjusted estimate.

Growth Model

Using test score data that are interval scaled and vertically equated to allow for growth modeling, we estimate a quadratic growth model with random intercept and slopes. Researchers in sociology, psychology, education, and criminology often use multilevel methods to account for within-subject inter-correlation, a wide range of covariance structures, and empirical Bayes estimation, which weights estimates by their reliability (the ratio of the true score variance to the observed score variance) (Bryk and Raudenbush 1987; Singer and Willett 2003). We formulate our quadratic growth model as:

$$A_{ti} = \beta_0 + \beta_1 Grade_{ti} + \beta_2 Grade_{ti}^2 + \beta_3 \mathrm{Z}_{ti} + \boldsymbol{\theta XT}_{ti} + \boldsymbol{\gamma X}_i + u_{0i} + u_{1j} Grade_{ti} + u_{2j} Grade_{ti}^2 + u_{3i} Z_{ti} + \varepsilon_{ti} \quad (2)$$

This model regresses a math achievement test score, A, at time t for student i on grade level, grade squared, a classroom poverty indicator, a vector of time-varying covariates, $\boldsymbol{XT}$, and a vector of time-invariant covariates, $\boldsymbol{X}$, with all covariates grand mean centered. Due to the problem of time-varying confounding, we omit from $\boldsymbol{XT}$ variables that could be affected by prior treatment status such as school mobility, and assignment to gifted, special education status, limited English proficiency, and grade retention. The model allows the intercept and slopes of *Grade*, *Grade*^{2}, and *Z* to randomly vary and the variance-covariance matrix, Σ, imposes no restrictions on the covariation of these random effects (i.e., the matrix is specified as unstructured). We also estimate a model that interacts variables in ***X*** and ***XT*** with *Grade* and *Grade*^{2} to ensure our estimates are not biased by differential growth rates across different subpopulations of students.

Random effects models such as the growth model shown in equation (2) produce a precision-weighted least-squares estimate that depends on within- and between-student variance components (σ_e^2 and σ_u^2, respectively) and the average number of periods per student ($\bar{T}$). In a generic panel regression of *y* on *x*, both sides of the equation are quasi-demeaned with a weighting parameter, λ:

$$(y_{it} - \lambda\bar{y}_i) = \beta_0(1-\lambda) + \beta_1(x_{it} - \lambda\bar{x}_i) + (e_{it} - \lambda\bar{e}_i), \text{ where} \tag{3}$$

$$\lambda = 1 - \sqrt{\frac{\sigma_e^2}{\sigma_u^2 + \bar{T}\sigma_e^2}} \tag{4}$$

As $\sigma_e^2 \to 0$, $\lambda \to 1$ and the random effects estimate converges toward the fixed effects estimate, discussed below. As $\sigma_u^2 \to 0$, $\lambda \to 0$ and the random effects estimate converges toward the pooled OLS estimate (Wooldridge 2003). Typically $0 < \lambda < 1$, with the random effects estimate falling between the pooled OLS and the fixed effect estimates.

The coefficient of interest in this model is β_3, the average effect of classroom poverty on achievement across grades three to eight. Parameters estimated with model (2) are unbiased and efficient assuming that given the covariates, the random effects and the student-level residual, ε_{ti}, are normally distributed with zero mean, are independent of one another, with the random effect independent across subjects and ε_{ti} independent across subjects and occasions.[9] The growth model produces an unbiased estimate of the effect of classroom poverty on test score growth if classroom poverty is uncorrelated with the random effects, if Z and the variables in $\boldsymbol{XT}$ are exogenous,[10] and if family background is adequately controlled and well measured.

As with the cross-sectional model, omitted variable bias could produce inconsistent parameter estimates, which could threaten the validity of this model. Although multilevel models can increase efficiency due to the use of both within and between variance, such models provide no solution for this type of confounding bias. If the between-student effects of classroom poverty are large relative to the within-student effects, it is possible that the omission of student and family background characteristics could bias estimates of classroom poverty contextual effects. In thinking about bias, it is helpful to return to our explanations of classroom poverty effects: contagion, relative deprivation, collective socialization, and institutions. Classroom effects can emerge either because students affect each other or because adults in schools affect students. The former pertains to contagion and relative deprivation explanations; the latter to an institutional or collective socialization explanation. In either case, the validity of inferences about contagion or institutional effects depends on removing the confounding effects of student and family

[9] We specify growth models as two-level models, occasions within subjects, rather than occasions nested within subjects cross-nested in classrooms due to computational limitations, our focus on the parameter estimates rather than the random effects, and the fact that with population level data efficient estimates of standard errors are a secondary concern. Moreover, simulation evidence suggests that ignoring cross-nesting is likely to affect the variance components and not the parameter estimates (see Luo, Wen, and Oi-man Kwok. 2009. "The Impacts of Ignoring a Crossed Factor in Analyzing Cross-Classified Data." *Multivariate Behavioral Research* 44(2):182-212.).

[10] In studies of peer achievement effects, but not classroom poverty effects, the direction of causality may be quite difficult to determine because student achievement at time *t* and peer achievement at time *t* are simultaneously determined (see Manski, Charles F. 1993. "Identification of Endogenous Social Effects - The Reflection Problem." *Review of Economic Studies* 60(3):531-42.). Determining the causal direction between classroom poverty and achievement is more straightforward. We posit that classroom poverty affects student achievement and that student achievement at time *t* does not affect classroom poverty at time *t*. This seems like a reasonable assumption given that student academic performance has no bearing on their parents' earning power.

background. We address the threat of adverse selection based on time invariant family and student background characteristics with a student fixed effects specification.

Student Fixed Effects Model

We estimate a student fixed effects model to control for fixed unobservables such as innate ability, mother's IQ, and early childhood experiences that might confound the effect of classroom poverty on test score. The fixed effects formulation uses each student as his/her own baseline, which holds constant all observable, unobservable, and mismeasured time-invariant student and family background characteristics. This approach eliminates all time-invariant between-student confounding in the classroom poverty effect and produces consistent parameter estimates when there is no within-student confounding of the classroom poverty effect. The student fixed effects model is specified as:

$$A_{ti} = \beta_0 + \beta_1 Grade_{ti} + \beta_2 Grade_{ti}^2 + \beta_3 \mathrm{Z}_{ti} + \boldsymbol{\theta X T}_{ti} + \alpha_i + \varepsilon_{ti} \quad (5)$$

Here we treat the subject-specific intercept as a fixed unknown parameter to be estimated, with α_i representing the deviation of subject i's intercept from the mean intercept β_0 with $\sum_{i=1}^{I} \alpha_i = 0$. This model is often estimated by "demeaning" both sides of the equation by the subject's panel mean, which removes between-student confounding by using only within-subject variation to estimate parameters. Omitted from equation (5) are time invariant covariates because these have no within-subject variance and are therefore not estimable with this approach (though their effects are subsumed into the subject-specific intercept).

The student fixed effects approach requires within-student variation on classroom poverty to identify parameters and is relatively inefficient relative to the random effects models. Due to its large sample size and the six-year panels within it, however, our data are well suited to this approach. We identify the classroom poverty effect from year-to-year variation in the poverty composition of students' classrooms. This changes due to school mobility and due to variations in the poverty compositions of student's assigned classrooms as they progress through grade levels in the same school. Because classroom poverty rates vary more between schools than within schools, school movers are somewhat more likely to experience a change in classroom poverty than students who remain in the same school. Nearly the entire sample makes some sort of school move during their panel: 85% of students make a structural move (e.g., moving from elementary to middle school), 35% of students make a non-structural move (e.g., moving due to residential mobility), and 91% of students make either a structural move or a non-structural move or both. The evidence suggests that across time variation exists to analyze for both school stayers and school movers, but that a larger portion of the variation that is analyzed appears to come from movers.

In total, about two-thirds of students either move into or out of a high poverty classroom at least once during their panel. About 17% of students make a change into or out of a high poverty classroom during the 3rd to 4th, 4th to 5th, 6th to 7th, or 7th to 8th grade transitions, whereas 20% of students make one of these changes during the 5th to 6th grade transition (a shift from elementary to middle school for most students in the sample). These changes are evenly split: 52% are changes into a high poverty classroom and 48% are changes out of a high poverty classroom. On average, the changes into a high poverty classroom are a grade-to-grade increase of 22% in peer poverty and the changes out of a high poverty classroom are a grade-to-grade decrease of 28% in peer poverty. Students who do not change on the binary high poverty classroom variable on average have a grade-to-grade decrease of 1.4% in peer poverty.

Estimates from model (5) can be considered causal assuming that selection into high poverty classrooms is based only on time-invariant unobservables. The model does not adjust for unobserved time-varying exogenous factors that could be related to attending a high poverty classroom. We must assume strict exogeneity, that for each *t*, the expected value of the idiosyncratic error given the explanatory variables in all time periods and the student fixed effect is zero: $E(\varepsilon_{ti}|\boldsymbol{X}_i, \alpha_i) = 0$, where **X** is a vector containing all variables appearing on the right hand side of equation (5).

Marginal Structural Model

Both the multilevel and fixed effects models outlined above are vulnerable to the threat of time-varying confounding, which arises when there is a time varying variable that is affected by prior treatment and is associated with subsequent treatment and the outcome. For example, consider the causal diagram in figure 1. In this diagram, X is a time-varying control variable, Z is treatment (high poverty classroom), Y is outcome (test score achievement), the subscript 0 represents baseline variables, 1 the subsequent time period variables, and U is an unobservable that affects both X_1 and Y_1. The variable in the shaded box, X_1, is a time varying confounder (e.g., assignment to gifted or special education), because it predicts future treatment, Z_1, and is associated with future outcome, Y_1, via U and directly (Robins, Hernan and Brumback 2000). Because X_1 is affected by prior treatment through the prior outcome (i.e., endogenous), standard models will produce biased treatment effect estimates.

Time-varying confounding presents a dilemma: X_1 is a confounder for later treatment and thus must be controlled, but may also be affected by earlier treatment and thus cannot be controlled (Robins, Hernan and Brumback 2000). Because X_1 is a collider (it is an effect of both Y_0 and U), controlling for X_1 introduces collider stratification bias (Cole et al. 2010; Greenland 2003; Hernán, Hernández-Díaz and Robins 2004). Marginal structural modeling (MSM) fit using inverse-probability-of-treatment (IPT) weighting can account for time-varying confounding and produce asymptotically consistent estimates of treatment effects in longitudinal analysis. This approach involves first computing IPT weights from each subject's probability of having their own treatment history and second estimating an IPT-weighted regression model. Our MSM is a weighted version of the growth model shown in equation (2).

Following standard practice, we compute stabilized weights which have lower variance than non-stabilized weights (Robins, Hernan and Brumback 2000):

$$IPTW_{ti} = \prod_{k=0}^{t} \frac{P[Z_k = z|Z_{k-1}, Z_0, Y_0, \boldsymbol{X}_0, G_k]}{P[Z_k = z|Z_{k-1}, Z_0, Y_{\mathrm{k}-1}, Y_0, \boldsymbol{X}_{k-1}, \boldsymbol{X}_0, G_k]} \tag{6}$$

where *t* indexes time, *i* indexes student, $Z_k = z$ is treatment actually received (classroom poverty exposure), *Y* is outcome, ***X*** is a vector of time-invariant and time-dependent confounds, and *G* is grade level, variables subscripted with a 0 represent baseline values, and variables subscripted with *k-1* are one period lags. In ***X*** we include student background characteristics (race/ethnicity, gender, family poverty status, parental education), school mobility variables, and academic classifications (gifted, special education, Limited English Proficiency, and grade retention). The denominator is, informally, a student's conditional probability of receiving her own observed treatment up to time *t*, given past treatment, outcome, covariate history, and grade level. The numerator is, informally, a student's conditional probability of receiving her own observed treatment up to time *t*, given past treatment, baseline outcome, baseline covariates, and grade level. This technique is a generalization of propensity score methods for longitudinal data. Rather than weighting inversely proportional to the probability of receiving treatment (*Z*=1), we instead weight inversely proportional to the probability of the treatment status actually received (*Z*=*z*). We truncate weights at the first and ninety-ninth percentiles by recoding observations above the 99th percentile to the 99th percentile weight and recoding observations below the 1st percentile to the 1st percentile weight (Cole and Hernán 2008).

Our MSM models also adjust for possible bias due to selective attrition (Hernan, Brumback and Robins 2000). We compute a stabilized censoring weight as:

$$CW_{ti} = \prod_{k=0}^{t} \frac{P[C_k = 0 | C_{k-1} = 0, Z_k, Z_{k-1}, Z_0, Y_0, \boldsymbol{X}_0, G_k]}{P[C_k = 0 | C_{k-1} = 0, Z_k, Z_{k-1}, Z_0, Y_{\mathrm{k}-1}, Y_0, \boldsymbol{X}_{k-1}, \boldsymbol{X}_0, G_k]} \tag{7}$$

where *t*, *i*, *Z*, *Y*, $\boldsymbol{X}$, and *G* are defined above, and C_k is an indicator for student became censored at time *t* (i.e., last observation in student's panel). To adjust for both the inverse of the probability of treatment and censoring, the weight used in the MSM models is the product of the IPT and censoring weights ($IPTW_{ti} * CW_{ti}$).

The IPT weighted version of the model shown in equation (2) produces a consistent estimate of treatment effect under the assumption of no unmeasured confounders or sequential strong ignorability (that treatment assignment is conditionally independent of the current and future potential outcomes given the measured past).

RESULTS

We begin by discussing descriptive analysis of the difference in the medians and distribution of test scores by grade and classroom poverty composition. We then summarize results from cross-sectional models which show substantial associations between classroom poverty and student test score, especially in the middle school grades. Following this, we present random coefficient growth model estimates and two alternative specifications with student fixed effects and inverse-probability-of-treatment weighting for our three measures of classroom poverty: exposure to high poverty classroom, cumulative exposure, and continuous classroom poverty. We then present a summary table of effect sizes and confidence intervals for all results.

Figure 2 displays a box plot of the distribution of math test scores by grade level and the high poverty classroom measure. The plot shows a general upward trend in scores and a reduction in the inter-quartile range for both groups as students increase in grade level. The gap in median test scores between students in high and low poverty classrooms in third grade is six points; by eighth grade it is seven points. The slight widening of the gap is more noticeable among eight graders always and never exposed to a high poverty classroom (the cumulative exposure measure), with the gap in median test scores growing from six points to ten points between third and eighth grade (results not shown, but available from authors upon request).

Cross-Sectional Estimates of Classroom Poverty and Cumulative Exposure

We produce cross-sectional estimates from the students-within-classrooms multilevel random intercept model shown in equation (1), which control for (but do not display) race, gender, parental education, family poverty status, gifted, special education, limited English proficiency status, grade retention, and structural and non-structural school mobility. We estimate models separately by grade level. Net of controls, the third grade cross-sectional effect of attending a high poverty classroom is -.877 scale points, or a standardized effect size of -.082σ (-.877/10.74=.082). Between fifth and sixth grade, this effect jumps from -.133σ to -.233σ. By eighth grade this estimate has grown to -.280σ. If causal, these results could reflect the differentiation of math curriculum in middle school (Dauber, Alexander and Entwisle 1996; Hallinan 1992) and the growing influence of peers for young adolescents (Buchmann and Dalton 2002; Furman 1982; Veronneau and Dishion 2011), or the cumulative nature of cognitive disadvantage through the life course of children and young adolescents. Replacing the high poverty classroom variable in equation (1) with the our cumulative exposure variable produces even larger point estimates, with the estimate growing from -.082σ to -.429σ between third and 8th grade (not shown, but results available from authors upon request). In summary, we find substantively large cross-sectional associations of high poverty classroom and cumulative exposure to high poverty classrooms that increase with grade level and become especially large by eighth grade, suggesting that test score trajectories may widen over time between students exposed to higher versus lower poverty classrooms. As we have argued above, however, a growth model is a more appropriate model to estimate the gap between higher and lower poverty classrooms than a cross-sectional one. A random coefficients growth model produces precision-weighted trajectories, which provide much more convincing evidence of the effect of context on student achievement than a point-in-time estimate.

Growth Model Estimates of High Poverty Classroom

Table 3 displays coefficients from time-nested-within-student random coefficient growth models for math (see equation 2, above).[11] Column 1 shows an unadjusted effect of high poverty classroom of -2.249. Including grade and grade^2 reduces the classroom poverty effect by about 70%. The coefficient shrinks from -2.249 to -.683 due to the fact that including grade level (a within-student parameter) greatly reduces unexplained within-student variation (σ_e) while leaving between-student variation (σ_u) essentially unchanged. As discussed above, when within-student variation falls between models 1 and 2, λ increases, which pulls the classroom poverty effect closer to the fixed effect estimate, which as we will see below, is quite close to zero. At the average of grade, the average student has a math test score of 352.5, an instantaneous growth rate of 4.166, and a negative curvature parameter of -0.651, which indicates that students' rate of

[11] Alternative models using age instead of grade produce similar results (available from authors upon request).

change in test score growth declines over time. Including covariates in column 3 further reduces the classroom poverty effect from -.683 to -.413, suggesting that the classroom poverty effect is due in part to compositional differences across classrooms. Most control variable coefficients conform to expectations, with negative effects for minorities, males, and poor students, and positive differences for students with highly educated parents. The classroom poverty effect indicates that the predicted gap between students in high and low poverty classrooms across grades three through eight is -.413 scale points, or .035σ. In column 4, we include interactions between all student background controls and grade and grade.[2] Allowing the effects of student background to vary with grade level does not significantly alter our treatment effect estimate. Table 4 shows these same models with reading test score as the outcome. Similar to the effects on math, the coefficient drops from -2.275 scale points in the unadjusted model to -0.399 scale points (.031σ) in the fourth model.

Fixed Effect and MSM Growth Model Estimates of High Poverty Classroom

Growth model estimates are unbiased assuming all confounders are controlled. A student fixed effects model controls for fixed pre-baseline unobservables such as innate ability, early childhood experiences, and mother's IQ that might confound the effect of classroom poverty on test scores. By using only within-student variation in classroom poverty, this approach eliminates all time-invariant between-student confounding and produces unbiased parameter estimates when all time-varying confounders are controlled. The strength of the student fixed effects model is adjustment for time-invariant unobservables. A weakness of both the growth model and the student fixed effects specification is that neither appropriately adjusts for time-dependent confounding. A marginal structural model (MSM) with inverse-probability-of-treatment weighting (IPTW) is designed to address time-dependent confounding. Table 5 presents in column 1 estimates from the primary coefficients of interest from the random effects growth model shown in tables 3 and 4, model 3, student fixed effects estimates in column 2 (see equation 5), and MSM with IPTW estimates in column 3 (the random effects growth model in column 1 estimated with the weights computed by equations 6 and 7).

In both math and reading, the absolute values of the student fixed effects estimates and the MSM estimates are much smaller than the random coefficients estimates. The effect of high poverty classroom on math in the MSM model is not statistically different from zero and the effect on reading is only significant at the $p<0.05$ level. Both of these alternative specifications produce estimates that are less than one percent of a standard deviation in effect size in both math and reading.

Growth Model Estimates of Cumulative Exposure to Classroom Poverty

To address the concern that our estimates presented thus far could potentially underestimate the effect of classroom context by ignoring the cumulative nature of such effects, in table 5 we present growth model estimates with cumulative exposure to high classroom poverty as the explanatory variable. The random effects growth model predicts a -2.037 point gap in math between students who up to a point in time were always and never exposed to high poverty classrooms. This represents 0.172σ of the math test score, a fairly large effect, which is due in part because it is an average of effects across grades three through eight. As shown in the cross sectional results, the effect of cumulative poverty in third grade are much smaller than effects in eighth grade. (By eighth grade, an always-exposed student has been in a high poverty classroom for six years, whereas an always-exposed student in third grade has been exposed only once.) The student fixed effects model, on the other hand, produces only a -0.192 point gap (0.016σ) in math between students who up to a point in time were always and never exposed to high poverty classrooms. The difference in the two estimates suggests that the large effect produced by the random coefficients model is largely due to baseline differences in students who become exposed to particularly high and low levels of classroom poverty. The MSM model produces a high poverty classroom gap of that is not significantly different from zero (0.0267), suggesting that time-dependent confounding is downwardly biasing the estimate in column 1. The pattern in reading is largely the same with a large negative effect from the random effects growth model (0.161σ) shrinking closer toward zero in the fixed effects (0.023σ) and MSM specifications (0.022σ). The unexpected result is that the sign of the cumulative poverty effect is positive in the fixed effect and MSM specifications rather than negative.

Growth Model Estimates of Continuous Poverty

For consistency with prior research that uses continuous measures of poverty context, we present the effect of a standardized measure of continuous classroom poverty (measured as percent of classroom that is on free or reduced priced lunch) in table 7. We present estimates from the random coefficients and fixed effects models.[12] In column 1, a one standard deviation increase in classroom poverty produces a 0.297 decrease in math test score (-.025σ). The estimate from a student fixed effects model produces an effect with the opposite sign, but much smaller in absolute value (+.003σ). In reading, the signs of the effects from the random coefficients and fixed effects specifications are also opposite signed, and both are approximately the same size, producing effect sizes of -.002σ and +.017σ, for a one standard deviation increase in classroom poverty.

[12] Due to the complexity of calculating IPT weights for continuous treatments, we do not present MSM results for standardized classroom poverty.

Effect Size Summary with 95% Confidence Intervals

Most of the reported effects are statistically significant, yet small in substantive terms. Our claim that the contextual effects found in this study are small would fail if confidence intervals contain values that could be considered large. Due to the large sample size used in this study, however, confidence intervals are quite narrow (table 8). For example, in panel A of table 8 we show that 95% confidence interval of standardized fixed effects high poverty classroom estimates (0 to 1 contrast) lie between -.009 and -.002 in math and .005 and .012 in reading. The MSM confidence intervals are also tightly arranged around zero. Panel B shows that the fixed effect cumulative poverty effects (0 to 1 contrast) in math and reading lie within the range of -.026 (lower bound for math) and .033 (upper bound for reading); the MSM results for the cumulative poverty effect lie between -.006 and .031. For continuous classroom poverty we report effects from two contrasts: a one and two standard deviation increase in classroom poverty. To put these contrasts into perspective, in our sample 32% of students experience no more than one instance of an increase in classroom poverty of 25%, which represents about one standard deviation in classroom poverty. A two standard deviation change represents a much more rare event. Only five percent of students have no more than one instance of an increase in classroom poverty of 50%. The effects for a one standard deviation effect from the random coefficient and fixed effects specifications lie within the range of -.027 to +.019, whereas the effects for a two standard deviation effect from these two specifications lie within -.054 to +.039. Given that most effects reported from the student fixed effects and MSM models are smaller than $.04\sigma$ in absolute value, we conclude that the contextual effects of classroom poverty on cognitive achievement are very small.

CONCLUSION

For decades scholars from a variety of disciplines have been debating the evidence of contextual effects on youth outcomes. What do we add to this rich literature? This study moves beyond a conception of contextual effects as correlations estimated on young people at one point in time to context shaping youth development over time. This study is designed to test the hypothesis that classroom contexts with high levels of poverty harm student achievement. Our findings challenge previous research based on cross-sectional designs which tend to report negative effects of peer poverty on student achievement. With cross-sectional models, we replicate past research by establishing that exposure to high poverty classrooms is negatively associated with math test score, with the strength of the association becoming quite large as students increase in grade level. The growth model evidence presented, however, shows very small negative effects of exposure to a high poverty classroom and continuous classroom poverty on math and reading test scores. The effects on cumulative exposure to a high poverty classroom, though, are larger (one-sixth of a standard deviation). Models with student fixed effects, which control for time-invariant student background unobservables, and with inverse-probability-of-treatment weighting, a method to properly adjust for observable time-dependent confounding, produce negligible effects of all three measures of classroom poverty on math and reading achievement. That exposure to classroom poverty has a strong association with test score in the cross-section, but has very small effects in models with weaker assumptions for causal inference, strongly suggests that selection bias is present in the cross-sectional estimates reported in studies based on point-in-time designs. The question of whether the effect of school poverty is causal or simply a function of either omitted variable bias or endogenous self-selection is a critical conceptual and empirical matter for both the theory of school effects and policies that seek to integrate students by socio-economic background (Duncan and Raudenbush 1999).

This study has important implications for both research and public policy. These findings suggest that standard estimates and prevailing theories about social influence among pre- and early-adolescents may not hold for test score achievement, one of the most important educational outcomes. This suggests that simply mixing students by poverty level and not altering important institutional resources such as high quality instruction and teacher expectations may not have the intended effect of increasing achievement because achievement is not simply a function of poverty context but of student and family background. The policy goal of mixing students by race-ethnicity or social background has been a mainstay in educational policy since the Brown vs. Board decision. Since the 1980s, school desegregation orders have been vacated by an increasingly conservative judiciary. The changing legal landscape has contributed to a resegregation of American schools (Orfield, Eaton and Desegregation 1996; Reardon and Yun 2005; Rumberger and Palardy 2005). These trends are likely to continue given that in the 2007 case of *Parents Involved v. Seattle School District No. 1*, the Supreme Court ruled that school districts may not use race in assigning students or granting transfers to achieve or maintain school integration. In response to increases in school racial segregation and the Supreme Court's prohibition on the use of race in making school assignments, some advocate for integrating students based on socio-economic background (Kahlenberg 2001), which is constitutionally permissible. Kahlenberg (2001) argues that the best way to ensure the presence of high standards, highly qualified teachers, and less crowded classes is to ensure a critical mass of middle class families to advocate for these resources. Various forms of SES integration have been implemented in more than 50 districts across the U.S., including Lacrosse, WI; Wake County, NC; Cambridge, MA; and San Francisco, CA. The findings of the present study suggest that simply mixing students by social background may not have the intended effects, unless such mixing can garner increased resources and support for proven teaching practices that can increase student achievement in impoverished contexts.

There are some limitations to this study that point the way for future work on school and classroom poverty effects. Although North Carolina is racially and economically diverse, the study covers only the public school students from one state, which limits the generalizability of our findings. Using population-level administrative data, we have pursued an identification strategy that privileges reduction of bias over national representativeness. The external validity of these results will hinge on cross-state replications using administrative data, and large nationally representative surveys with rich contextual information and interval metric test scores designed to measure growth over time. We must be careful to stress that we use a research design that reduces, but may not entirely eliminate, bias from unobservables. For example, our inability to account for time-varying student or school unobservables could prove these estimates to be biased. Future work should carefully theorize and measure time-varying factors that predict test scores. We cannot empirically examine whether changes in classroom poverty correspond to substantial differences in micro-level interaction between students and between students and teachers. An important next step for school contextual effects research is to examine the effect of school or classroom poverty on test score growth. Although these models suggest negligible effects of classroom poverty on average test score achievement across students' third through eighth grade panels, it is possible that school or classroom poverty negatively deflects students' growth rates. Moreover, contextual effects may vary by a number of demographic characteristics such as race, individual poverty status, or gender (Clampet-Lundquist et al. 2011; Legewie and DiPrete 2012). Finally, test scores may be mostly impervious to the influence of peers and socialization processes. Other outcomes such as pregnancy, drug use, school completion, and college attendance, may be more amenable to these factors than a test score, which is a discrete cognitive task rather than a behavioral event.

Much of the existing research base on contextual effects has examined the experiences and outcomes of high school students. This study represents one of the first sociological examinations of contextual effects among elementary and middle school students. Despite this contribution, it may be that by third grade, the earliest time point in this study, early childhood experiences have largely determined a student’s potential for achievement. If test score gaps among socioeconomic groups are essentially stable by third grade and variations in school quality have little effect on these gaps over time (Heckman 2006), then policies to mix students by social background may be of limited utility. On the other hand, research has found neighborhood effects on birth weight and other early childhood development experiences (e.g. Chase-Lansdale et al. 1997; Masi et al. 2007; Morenoff 2003). Therefore, a promising avenue for future research may be the investigation of younger children's sensitivity to impoverished contexts, preferably with research designs that permit accounting for unobserved family or individual heterogeneity and time-dependent confounding.

REFERENCES

Alexander, Karl, and Bruce K. Eckland. 1975. "Contextual Effects in High-School Attainment Process." *American Sociological Review* 40(3):402-16.

Alexander, Karl L., James Fennessey, Edward L. McDill, and Ronald J. D'Amico. 1979. "School SES Influences - Composition or Context." *Sociology of Education* 52(4):222-37.

Allison, Paul D. 2009. *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage.

Attewell, Paul. 2001. "The Winner-Take-All High School: Organizational Adaptations to Educational Stratification." *Sociology of Education* 74(4):267-95.

Bachman, Jerald G., and Patrick M. O'Malley. 1986. "Self-Concepts, Self-Esteem, and Educational-Experiences - The Frog Pond Revisited (Again)." *Journal of Personality and Social Psychology* 50(1):35-46.

Barr, Rebecca, and Robert Dreeben. 1983. *How Schools Work*. Chicago, IL: University of Chicago Press.

Battistich, Victor, Daniel Solomon, Dong-il Kim, Marilyn Watson, and Eric Schaps. 1995. "Schools as Communities, Poverty Levels of Student Populations, and Students Attitudes, Motives, and Performance - a Multilevel Analysis." *American Educational Research Journal* 32(3):627-58.

Bazelon, Emily. 2008. "The Next Kind of Integration." in *New York Times Magazine*: New York Times.

Brooks-Gunn, Jeanne, Greg J. Duncan, Pamela Kato Klebanov, and Naomi Sealand. 1993. "Do Neighborhoods Influence Child and Adolescent Development?" *American Journal of Sociology* 99(2):353-95.

Bryk, Anthony S., and Stephen W. Raudenbush. 1987. "Application of Hierarchical Linear-Models to Assessing Change." *Psychological Bulletin* 101(1):147-58.

Buchmann, Claudia, and Ben Dalton. 2002. "Interpersonal Influences and Educational Aspirations in 12 Countries: The Importance of Institutional Context." *Sociology of Education* 75(2):99-122.

Chase-Lansdale, P. Lindsay, Rachel A. Gordon, Jeanne Brooks-Gunn, and Pamela Kato Klebanov. 1997. "Neighborhood and Family Influences on the Intelllectual and Behavioral Competence of Preschool and Early School-Age Children." Pp. 79-118 in *Neighborhood Poverty: Context and Consequences for Children*, edited by J. Brooks-Gunn, Greg J. Duncan, and J. Aber. New York, NY: Russell Sage Foundation.

Choi, Kate H., R. Kelly Raley, Chandra Muller, and Catherine Riegle-Crumb. 2008. "Class Composition: Socioeconomic Characteristics of Coursemates and College Enrollment." *Social Science Quarterly* 89(4):846-66.

Clampet-Lundquist, Susan, Kathryn Edin, Jeffrey R. Kling, and Greg J. Duncan. 2011. "Moving Teenagers Out of High-Risk Neighborhoods: How Girls Fare Better Than Boys." *American Journal of Sociology* 116(4):1154-89.

Cole, S.R., and M.A. Hernán. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American journal of epidemiology* 168(6):656.

Cole, S.R., R.W. Platt, E.F. Schisterman, H. Chu, D. Westreich, D. Richardson, and C. Poole. 2010. "Illustrating Bias due to Conditioning on a Collider." *International journal of epidemiology* 39(2):417.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander J. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. "Equality of Educational Opportunity." Washington: USGPO.

Crane, Jonathan. 1991. "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing." *American Journal of Sociology* 96(5):1226-59.

Crosnoe, Robert. 2009. "Low-Income Students and the Socioeconomic Composition of Public High Schools." *American Sociological Review* 74(5):709-30.

Crowder, Kyle, and Scott J. South. 2011. "Spatial and Temporal Dimensions of Neighborhood Effects on High School Graduation." *Social Science Research* 40(1):87-106.

Dauber, Susan L., Karl L. Alexander, and Doris R. Entwisle. 1996. "Tracking and Transitions through the Middle Grades: Channeling Educational Trajectories." *Sociology of Education* 69(4):290-307.

Davis, James A. 1966. "Campus as a Frog Pond - Application of Theory of Relative Deprivation to Career Decisions of College Men." *American Journal of Sociology* 72(1):17-31.

DeLuca, Stefanie, and Elizabeth Dayton. 2009. "Switching Social Contexts: The Effects of Housing Mobility and School Choice Programs on Youth Outcomes." *Annual Review of Sociology* 35:457-91.

Duncan, Greg J. 1994. "Families and Neighbors as Sources of Disadvantage in the Schooling Decisions of White and Black Adolescents." *American Journal of Education* 103(1):20-53.

Duncan, Greg J., and Jeanne Brooks-Gunn. 1997. *Consequences of Growing up Poor*. New York: Russell Sage Foundation.

Duncan, Greg J., and Stephen W. Raudenbush. 1999. "Assessing the Effects of Context in Studies of Child and Youth Development." *Educational Psychologist* 34(1):29-41.

England, Paula, Paul D. Allison, and Yuxiao Wu. 2007. "Does Bad Pay Cause Occupations to Feminize, Does Feminization Reduce Pay, and How Can We Tell with Longitudinal Data?" *Social Science Research* 36(3):1237-56.

Entwisle, Doris R., Karl L. Alexander, and Linda S. Olson. 1994. "The Gender Gap in Math: Its Possible Origins in Neighborhood Effects." *American Sociological Review* 59(6):822-38.

—. 2005. "First Grade and Educational Attainment by Age 22: A New Story." *American Journal of Sociology* 110(5):1458-502.

Evans, William N., Wallace E. Oates, and Robert M. Schwab. 1992. "Measuring Peer Group Effects - A Study of Teenage Behavior." *Journal of Political Economy* 100(5):966-91.

Felmlee, Diane, and Donna Eder. 1983. "Contextual Effects in the Classroom - The Impact of Ability Groups on Student Attention." *Sociology of Education* 56(2):77-87.

Furman, Wyndol. 1982. "Children's Friendships." Pp. 327-39 in *Review of Human Development*, edited by Tiffany M. Field. New York, NY: Wiley.

Galster, George, Dave E. Marcotte, Marv Mandell, Hal Wolman, and Nancy Augustine. 2007. "The Influence of Neighborhood Poverty during Childhood on Fertility, Education, and Earnings Outcomes." *Housing Studies* 22(5):723-51.

Gamoran, Adam. 1987. "The Stratification of High-School Learning Opportunities." *Sociology of Education* 60(3):135-55.

Garner, Catherine L., and Stephen W. Raudenbush. 1991. "Neighborhood Effects on Educational Attainment: A Multilevel Analysis." *Sociology of Education* 64(4):251-62.

Grant, Gerald. 2009. *Hope and Despair in the American City*. Cambridge, MA: Harvard University Press.
Greenland, S. 2003. "Quantifying biases in causal models: classical confounding vs collider-stratification bias." *Epidemiology* 14(3):300.
Halaby, Charles N. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30:507-44.
Hallinan, Maureen T. 1992. "The Organization of Students for Instruction in Middle School." *Sociology of Education* 65(2):114-27.
Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2009. "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics* 27(3):349-83.
Harding, David J. 2003. "Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping out and Teenage Pregnancy." *American Journal of Sociology* 109(3):676-719.
Harris, Douglas N. 2010. "How Do School Peers Influence Student Educational Outcomes? Theory and Evidence from Economics and Other Social Sciences." *Teachers College Record* 112(4).
Hauser, Robert M. 1970. "Context and Consex - Cautionary Tale." *American Journal of Sociology* 75(4):645-&.
Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312(5782):1900-02.
Hernán, M.A., S. Hernández-Díaz, and J.M. Robins. 2004. "A structural approach to selection bias." *Epidemiology* 15(5):615.
Hernan, Miguel Angel, Babette Brumback, and James M. Robins. 2000. "Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-positive Men." *Epidemiology* 11(5):561-70.
Hong, Guanglei, and Stephen W. Raudenbush. 2008. "Causal Inference for Time-Varying Instructional Treatments." *Journal of Educational and Behavioral Statistics* 33(3):333-62.
Hoxby, Caroline M., and Gretchen Weingarth. 2005. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." *Working Paper*.
Jackson, Margot I., and Robert D. Mare. 2007. "Cross-Sectional and Longitudinal Measurements of Neighborhood Experience and Their Effects on Children." *Social Science Research* 36(2):590-610.
Jacobs, David, and Jason T. Carmichael. 2001. "The Politics of Punishment across Time and Space: A Pooled Time-Series Analysis of Imprisonment Rates." *Social Forces* 80(1):61-89.
Jacobs, David, and Daniel Tope. 2007. "The Politics of Resentment in the Post-Civil Rights Era: Minority Threat, Homicide, and Ideological Voting in Congress." *American Journal of Sociology* 112(5):1458-94.
Jencks, Christopher, and Susan E. Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." Pp. 111-86 in *Inner-City Poverty in the United States*, edited by Laurence E. Lynn and Michael G.H. McGeary. Washington, D.C.: National Academy Press.
Kahlenberg, Richard D. 2001. *All Together Now: Creating Middle-Class Schools Through Public School Choice*. Washington, D.C.: Brookings Institution Press.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1):83-119.
Kocak, Ozgecan, and Glenn R. Carroll. 2008. "Growing Church Organizations in Diverse US Communities, 1890-1926." *American Journal of Sociology* 113(5):1272-315.
Lee, Valerie E., Anthony S. Bryk, and Julia B. Smith. 1993. "The Effects of High School Organization on Teachers and Students." Pp. 171-268 in *Review of Research in Education*, edited by Linda Darling-Hammond. Washington, DC: American Educational Research Association.
Legewie, Joscha, and Thomas A. DiPrete. 2012. "School Context and the Gender Gap in Educational Achievement." *American Sociological Review* 77(3):463-85.
Leventhal, Tama, and Jeanne Brooks-Gunn. 2000. "The Neighborhoods They Live in: The Effects of Neighborhood Residence on Child and Adolescent Outcomes." *Psychological Bulletin* 126(2):309.
Luo, Wen, and Oi-man Kwok. 2009. "The Impacts of Ignoring a Crossed Factor in Analyzing Cross-Classified Data." *Multivariate Behavioral Research* 44(2):182-212.
Manski, Charles F. 1993. "Identification of Endogenous Social Effects - The Reflection Problem." *Review of Economic Studies* 60(3):531-42.
Marsh, Herbert W. 1987. "The Big-Fish Little-Pond Effect on Academic Self-Concept." *Journal of Educational Psychology* 79(3):280-95.
Marsh, Herbert W., and John W. Parker. 1984. "Determinants of Student Self-Concept - Is It Better to Be a Relatively Large Fish in a Small Pond Even If You Dont Learn to Swim as Well." *Journal of Personality and Social Psychology* 47(1):213-31.
Masi, Christopher M., Louise C. Hawkley, Z. Harry Piotrowski, and Kate E. Pickett. 2007. "Neighborhood Economic Disadvantage, Violent Crime, Group Density, and Pregnancy Outcomes in a Diverse, Urban Population." *Social Science & Medicine* 65(12):2440-57.
Morenoff, Jeffrey D. 2003. "Neighborhood mechanisms and the spatial dynamics of birth weight." *American Journal of Sociology* 108(5):976-1017.
Mouw, Ted. 2003. "Social capital and finding a job: Do contacts matter?" *American Sociological Review* 68(6):868-98.
—. 2006. "Estimating the causal effect of social capital: A review of recent research." *Annual Review of Sociology* 32:79-102.
Orfield, Gary, Susan E. Eaton, and Harvard Project on School Desegregation. 1996. *Dismantling Desegregation: The Quiet Reversal of Brown v. Board of Education*. New York: New Press : Distributed by W.W. Norton & Company.
Orr, Larry, Judith D. Feins, Rrobin Jacob, Erik Beercroft, Lisa Sanbonmatsu, Lawrence F. Katz, Jeffrey B. Liebman, and Jeffrey R. Kling. 2003. "Moving to Opportunity: Interim Impacts Evaluation." Washington, D.C.: U..S. Department of Housing and Urban Development.
Owens, Ann. 2010. "Neighborhoods and Schools as Competing and Reinforcing Contexts for Educational Attainment." *Sociology of Education* 83(4):287-311.
Quigley, John M., and Steven Raphael. 2008. "Neighborhoods, Economic Self-Sufficiency, and the MTO." *Brookings-Wharton Papers on Urban Affairs* 3:1-46.
Reardon, Sean F. , and John T. Yun. 2005. "Integrating Neighborhoods, Segregating Schools: The Retreat from School Desegregation in the South, 1990-2000." Pp. 51-69 in *School Resegregation: Must the South Turn Back?*, edited by John Charles Boger and Gary Orfield. Chapel Hill: University of North Carolina Press.

Robins, James M. 1999. "Association, Causation, and Marginal Structural Models." *Synthese* 121:151-79.
Robins, James M., Miguel Angel Hernan, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11(5):550-60.
Rumberger, Russell W., and Gregory J. Palardy. 2005. "Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High School." *The Teachers College Record* 107(9):1999-2045.
Rumberger, Russell W., and J. Douglas Willms. 1992. "The Impact of Racial and Ethnic Segregation on the Achievement Gap in California High Schools." *Educational Evaluation and Policy Analysis* 14(4):377-96.
Sampson, Robert J. 2008. "Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure." *American Journal of Sociology* 118(1):189-231.
Sampson, Robert J., Patrick Sharkey, and Stephen W. Raudenbush. 2008. "Durable Effects of Concentrated Disadvantage on Verbal Ability among African-American Children." *Proceedings of the National Academy of Sciences of the United States of America* 105(3):845-52.
Sanbonmatsu, Lisa, Jeffrey R. Kling, Greg J. Duncan, and Jeanne Brooks-Gunn. 2006. "Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment." *Journal of Human Resources* 41(4):649-91.
Saporito, Salvatore, and Deenesh Sohoni. 2007. "Mapping Educational Inequality: Concentrations of Poverty among Poor and Minority Students in Public Schools." *Social Forces* 85(3):1227.
Schneiberg, Marc, Marissa King, and Thomas Smith. 2008. "Social movements and organizational form: Cooperative alternatives to corporations in the American insurance, dairy, and grain industries." *American Sociological Review* 73(4):635-67.
Sedlak, Michael W., Christopher W. Wheeler, Diana C. Pullin, and Philip A. Cusick. 1986. *Selling Students Short: Classroom Bargains and Academic Reform in the American Classroom*. New York, NY: Teachers College Press.
Sharkey, Patrick, and Felix Elwert. 2011. "The Legacy of Disadvantage: Multigenerational Neighborhood Effects on Cognitive Ability." *American Journal of Sociology* 116(6):1934-81.
Singer, Judith D., and John B. Willett. 2003. *Applied Longitudinal Data Analysis*. New York, NY: Oxford University Press.
Solon, Gary, Marianne E. Page, and Greg J. Duncan. 2000. "Correlations between neighboring children in their subsequent educational attainment." *Review of Economics and Statistics* 82(3):383-92.
South, Scott J., Eric P. Baumer, and Amy Lutz. 2003. "Interpreting community effects on youth educational attainment." *Youth & Society* 35(1):3-36.
Veronneau, Marie-Helene, and Thomas J. Dishion. 2011. "Middle School Friendships and Academic Achievement in Early Adolescence: A Longitudinal Analysis." *Journal of Early Adolescence* 31(1):99-124.
Willms, J. Douglas. 1986. "Social Class Segregation and Its Relationship to Pupils' Examination Results in Scotland." *American Sociological Review* 51(2):224-41.
—. 2010. "School Composition and Contextual Effects on Student Outcomes." *Teachers College Record* 112(4).

Wilson, Alan B. 1959. "Residential Segregation of Social-Classes and Aspirations of High-School Boys." *American Sociological Review* 24(6):836-45.
Wodtke, Geoffrey T., David J. Harding, and Felix Elwert. 2011. "Neighborhood Effects in Temporal Perspective: The Impact of Long-Term Exposure to Concentrated Disadvantage on High School Graduation." *American Sociological Review* 76(5):713-36.
Wooldridge, Jeffrey M. 2003. *Introductory Econometrics: A Modern Approach*. Cincinnati, OH: South-Western College Publishing.

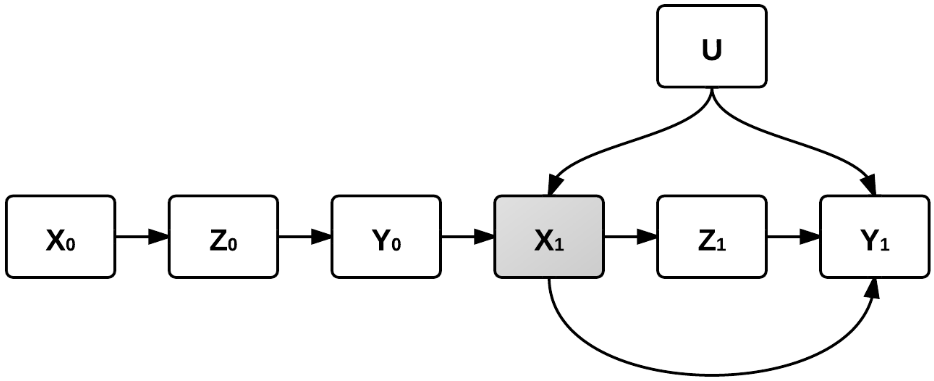

Figure 1. Causal diagram showing a time-varying confounder (X_1) on the causal pathway between treatment occasions and the outcome.

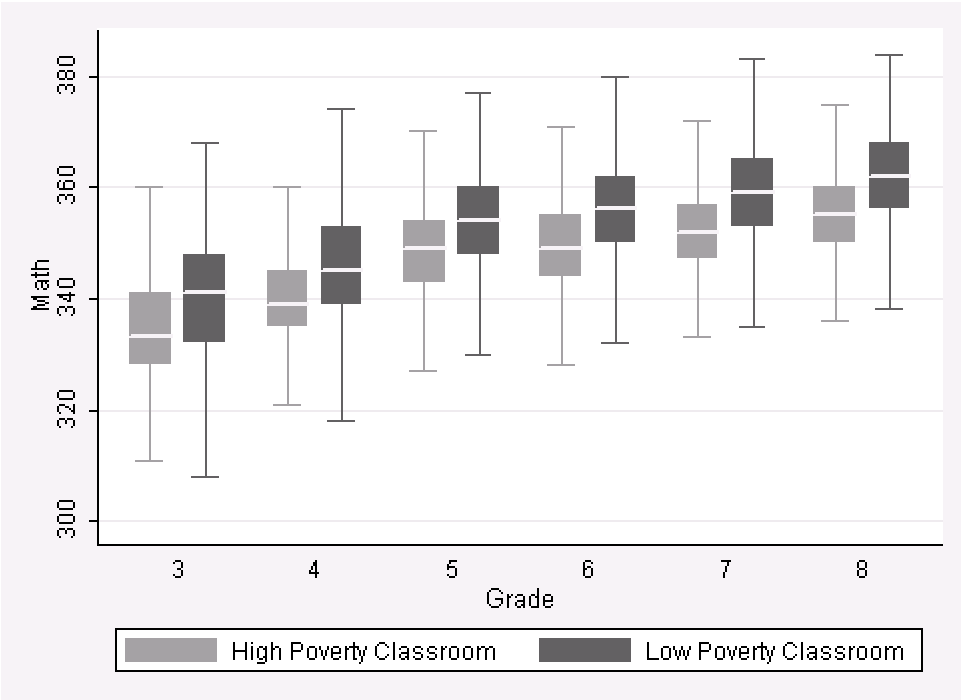

Figure 2. Math Score by Classroom Poverty and Grade
Note: Each box contains the 25th to 75th percentile with the white line in each box at the median. Whiskers indicate the 5th and 95th percentiles

Table 1. Descriptive Statistics

	Obs	Mean	SD	Min	Max
Dependent Variables					
Math test score	537,653	350.78	11.84	303	388
Reading test score	537,653	348.04	12.89	303	384
Student Background					
Parent has high school degree or less	537,653	0.53	0.42	0	1
Parent has some postsecondary education	537,653	0.21	0.41	0	1
Parent has bachelor's degree or higher	537,653	0.27	0.44	0	1
Black student	537,653	0.30	0.46	0	1
Hispanic student	537,653	0.05	0.22	0	1
Other racial/ethnic background	537,653	0.05	0.22	0	1
Male student	537,653	0.50	0.50	0	1
Student was designated gifted	537,653	0.14	0.35	0	1
Student received special education services	537,653	0.11	0.32	0	1
Student showed Limited English Proficiency	537,653	0.02	0.15	0	1
Student was ever retained	537,653	0.11	0.32	0	1
Student received free or reduced price lunch	537,653	0.45	0.50	0	1
Student made a structural school move	537,653	0.16	0.37	0	1
Student made a non-structural school move	537,653	0.09	0.28	0	1
Classroom Poverty Measures					
High poverty classroom (top quintile)	537,653	0.24	0.43	0	1
Cumulative exposure to high poverty classrooms	537,653	0.23	0.35	0	1
Continuous classroom poverty (pct free/reduced lunch)	537,653	0.45	0.26	0	1
Time Variables					
Grade level	537,653	5.34	1.69	3	8
Grade level ^2	537,653	31.41	18.50	9	64

Note: Observations reported are student-year observations.

Table 2. Classroom Poverty and Math Achievement: Cross-Sectional Multilevel Models, 2001-2006

	3rd Grade	4th Grade	5th Grade	6th Grade	7th Grade	8th Grade
High Poverty	-0.877***	-0.848***	-1.155***	-2.088***	-2.210***	-2.484***
Classroom	(0.0976)	(0.0819)	(0.0799)	(0.0888)	(0.0880)	(0.107)
Effect Size	-0.082	-0.091	-0.133	-0.233	-0.249	-0.280
SD (Math)	10.74	9.34	8.66	8.95	8.87	8.86
Observations	97,995	96,144	90,904	89,071	84,585	74,525

Note: Each model controls for race, gender, parental education, individual poverty status, gifted, special education, limited English proficiency, structural school mobility, and non-structural school mobility. Each model also includes a random intercept for classroom. High poverty classroom is the top quartile of a standardized measure of percent classroom peers' poverty status. Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3. Classroom Poverty and Math Achievement: Random Coefficient Growth Models, 2001-2006

	(1) Classroom Pov	(2) w/ Growth	(3) w/ Student Chars	(4) w/ Student Chars Interactions
High Poverty Classroom	-2.249***	-0.683***	-0.413***	-0.407***
	(0.0413)	(0.0191)	(0.0191)	(0.0191)
Grade		4.166***	4.148***	4.142***
		(0.00451)	(0.00453)	(0.00451)
Grade^2		-0.651***	-0.651***	-0.653***
		(0.00228)	(0.00229)	(0.00229)
Parent Has Some Postsec Educ			0.514***	0.419***
			(0.0174)	(0.0230)
Parent Has Bach Degree or More			1.057***	0.764***
			(0.0217)	(0.0265)
Black			-6.324***	-6.358***
			(0.0523)	(0.0538)
Hispanic			-3.507***	-3.661***
			(0.108)	(0.111)
Other Race/Ethnicity			-1.313***	-1.454***
			(0.120)	(0.122)
Male			-0.419***	-0.302***
			(0.0480)	(0.0490)
Student Poverty			-0.706***	-0.601***
			(0.0236)	(0.0275)
Constant	350.2***	352.5***	352.5***	352.5***
	(0.0283)	(0.0269)	(0.0243)	(0.0245)
σ_u	7.809***	7.968***	7.118***	7.175***
	(0.370)	(0.322)	(0.268)	(0.274)
σ_e	8.461***	3.722***	3.736***	3.733***
	(0.160)	(0.0392)	(0.0397)	(0.0396)
SD(High Pov Classroom)	5.596***	1.027***	0.930***	0.907***
	(0.666)	(0.116)	(0.114)	(0.112)
SD(Grade)		0.978***	0.976***	0.955***
		(0.00906)	(0.00906)	(0.00888)
SD(Grade^2)		0.279***	0.280***	0.257***
		(0.00251)	(0.00252)	(0.00246)
Observations	537,653	537,653	537,653	537,653

Note: Random coefficient models (see equation 2) with an unstructured covariance matrix. Covariances of random effects not shown. Model 4 includes interactions between listed student background controls and grade and grade^2 (coefficients not shown). Robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4. Classroom Poverty and Reading Achievement: Random Coefficient Growth Models, 2001-2006

	(1) Classroom Pov	(2) w/ Growth	(3) w/ Student Chars	(4) w/ Student Chars Interactions
High Poverty Classroom	-2.275***	-0.727***	-0.418***	-0.399***
	(0.0469)	(0.0220)	(0.0221)	(0.0220)
Grade		4.656***	4.631***	4.630***
		(0.00517)	(0.00520)	(0.00513)
Grade^2		-0.570***	-0.573***	-0.570***
		(0.00260)	(0.00261)	(0.00263)
Parent Has Some Postsec Educ			0.584***	0.525***
			(0.0203)	(0.0274)
Parent Has Bach Degree or More			1.278***	1.145***
			(0.0248)	(0.0311)
Black			-6.097***	-6.662***
			(0.0550)	(0.0591)
Hispanic			-3.905***	-5.074***
			(0.116)	(0.123)
Other Race/Ethnicity			-1.769***	-2.392***
			(0.115)	(0.124)
Male			-1.617***	-1.702***
			(0.0481)	(0.0519)
Student Poverty			-0.951***	-0.869***
			(0.0271)	(0.0325)
Constant	347.5***	349.6***	349.6***	349.6***
	(0.0303)	(0.0290)	(0.0259)	(0.0260)
σ_u	8.237***	8.479***	7.460***	7.483***
	(0.425)	(0.373)	(0.305)	(0.310)
σ_e	9.481***	4.448***	4.466***	4.465***
	(0.201)	(0.0556)	(0.0563)	(0.0562)
SD(High Pov Classroom)	6.407***	1.206***	1.086***	1.095***
	(0.829)	(0.153)	(0.150)	(0.150)
SD(Grade)		1.103***	1.097***	1.052***
		(0.0121)	(0.0121)	(0.0117)
SD(Grade^2)		0.247***	0.247***	0.237***
		(0.00331)	(0.00333)	(0.00331)
Observations	537,653	537,653	537,653	537,653

Note: Random coefficient models (see equation 2) with an unstructured covariance matrix. Covariances of random effects not shown. Model 4 includes interactions between listed student background controls and grade and grade^2 (coefficients not shown). Robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5. High Poverty Classroom and Achievement: Comparison of Alternative Specifications, 2001-2006

	(1) Random Coefficients	(2) Student FE	(3) MSM w/ IPTW (treatment and censoring)
A. Math			
High Poverty Classroom	-0.413^{***}	-0.0629^{***}	0.00784
	(0.0191)	(0.0204)	(0.0180)
Grade	4.148^{***}	4.154^{***}	4.199^{***}
	(0.00453)	(0.00455)	(0.00466)
Grade^2	-0.651^{***}	-0.655^{***}	-0.665^{***}
	(0.00229)	(0.00230)	(0.00235)
Constant	352.5^{***}	352.6^{***}	352.3^{***}
	(0.0243)	(0.00654)	(0.0177)
Observations	537,653	537,653	537,653
B. Reading			
High Poverty Classroom	-0.418^{***}	0.106^{***}	0.0545^{*}
	(0.0221)	(0.0237)	(0.0212)
Grade	4.631^{***}	4.623^{***}	4.685^{***}
	(0.00520)	(0.00524)	(0.00540)
Grade^2	-0.573^{***}	-0.592^{***}	-0.589^{***}
	(0.00261)	(0.00265)	(0.00266)
Constant	349.6^{***}	349.7^{***}	349.4^{***}
	(0.0259)	(0.00753)	(0.0205)
Observations	537,653	537,653	537,653

Note: Model 1 panel A is model 3 from table 3 and model 1 panel B is model 3 from table 4 (with the same covariates, though only a selection are shown here) reprinted here for comparison purposes. All models control for parent's education, race/ethnicity, gender, and poverty status; race/ethnicity and gender are subsumed into the student-specific intercept in the student fixed effects model; MSM weights also adjust for school mobility, gifted, special education, LEP, and grade retention. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6. Cumulative Exposure to Classroom Poverty and Achievement: Comparison of Alternative Specifications, 2001-2006

	(1) Random Coefficients	(2) Student FE	(3) MSM w/ IPTW (treatment and censoring)
A. Math			
Cumulative Pov Exposure	-2.037***	-0.192***	0.0267
	(0.0462)	(0.0567)	(0.0487)
Grade	4.144***	4.154***	4.199***
	(0.00455)	(0.00455)	(0.00465)
$Grade^2$	-0.649***	-0.655***	-0.665***
	(0.00229)	(0.00230)	(0.00235)
Constant	352.5***	352.6***	352.3***
	(0.0241)	(0.00654)	(0.0178)
Observations	537,653	537,653	537,653
B. Reading			
Cumulative Pov Exposure	-2.085***	0.292***	0.263***
	(0.0529)	(0.0668)	(0.0569)
Grade	4.627***	4.623***	4.685***
	(0.00522)	(0.00524)	(0.00540)
$Grade^2$	-0.572***	-0.593***	-0.589***
	(0.00261)	(0.00265)	(0.00266)
Constant	349.6***	349.7***	349.4***
	(0.0257)	(0.00753)	(0.0205)
Observations	537,653	537,653	537,653

Note: All models control for parent's education, race/ethnicity, gender, and poverty status; race/ethnicity and gender are subsumed into the student-specific intercept in the student fixed effects model; MSM weights also adjust for school mobility, gifted, special education, LEP, and grade retention. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 7. Continuous Classroom Poverty and Achievement: Comparison of Alternative Specifications, 2001-2006

	(1) Random Coefficients	(2) Student FE
A. Math		
Continuous Classroom Pov	-0.297^{***}	0.0328^{**}
	(0.0104)	(0.0112)
Grade	4.123^{***}	4.156^{***}
	(0.00459)	(0.00460)
$Grade^2$	-0.648^{***}	-0.655^{***}
	(0.00229)	(0.00230)
Constant	352.5^{***}	352.6^{***}
	(0.0242)	(0.00654)
Observations	537,653	537,653
B. Reading		
Continuous Classroom Pov	-0.260^{***}	0.224^{***}
	(0.0118)	(0.0131)
Grade	4.610^{***}	4.637^{***}
	(0.00529)	(0.00529)
$Grade^2$	-0.571^{***}	-0.592^{***}
	(0.00261)	(0.00265)
Constant	349.6^{***}	349.7^{***}
	(0.0258)	(0.00753)
Observations	537,653	537,653

Note: All models control for parent's education, race/ethnicity, gender, and poverty status; race/ethnicity and gender are subsumed into the student-specific intercept in the student fixed effects model; MSM weights also adjust for school mobility, gifted, special education, LEP, and grade retention. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 8. Effect Size Summary Table

	(1) Random Coefficients			(2) Student FE			(3) MSM w/ IPTW		
	Lower bound	Estimate	Upper bound	Lower bound	Estimate	Upper bound	Lower bound	Estimate	Upper bound
A. High Poverty Classroom									
Math	-0.038	-0.035	-0.032	-0.009	-0.005	-0.002	-0.004	0.000	0.003
Reading	-0.036	-0.032	-0.029	0.005	0.008	0.012	0.001	0.006	0.009
B. Cumulative Pov Exposure									
Math	-0.180	-0.172	-0.164	-0.026	-0.016	-0.007	-0.006	0.000	0.008
Reading	-0.170	-0.162	-0.154	0.012	0.023	0.033	0.012	0.022	0.031
C. Continuous Classroom Pov									
One standard deviation increase									
Math	-0.027	-0.025	-0.023	0.001	0.003	0.005	--	--	--
Reading	-0.022	-0.020	-0.018	0.015	0.017	0.019	--	--	--
Two standard deviation increase									
Math	-0.054	-0.050	-0.047	0.002	0.006	0.009	--	--	--
Reading	-0.044	-0.040	-0.037	0.031	0.035	0.039	--	--	--

Note: These estimates use coefficients and 95% confidence intervals from Tables 5, 6, and 7, divided by the standard deviation of math (11.84) and reading (12.89) to provide standardized effect sizes. High poverty classroom effects represent the average difference between 0 (not in a high poverty classroom) and 1 (in a high poverty classroom) across all years. Cumulative poverty exposure effects represent the average difference between 0 (never in a high poverty classroom) and 1 (always in a high poverty classroom) across all years. MSM model for continuous classroom poverty not estimated.

Appendix Table A1. Pre-Imputation Descriptive Statistics

	Pre-Imputation			Post-Imputation		
	Obs	**Mean**	**SD**	**Obs**	**Mean**	**SD**
Dependent Variables						
Math test score	550,147	350.72	11.86	537,653	350.78	11.84
Reading test score	548,301	348.11	12.91	537,653	348.04	12.89
Student Background						
Parent has high school degree or less	551,906	0.53	0.50	537,653	0.53	0.42
Parent has some postsecondary education	551,906	0.21	0.41	537,653	0.21	0.41
Parent has bachelor's degree or higher	551,906	0.26	0.44	537,653	0.27	0.44
Black student	558,353	0.31	0.46	537,653	0.30	0.46
Hispanic student	558,353	0.05	0.22	537,653	0.05	0.22
Other racial/ethnic background	558,353	0.05	0.23	537,653	0.05	0.22
Male student	558,353	0.51	0.50	537,653	0.50	0.50
Student was designated gifted	558,353	0.14	0.35	537,653	0.14	0.35
Student received special education services	557,862	0.12	0.33	537,653	0.11	0.32
Student showed Limited English Proficiency	557,931	0.02	0.15	537,653	0.02	0.15
Student was ever retained	552,791	0.11	0.32	537,653	0.11	0.32
Student received free or reduced price lunch	544,289	0.46	0.50	537,653	0.45	0.50
Student made a structural school move	558,353	0.16	0.36	537,653	0.16	0.37
Student made a non-structural school move	558,353	0.09	0.29	537,653	0.09	0.28
Peer Poverty						
High poverty classroom	558,353	0.24	0.43	537,653	0.24	0.43
Cumulative exposure to high poverty classrooms	558,353	0.24	0.35	537,653	0.23	0.35
Continuous classroom poverty	558,353	0.45	0.27	537,653	0.45	0.26
Time Variables						
Grade level	558,353	5.34	1.70	537,653	5.34	1.69
Grade level ^2	558,353	31.43	18.58	537,653	31.41	18.50

Note: Observations reported are student-year observations. 4,108 student year observations were dropped pre-imputation due to a student having < 50% valid math scores in their panel (1,266 for reading).