Are Neighborhood Effects Explained by Differences in School Quality?

Geoffrey T. Wodtke
University of Chicago

Ugur Yildirim
University of California-Berkeley

David J. Harding
University of California-Berkeley

Felix Elwert
University of Wisconsin-Madison

Corresponding Author: Geoffrey T. Wodtke, Department of Sociology, University of Chicago, 1126 E. 59th St. Chicago, IL 60637. Email: wodtke@uchicago.edu.

Acknowledgements: The authors thank Ken Wodtke, Steven Durlauf, James Heckman, Stephen Raudenbush, Guanglei Hong, Scott Davies, Katyn Chmielewski, Zahide Alaca, Xiang Zhou, Eric Grodsky, Kristin Perkins, and Jared Schachner for helpful comments and discussions. This study was supported by grants from the Social Sciences and Humanities Research Council of Canada (No. 435-2018-0736), the U.S. National Science Foundation (No. 2015613), and the Institute for Research on Labor and Employment at the University of California-Berkeley.

Are Neighborhood Effects Explained by Differences in School Quality?

Abstract

It is widely hypothesized that neighborhood effects on academic achievement are explained by differences in the quality of schools attended by resident children. We evaluate this hypothesis by examining whether elementary school quality mediates or interacts with the effects of neighborhood poverty using data from the Early Childhood Longitudinal Study. With a diverse set of measures that variously capture a school's effectiveness, resources, and climate, we implement a novel decomposition that separates the overall effect of neighborhood poverty into components due to mediation versus interaction via these different factors. Results indicate that living in a disadvantaged neighborhood reduces academic achievement. But, contrary to expectations, we find no evidence that neighborhood effects are mediated by or interact with any of our measures for school quality. Differences in the quality of elementary schools do not appear to mediate the effects of neighborhood context because they are not, in fact, strongly linked with the socioeconomic composition of neighborhoods. Elementary school quality also does not appear to interact with neighborhood context because it has similar effects on achievement whether children reside in advantaged or disadvantaged neighborhoods. We discuss the implications of these findings for theory, research, and policy addressing the link between concentrated poverty and educational inequality.

Keywords: neighborhoods, schools, achievement, poverty, inequality, mediation, interaction

1. Introduction

Are neighborhood effects on academic achievement explained by differences in the quality of schools that resident children attend? Although a large volume of evidence indicates that neighborhood poverty affects academic achievement (Chetty et al. 2016; Harding 2003; Rosenbaum 1995; Wodtke et al. 2011, 2016), relatively little is known about the causal processes through which these effects may be transmitted. Indeed, a frequent criticism of research on concentrated poverty is that "the social mechanisms…accounting for neighborhood effects have remained largely a black box" (Sampson 2012:46).

Analyses of the social mechanisms linking concentrated poverty to academic achievement are important for several reasons. First, they are important for testing, refining, and building theoretical models for the effects of residential segregation on child development, the reproduction of poverty, and the intergenerational transmission of social status (Sharkey and Faber 2014; Harding et al. 2011). Second, they also help to identify promising points of intervention for policies aimed at attenuating the harms of concentrated poverty and promoting upward social mobility (Sampson 2012). Without knowledge of the mechanisms that explain neighborhood effects, it is difficult to diagnose the failures of existing urban policy, and by extension, to design more effective interventions in the future. Finally, because causal inference is a holistic enterprise in research on neighborhood effects, where randomized experiments are difficult to design and implement (Clampet-Lundquist and Massey 2008; Sampson 2008), inferences about the consequences of neighborhood poverty are buttressed when it is possible to explain them by tracing out the mechanisms connecting residential conditions to individual outcomes.

Whether the effects of neighborhood poverty can be explained by differences in school quality depends on two causal processes: mediation and interaction. Mediation refers to the operation of a causal chain whereby differences in neighborhood context engender differences in access to higher versus lower quality schools, which in turn generate differences in academic achievement. Interaction, by contrast, refers to a causal process whereby the effects of school quality on achievement are dampened or amplified by residence in an advantaged versus disadvantaged neighborhood. Mediation may occur in the absence of interaction, interaction may occur in the absence of mediation, or both may occur together (VanderWeele 2015). In other words, neighborhood poverty may influence academic achievement not only by changing the school environment to which children are exposed but also by altering the effects of this environment on student learning.

It is widely hypothesized that neighborhood effects on academic achievement are *mediated* by differences in school quality. For example, according to institutional resource theory, children in disadvantaged neighborhoods are more likely to attend lower quality schools because schools in poor communities may have fewer experienced teachers, more disorderly classrooms, and a slower pace of instruction (Arum 2000; Jencks and Mayer 1990; Johnson 2012; Leventhal and Brooks-Gunn 2000; Sanbonmatsu et al. 2006; Wilson 1987). These educational deficiencies are, in turn, thought to inhibit student learning.

It is also widely hypothesized that differences in school quality *interact* with neighborhood effects on achievement. For example, compound disadvantage theory suggests that the effects of attending a higher versus lower quality school may be more pronounced for children in disadvantaged neighborhoods because residents of these communities rely more heavily on local institutions than children from advantaged areas (Jencks and Mayer 1990;

Wodtke et al. 2016). By contrast, relative deprivation theory suggests that the effects of school quality are less pronounced when children live in disadvantaged neighborhoods because children from poor communities may struggle to capitalize on the instructional advantages available in higher quality schools (Crosnoe 2009; Jencks and Mayer 1990).

Several prior studies have investigated the joint effects of neighborhood and school contexts on educational outcomes. Some report mainly neighborhood effects (Ainsworth 2002; Card and Rothstein 2007; Wodtke and Parbst 2017); some report mainly school effects (Goldsmith 2009; Carlson and Cohen 2014; Cook et al. 2002); and others report both (Owens 2010; Rendón 2014). All of these prior studies, however, suffer from two important limitations. First, none properly evaluate the explanatory role of schools by decomposing the overall effect of neighborhood context into components due to mediation versus interaction. Second, they rely primarily on measures of school composition, such as the proportion of students who are eligible for a lunch subsidy or who identify as black, that are distinct from measures of school quality and are, at best, noisy proxies for the concept.

In this study, we investigate whether differences in the quality of elementary schools, and specifically, the 1st grade classes within them, explain the effect of neighborhood poverty on academic achievement using a large and diverse set of measures for school quality together with novel decomposition methods. The concept of school quality is theoretically amorphous and sometimes controversial. In prior research, it is variously measured in terms of school inputs, such as financial resources and teacher qualifications, or in terms of characteristics internal to schools, such as academic climate and classroom disorder (Ladd and Loeb 2013). More recently, school quality is increasingly measured in terms of school outputs, such as student proficiencies or value added (Downey et al. 2019; Raudenbush and Eschmann 2015). We do not attempt to adjudicate or unite these different approaches. Instead, we investigate whether neighborhood effects can be explained by *any* of these multiple dimensions of school quality.

With data from the Early Childhood Longitudinal Study - Kindergarten Class of 1998 (ECLS-K), we operationalize school quality in three ways. First, we use an output-based, or value-added, measure that is equal to the difference between a school's average learning rate among its 1st grade students during the school year and the average learning rate among those same students during the previous summer (Downey et al. 2008, 2019; Raudenbush and Eschmann 2015). By this measure, which we refer to as "school effectiveness," a high-quality elementary school is one that raises its students' abilities above what they might otherwise be without the benefit of attending school. Second, we operationalize school quality using an input-based measure that combines information on per-pupil expenditures, classroom size, and teacher qualifications (Hanushek 2006; Jackson et al. 2016). By this measure, which we refer to as "school resources," a high-quality elementary school is one that is well-funded, has small class sizes, and employs the most qualified and experienced teachers. Finally, we operationalize school quality using an internal measure based on levels of student absenteeism and disruptive classroom behavior (Figlio 2007; Gottfried 2019), which we refer to as "school disorder." With these measures, we then use novel methods of causal inference to decompose the overall effects of neighborhood context on achievement at the end of 3rd, 5th, and 8th grade into components due to mediation versus interaction (Wodtke and Zhou 2020; Zhou and Wodtke 2019).

Our results indicate that exposure to a disadvantaged neighborhood during kindergarten has negative effects on both reading and mathematics achievement that persist through the end of 8th grade. Contrary to expectations, however, we find little evidence that these effects are mediated by or interact with elementary school quality, whether measured in terms of effectiveness, resources, or disorder. Differences in the quality of elementary schools do not appear to mediate the effects of neighborhood context because they are not, in fact, strongly linked with the socioeconomic composition of neighborhoods. Elementary school quality also does not appear to interact with neighborhood context because attending a higher versus lower quality school has similar effects on achievement whether children reside in advantaged or disadvantaged neighborhoods. Furthermore, a formal sensitivity analysis indicates that these results are stable under several different patterns of unobserved confounding, and a set of additional robustness checks shows that our conclusions are also unaffected by several forms of measurement error, by the use of many alternative measures for school quality, and by the use of multiple different model specifications.

This study makes three contributions to research on concentrated poverty and neighborhood effects. First, empirically, it provides evidence that the effects of neighborhood poverty on academic achievement are most likely not explained by differences in the quality of elementary schools attended by resident children, despite widely held assumptions to the contrary. Second, theoretically, these results help to adjudicate between institutional resource theory, compound disadvantage theory, and relative deprivation theory, on the one hand, and an alternative perspective that views elementary schools as neutral or perhaps even equalizing institutions with respect to the educational inequalities engendered by socioeconomic segregation. Third, methodologically, this study introduces novel methods for decomposing causal effects into components due to mediation versus interaction and for estimating these components in the presence of complex selection processes.

2. Neighborhood Effect Mediation via School Quality

The mechanisms through which poor neighborhoods are hypothesized to affect academic achievement include social and cultural isolation (Wilson 1987), peer influence and the socialization of children by adolescents and young adults (Anderson 1999; Harding 2009; Jencks and Mayer 1990), a breakdown of collective trust among residents and proximity to violent crime (Sampson 2001; Sharkey 2010), exposure to environmental health hazards (Crowder and Downey 2010; Rosenfeld et al. 2010), and institutional resource deprivation (Galster 2012; Jencks and Mayer 1990; Wilson 1987). Elementary schools are one particularly important type of institutional resource, and differences in their quality are widely thought to explain neighborhood effects on academic achievement (e.g., Arum 2000; DeLuca and Rosenblatt 2010; Ferryman et al. 2008; Galster 2012; Johnson 2012).

Consider, for example, the Moving to Opportunity (MTO) field experiment, which found that children in an experimental group who received housing vouchers to move into low-poverty neighborhoods performed no better academically than children in a control group who did not receive housing assistance (Orr et al. 2003, Sanbonmatsu et al. 2011). Although the MTO experiment was limited in a variety of ways (Clampet-Lundquist and Massey 2008; Sampson 2008), many observers have attempted to explain its findings by pointing out that children in the experimental group did not end up attending schools with higher average test scores compared with children in the control group (Dobbie and Fryer 2009; Ferryman et al. 2008, Sanbonmatsu et al. 2006). The small differences in school-level test scores observed across MTO treatment groups prompted Dobbie and Fryer (2011:179) to conjecture that “a better community, as measured by the poverty rate, does not significantly raise test scores if school quality remains essentially unchanged.”

Neighborhood context directly affects the socioeconomic composition of the schools to which children have access because, in most districts, school assignment rules are based on a student's residential location. As a result, children in disadvantaged neighborhoods typically attend schools with a greater number of low-income students than children in advantaged neighborhoods. In total, about 70 percent of the variance in the socioeconomic composition of public schools can be explained by the composition of the catchment areas they serve, despite the proliferation of magnet schools, charter schools, and intra-district open enrollment policies (Saporito and Sohoni 2007).

Elementary schools with a large proportion of low-income students are thought to provide a lower quality of instruction because they may suffer from multiple educational deficiencies. First, schools with a large proportion of low-income students tend to enroll children with lower ability levels, more unstable home environments, and more behavioral problems. Consequently, these schools may have a slower pace of instruction, greater absenteeism, more disorderly classrooms, and a mix of student peers who may struggle to assist one another with learning (Kahlenberg 2001; Willms 2010). Second, elementary schools with a large proportion of low-income students suffer from higher rates of teacher attrition, and they may have difficulty recruiting and retaining the most qualified teachers (Borman and Dowling 2008; Boyd et al. 2005). Finally, schools with many poor students enroll fewer high-achieving children, who may help to engender an academic climate that prioritizes creativity and scholastic excellence rather than obedience and discipline (Esposito 1999; Kahlenberg 2001).

Neighborhood context may also directly affect school quality, apart from its link with the socioeconomic composition of students. For example, elementary schools serving poor communities may have fewer resources because school funding is determined in part by local property tax revenues and because low-income residents are ill-equipped to raise private funds or to provide in-kind benefits for their children's school (Kahlenberg 2001; Steinberg 1997). At the same time, however, both state and federal governments provide compensatory disbursements that tend to offset financial disparities that emerge between schools at the local level (Heuer and Stullich 2011). Elementary schools located in disadvantaged neighborhoods may also experience additional difficulties recruiting and retaining high-quality teachers if, for example, criminal activity in the surrounding area prompts concerns about safety at work or in transit (Boyd et al. 2011). Similarly, violent crime in disadvantaged neighborhoods may negatively influence a school's academic climate if it erodes interpersonal trust and promotes a more authoritarian disciplinary environment (Arum 2005; Devine 1996; Nolan 2011).

In sum, neighborhood context is widely thought to affect elementary school quality both directly and indirectly through its link with the socioeconomic composition of students. And school quality is, in turn, expected to have a lasting influence on academic achievement.

3. Effect Interaction between Neighborhood and School Contexts

Neighborhood context is also widely thought to interact with school quality. Different theoretical perspectives, however, yield divergent hypotheses about whether living in an advantaged versus disadvantaged neighborhood intensifies or attenuates the effects of attending a higher versus lower quality elementary school.

Compound disadvantage theory contends that the experience of deprivation in one social context exacerbates the harmful consequences of deprivation in other contexts (Jencks and Mayer 1990; Wodtke et al. 2016). This suggests that living in a disadvantaged neighborhood intensifies the harmful effects of attending a lower quality elementary school, or equivalently, that it amplifies the benefits of attending a higher quality school. These effects may be more pronounced when children live in a poor neighborhood because the experience of material deprivation across multiple social contexts could engender a negative, fatalistic outlook about one's life chances and the value of education (Wilson 1987). Similarly, when attending a lower quality school, children from poor neighborhoods may become less resilient to the cognitive effects of violent crime or environmental health hazards if, for example, the school does not provide adequate coping, counseling, or health services. Children in poor neighborhoods may also rely more heavily on their local public schools to acquire important academic skills and develop their vocabulary, whereas children in advantaged neighborhoods may have ample opportunities to learn these skills elsewhere.

Relative deprivation theory, by contrast, suggests that living in a disadvantaged neighborhood may actually moderate the harmful effects of attending a lower quality elementary school, or equivalently, that it may dampen the positive effects of attending a higher quality school (Crosnoe 2009; Davis 1966; Jencks and Mayer 1990; Owens 2010). This is because children living in disadvantaged neighborhoods are thought to be poorly equipped to benefit from the resources and instruction provided at high-quality schools. For example, compared to students from advantaged neighborhoods, children from disadvantaged neighborhoods may not come as well prepared for class and may begin elementary school with fewer academic or social skills. Consequently, in higher rather than lower quality schools, they may struggle if the pace of instruction is faster and the curriculum more demanding (Crosnoe 2009; Owens 2010), or they may have difficulty making friends and becoming socially integrated. Children from disadvantaged neighborhoods may also develop negative self-perceptions when they attend higher rather than lower quality schools, where they are more likely to evaluate themselves, and to be evaluated by school staff, against higher academic standards that may be more difficult for them to achieve.

In sum, neighborhood context is widely thought to interact with the effects of elementary school quality on achievement, but whether living in a more versus less disadvantaged neighborhood will dampen or amplify these effects is contested by alternative theoretical perspectives.

4. Schools as Equalizing Institutions?

Although it is commonly hypothesized that neighborhood effects are explained by differences in school quality, few prior studies investigate this causal process. Moreover, results from what limited research exists are generally disconfirming. For example, Wodtke and Parbst (2017) found that the indirect effects of neighborhood context on reading and math achievement, as mediated by the proportion of a school's students who are eligible for a free lunch, were substantively small and statistically insignificant, whereas the direct effects operating independently of school composition were large and significant at stringent thresholds. Similarly, Cook et al. (2002) found that the effects of neighborhood and school characteristics on several different measures of achievement were additive rather than multiplicative, providing little evidence of interaction.

In contrast to the assumptions that pervade theory and research on the effects of concentrated poverty, an emerging body of work in the sociology of education suggests that elementary schools may actually play a neutral or perhaps even an equalizing role in the etiology of academic disparities (Downey 2020; Downey and Condron 2016; Raudenbush and Eschmann 2015). Recent studies of socioeconomic gaps in achievement, for example, raise questions about the degree to which differences in elementary school quality might explain neighborhood effects, as the achievement gap between high- versus low-income students is largest at the start of kindergarten and then shrinks throughout the primary grades (Labaree 2010; Rothstein 2004; von Hippel et al. 2018). This trajectory is difficult to explain without admitting the possibility that low-income students may be served by elementary schools that are actually quite effective. Consistent with this pattern, several prior studies also indicate that disadvantaged elementary schools perceived to be "failing" are not, in fact, typically among the least impactful schools when evaluated in terms of their contributions to student learning (Downey et al. 2008, 2019; von Hippel 2009).

In addition, although there are well-documented differences in the resources available to schools serving more versus less disadvantaged neighborhoods, these disparities are often rather small. For example, Owens and Candipan (2019) report that, in schools serving the wealthiest quintile of American neighborhoods, about 1 percent of teachers are uncertified and about 9 percent have less than 3 years of experience; in the poorest quintile, about 2 percent are uncertified and about 14 percent have under 3 years of experience. Consistent with these relatively small disparities, other research suggests that differences in exposure to less experienced and less effective teachers contributes just 1 percentile point to socioeconomic gaps in achievement (Isenberg et al. 2013).

Similarly, funding formulae, staffing guidelines, compensation schedules, curricular vetting, and physical plant requirements are all typically standardized at an administrative level that would tend to compress differences in elementary schools across neighborhoods within the same district or even within the same state (Downey 2020; Guppy and Davies 2006). Consequently, variation in per-pupil expenditures, class size, and teacher salaries between different neighborhood contexts is not as pronounced as is often assumed in the literature on concentrated poverty and child development. For example, teachers employed by schools in the wealthiest quintile of American neighborhoods only earn about 3 to 4 percent more than teachers in other communities, on average; the teacher-pupil ratio among schools serving wealthy, poor, and middle income neighborhoods hovers around an average of 17 in all of these settings; and per-pupil spending differs by no more than about $300, on average, between any two neighborhood income quintiles (Owens and Candipan 2019). In general, inequality in school funding is relatively low—much lower than inequality in household incomes, for example (Corcoran et al. 2004).

Evidence also indicates that teachers devote greater effort toward increasing the number of students who meet minimum proficiency targets as opposed to challenging more advanced students with new content. For example, many teachers report spending considerably more time helping "struggling students" than engaging with "advanced students" whose success is taken for granted (Duffett et al. 2008; von Hippel et al. 2018). Elementary curricula, while variable across districts and schools, also typically focus on a fairly uniform set of foundational abilities during the early years (e.g., letter and word identification, decoding, counting, arithmetic). This curricular focus may be partly redundant with what advantaged children have already learned elsewhere and thus better tailored to the learning needs of students from disadvantaged communities.

The link between neighborhood context and elementary school quality may also be weaker than is commonly assumed because parents act upon limited information when choosing where to enroll their children. Although a large volume of evidence indicates that more advantaged families sort into neighborhoods in pursuit of quality schools (e.g., Hoxby 2003;

Lareau 2003; Owens 2016), they often make these decisions on the basis of factors that do not accurately reflect which schools provide the best education to their students (Abdulkadiroglu et al. 2014; Billingham and Hunt 2016).

Among the most powerful drivers of school choice are the racial composition and ability levels of the student body (Billingham and Hunt 2016), but neither of these characteristics may be very closely related to more defensible measures for the quality of a school's instructional regime. If parents are poor judges of what constitutes a high-quality education or they prioritize characteristics of schools that are largely unrelated to their effectiveness, this would attenuate the link between neighborhood composition and school quality (DeLuca and Rosenblatt 2010).

In New York City, for example, the most oversubscribed high schools are those whose students exhibit high achievement levels, even though there is little evidence that attending an oversubscribed school, compared with attending a less competitive school, improves student performance on advanced placement exams or state standardized tests (Abdulkadiroglu et al. 2014; Dobbie and Fryer 2014). Similarly, in the primary schools of Charlette-Mecklenberg, "average test scores alone contain almost no information about a school's causal impact on achievement" (Deming 2014:409), while in Boston, "school average test scores are only weakly related to school effectiveness" (Angrist et al. 2017:906). Findings consistent with these conclusions are also reported by Downey et al. (2008), Downey et al. (2019), and Hanselman and Fiel (2017), who show that school demographics are poor predictors of a school's influence on student learning at the elementary level. Thus, although research in this area is still sparse and somewhat mixed (c.f., Abdulkadiroglu et al. 2020, who find a closer association between average scores and value added at the high school level), the weight of the existing evidence suggests that the relationship between student composition and elementary school quality may not be so strong after all.

To summarize, the findings reviewed in this section conjure the controversial conclusion of Coleman et al. (1966:325) in their seminal study of educational inequality: that schools do not seem to contribute very much to socioeconomic disparities in achievement and "that the inequalities imposed on children by their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life at the end of school." This points toward the possibility that elementary schools might play a neutral or even an equalizing role in the transmission of neighborhood effects on student outcomes—a possibility that is rarely considered in the literature on concentrated poverty and academic achievement.

5. School Quality and its Measurement

Few prior studies investigate whether neighborhood effects are explained by differences in school quality, and among those that do, results fail to consistently provide evidence of mediation or interaction. These mixed results, however, may be due to potentially severe limitations of measurement, as prior studies have relied heavily on measures that do not accurately reflect school quality.

In general terms, school quality can be conceptually defined as the investment and consumption value of the education provided to students (Ladd and Loeb 2013). Investment value here refers to benefits in the form of greater knowledge, more advanced abilities, higher earnings, and so on, while consumption value refers to the immediate gratification that comes from attending school. Measuring a school's quality directly as the sum of its investment and consumption value is prohibitively difficult, as consumption benefits are often impossible to quantify and investment benefits may take years to realize. Consequently, all research must rely on proxies for school quality. But some proxies are better than others.

Prior research on the joint effects of neighborhood and school contexts has focused mainly on measures of *school composition*. For example, Wodtke and Parbst (2017), among others (e.g., Card and Rothstein 2007; Dobbie and Fryer 2011; Goldsmith 2009; Owens 2010), analyze whether neighborhood effects are explained by measures of the socioeconomic composition of schools to which resident children have access. Demographic characteristics of the student body, however, are not very strongly associated with the most important investment benefits of schooling (Coleman et al. 1966; Lauen and Gaddis 2013; Raudenbush 2004). Even if they were, school quality is still conceptually and empirically distinct from school composition, and thus research on contextual effects should not conflate their contribution to place-based educational disparities.

The three most common proxies for school quality in social science research involve measures of school inputs, school outputs, and processes internal to schools (Ladd and Loeb 2013). Input-based measures focus on school resources that are thought to influence their quality, such as spending per pupil, the number of teachers relative to the size of the student body, and teacher human capital. One advantage of input-based measures is that they are intuitive. They also circumvent any need to impose assumptions about which practices or student outcomes a school should prioritize when deploying its resources.

Nevertheless, measures based on school inputs suffer from several drawbacks. For example, per-pupil expenditures do not account for cost differences across districts, for differences in how money is spent on tangible resources, or for how spending on one versus another resource differentially contributes to the quality of the school environment (Hanushek

2003). Similarly, the teacher-pupil ratio accounts only for the quantity, and not the quality, of but one school input. In general, any proxy based on school inputs is limited by the difficulty associated with capturing all relevant inputs and appropriately weighting their contributions to the educational benefits of interest.

To avoid these limitations, an alternative measurement strategy uses a school's outputs to assess its quality. The outputs most widely used to assess school quality are achievement test scores. Although test scores certainly do not capture all of the investment and consumption benefits of interest, their use is justified on the grounds that they reflect one particularly important benefit—that is, the acquisition of knowledge and abilities—that predicts many others, such as higher earnings and better health in adulthood (Auld and Sidhu 2005; Ladd and Loeb 2013; Murnane and Levy 2006).

The challenge associated with using outputs, like test scores, as a proxy for school quality is that it can be difficult to isolate a school's contribution to these outcomes from other aspects of students' lives. Because children select into schools on the basis of many different factors that affect their outcomes, differences in achievement levels or proficiency rates across schools cannot simply be equated with differences in quality, as this would confound the contribution of the school environment with that of the family and other influences on children (Downey et al. 2008, 2019; Raudenbush and Eschmann 2015). Thus, a defensible proxy for school quality based on student outputs must correctly isolate a school's contribution to producing them.

These considerations motivate "value-added" approaches to measuring school quality, which attempt to estimate the gains in student achievement that can be uniquely attributed to attendance at a given school or to instruction from a particular teacher (Downey et al. 2008, 2019; Ladd and Loeb 2013). Value-added measures accurately capture the short-term benefits of schooling with respect to children's tested abilities (Angrist et al. 2017; Chetty et al. 2014a), and they also appear to predict several other long-term outcomes, such as college attendance and delayed childbearing (Chetty et al. 2014b). Despite these advantages, value-added measures are limited in that they do not capture the full breadth of outputs valued by individuals or society, and they do not reveal how schools actually produce the subset of outputs that are measured.

A third approach to evaluating school quality is based on factors internal to schools. These factors range from curricular offerings, administrative practices, and organizational structures to many different characteristics of a school's "climate," such as levels of parental involvement, students' sense of belonging, and classroom disorder (Ladd and Loeb 2013). During the early elementary years, the degree to which schools cultivate stable and orderly classroom environments would seem to be a basic precondition for effective teaching and learning. But there is little consensus about which internal factors are most important for improving child outcomes, and of the myriad processes, practices, and climatic features of schools considered in prior research, many appear to be weak predictors of academic achievement, net of other factors (Caldas 1993; Coleman et al. 1966; Wang and Degol 2016).

In conclusion, the concept of school quality is multidimensional and complex. To accommodate this complexity, we examine whether neighborhood effects are explained by several different measures capturing each of the three dimensions outlined previously: an output-based measure of school effectiveness, an input-based measure of school resources, and an internal measure of school disorder. We focus first on isolating the explanatory role of these factors from the potentially confounding influence of school composition. Then, for completeness, we also consider whether the quality and the composition of elementary schools may jointly explain neighborhood effects on academic achievement.

6. A Graphical Causal Model

Figure 1 presents a directed acyclic graph (DAG; Pearl 2009) that depicts a set of hypothesized causal relationships between neighborhood context, the elementary school environment, and academic achievement. In this figure and henceforth, A denotes the socioeconomic composition of a child's neighborhood, M denotes the quality of a child's elementary school, and Y denotes academic achievement. There is also a set of potentially confounding variables measured at a baseline time period, which are collectively denoted by C and include both individual and family characteristics like race, household income, and parental education. Lastly, there are measures that capture the socioeconomic composition of a child's elementary school, denoted by Z.

As indicated in Figure 1, neighborhood context is hypothesized to have an indirect effect on academic achievement via school quality, which is represented by the $A \rightarrow M \rightarrow Y$ and $A \rightarrow Z \rightarrow M \rightarrow Y$ paths. In other words, elementary school quality is thought to mediate, at least in part, the effect of neighborhood poverty on achievement.[1] Moreover, because A and M are both depicted to directly affect the outcome, Y, this figure is consistent with an interaction effect between neighborhood context and school quality.[2]

This figure also illustrates two methodological challenges that complicate analyses of mediation and interaction effects. The first is that consistently estimating these effects requires

[1] It is possible that, over a longer time horizon than is considered in the present study, a change in school quality could lead to changes in its composition and to the demographics of the local neighborhood. But given that neighborhood and school composition are relatively slow to turnover, this pattern of reverse causality is likely minimal in our data, and thus it is not represented in our graphical model.

[2] Effect interaction is sometimes depicted stylistically with a graph that includes an arrow from the exposure into the arrow representing the direct effect of the mediator on the outcome. In a DAG, however, interactions are represented implicitly, and "arrows into arrows" are not defined.

that both neighborhood and school selection are not confounded by unobserved determinants of child achievement, but confounding by unmeasured factors, denoted by U in Figure 1, is a ubiquitous threat to causal inference in contextual effects research. The second challenge is that, even if there were not any unobserved confounding, analyses of mediation and interaction must still resolve the problem of exposure-induced confounding by observed covariates. Exposure-induced confounding occurs when a variable affected by the exposure of interest confounds the effect of a focal mediator on the outcome. It arises in this study because the demographic composition of elementary schools is strongly affected by neighborhood context, as indicated by the $A \rightarrow Z$ path, and because school composition may in turn influence both the quality of the instructional regime and individual student achievement, as indicated by the $M \leftarrow Z \rightarrow Y$ path. It is problematic because consistently estimating mediation and interaction effects requires adjustment for exposure-induced confounders, but conventional methods that do so naively are biased (VanderWeele 2015).

We address these challenges in several ways. First, we use a new estimation procedure, termed regression-with-residuals (RWR), that can accurately evaluate mediation and interaction via school quality in the presence of exposure-induced confounders, like school composition (Wodtke and Almirall 2017; Wodtke et al. 2020; Wodtke and Zhou 2020; Zhou and Wodtke 2019). Second, we use this approach not only to isolate the explanatory role of school quality from that of school composition but also to evaluate whether these factors may jointly explain neighborhood effects on academic achievement. Third, we measure and adjust for the most powerful joint predictors of neighborhood attainment, school selection, and academic achievement together with several factors that proxy for unmeasured determinants of contextual selection. And finally, we combine RWR with a formal sensitivity analysis to construct a range of effect estimates under different hypothetical patterns of unobserved confounding. These features of our research design are outlined in detail below.

7. Methods

7.1. Data

To investigate whether neighborhood effects are explained by differences in the quality of elementary schools, we use data from the ECLS-K linked to information from the U.S. Census, the National Center for Education Statistics Common Core of Data, and the Private School Universe Survey.[3] The ECLS-K is a longitudinal study based on a nationally representative sample of schools and the children within them. All schools, public or private, that offered a kindergarten program were eligible for sample selection. Within selected schools, children of kindergarten age were sampled with approximately equal probability, except for those classified as Asian or Pacific Islanders, who were oversampled to meet precision goals for all racial and ethnic subgroups. The total base-year sample in the ECLS-K consists of 22,670 children attending 1,270 schools.[4]

For a random subset of the total base-year sample, the ECLS-K collected information on academic achievement in both the fall and spring of kindergarten (1998-99) and again in both the fall and spring of 1st grade (1999-2000). By collecting data at the beginning and at the end of the school year in kindergarten and 1st grade, the ECLS-K allows for seasonal learning comparisons, which we use to construct our output-based measure of school effectiveness. The analytic sample

[3] The data used in this analysis are based on restricted-access files from the U.S. Institute for Education Sciences, which were obtained under special contractual arrangements designed to protect the privacy of respondents. These data are not available from the authors.

[4] All sample sizes are rounded to the nearest ten in accordance with U.S. Department of Education disclosure risk guidelines.

for this study therefore includes the subset of $n = 6{,}040$ children in $k = 310$ schools that were selected for participation in both the fall and spring assessments during the first two years of elementary school. Following these early grade assessments, sample members were then tracked through the end of middle school, with additional tests administered during the spring of 3rd grade (2002), the spring of 5th grade (2004), and the spring of 8^{th} grade (2007), which we use to construct our outcome measures.

7.2. Measures

The outcome of interest in this study is academic achievement. We measure achievement with item-response theory (IRT) theta scores on ECLS-K assessments of math and reading abilities. IRT theta scores provide an equal-interval, vertically scaled measure of achievement that is capable of capturing student learning over time.[5] Both the math and reading assessments have desirable psychometric properties, including high reliability, high validity, and low differential item functioning (Pollock et al. 2005).

The exposure of interest is the socioeconomic composition of a child's home census tract, which we use to approximate their neighborhood. To construct this measure, we match children in the ECLS-K to their census tracts using a restricted-access geocode file. Demographic information on census tracts comes from the GeoLytics Neighborhood Change Database (NCDB), which contains tract-level data from the U.S. Census that have been harmonized over

[5] These scores are estimated from an item-response model in which the probability that a child answers a test question correctly is a function of their ability (theta) and then the question's difficulty, discrimination, and guessability. Theta scores avoid the scaling problems that afflict prior analyses of the ECLS-K because they properly isolate changes in a child's ability from differences in the properties of test questions (von Hippel and Hamrock 2019).

time (GeoLytics 2013).[6] With these data, we apply principal components analysis to compute a composite index of neighborhood disadvantage based on the following tract characteristics: the poverty rate, the unemployment rate, the proportion of families receiving cash assistance, median household income, the proportion of households that are female-headed, aggregate levels of education, and the occupational structure. This measure is standardized to have zero mean and unit variance, and it is scaled so that higher values represent more disadvantaged neighborhoods.

The mediator of interest is elementary school quality, which we operationalize using three different measures: effectiveness, resources, and disorder. First, we measure school effectiveness as the difference between a school's average learning rate among its 1st grade students during the school year and the average learning rate among those same students during the previous summer. This measure captures the degree to which a school increases its students' learning rates above those that would prevail were its students not in school under the following assumptions: (i) any effects of non-school factors on achievement must operate similarly during both the school year and the summer, and (ii) schools must not have sizeable "spillover" effects on summer learning. By isolating the impact of each school on its students' learning from potentially contaminating non-school factors, this measure reflects the effectiveness of a school's instructional regime more accurately than other output-based measures, such as average achievement levels or the proportion of students meeting targeted proficiency standards (Raudenbush 2004; von Hippel 2009).[7]

[6] For intercensal years, we impute tract characteristics using linear interpolation.
[7] In Part F of the Online Supplement, we present results based on multiple alternative measures of school effectiveness. Our preferred measure that we describe here is prioritized throughout the main text.

Following Downey et al. (2008, 2019), we estimate our measure of school effectiveness from the ECLS-K using a multilevel model of test score growth in which scores on tests administered during kindergarten and 1st grade are nested within children who are in turn nested within schools. From this model, we predict the monthly learning rates of students in each school during 1st grade and during the previous summer, and then school effectiveness is measured by taking the difference between them.[8] We compute separate measures for reading and math achievement to allow for the possibility that an elementary school's effectiveness may differ depending on the subject matter. In all multivariate analyses, these measures are standardized to have zero mean and unit variance, and they are scaled so that higher values represent more effective schools. Technical details underlying this measurement strategy are presented in Part A of the Online Supplement.

Second, we use principal components analysis to compute a composite index of school resources based on the following inputs to elementary schools: expenditures per pupil, the ratio of full-time equivalent teachers to the number of students, the average years of work experience among teachers, the proportion of teachers with an advanced degree, and average teacher base salary. This measure is standardized to have zero mean and unit variance, and it is scaled so that higher values represent elementary schools with greater resources—that is, better funding, smaller classes, and teachers with more experience and training.[9]

---

[8] The ECLS-K assessments of reading and math abilities are based on two-stage adaptive tests designed for students in grades K-8. Our measure of school effectiveness during 1st grade is therefore not distorted by any ceiling or floor effects. But even when value-added measures are based on test scores that suffer from fairly extreme censoring, research suggests that they are highly robust (Koedel and Betts 2010).

[9] A parallel analysis that examines each school input separately rather than combined as part of a composite index yields substantively similar results. For parsimony, we focus on the composite index throughout.

Third, we construct a measure of school disorder using teacher assessments of absenteeism and classroom misbehavior. The ECLS-K asked teachers how many of the children in their class were absent on a typical school day. It also asked them to rate the behavior of their class on a five-point scale with responses ranging from “the group misbehaves very frequently” to “the group behaves exceptionally well.” We aggregate, standardize, and then average responses to these questions within schools to generate a composite measure, where higher values denote more “disorderly” learning environments with frequent misbehavior and chronic absenteeism.[10]

We also measure and adjust for a set of baseline confounders that include both child and family characteristics. Specifically, we include adjustments for a child’s gender, race, and birth weight. Gender is coded as an indicator variable, one for male and zero for female. Race is expressed as a series of indicator variables that capture whether a child identifies as white, black, Hispanic, Asian, or another race. Birth weight is also coded as a binary indicator, one if a child weighed less than 88 ounces at birth and zero otherwise.

We additionally adjust for the following family characteristic at baseline: a mother’s age and marital status at the time of her child’s birth, family income, parental education and employment status, the level of cognitive stimulation a child received at home, an indicator of parental involvement with their child’s education, and maternal depressive symptoms. Maternal age is measured in years. Parental employment status is expressed as a series of indicator variables capturing whether each parent is “working at least 35 hours per week,” “working less than 35 hours per week,” or involved in some other arrangement. Family income is measured in

[10] In addition to the composite index described here, we also performed analyses based on separate measures for absenteeism and classroom misbehavior, and the results were similar.

dollars, which we transform using the natural log in all multivariate analyses. The highest level of education attained by either parent is expressed as a series of indicator variables for having "less than a high school diploma," "a high school diploma," "a vocational or technical degree," "some college education," "a bachelor's degree," or a "graduate degree." The level of cognitive stimulation provided in the household is measured using the HOME inventory (Caldwell and Bradley 1984). The level of parental involvement in their child's education is measured as a count of more than 20 different activities in which a parent may be engaged, such as attending parent-teacher association meetings or participating in extracurricular activities (Greenman et al. 2011). Maternal depressive symptoms are measured using an abbreviated version of the Center for Epidemiological Studies - Depression Scale (CES-D; Radloff 1977). In all multivariate analyses, the baseline confounders are centered at their sample means.

Finally, we measure and adjust for the socioeconomic and racial composition of a child's elementary school, which are potentially exposure-induced confounders. Specifically, we adjust for the percentage of students in a school who are eligible for a free lunch through the U.S. National School Lunch Program. This measure is an approximate school-level poverty rate, as a student's family must have an income at or below 130 percent of the federal poverty threshold in order to qualify for a free lunch. In addition, we also adjust for the percentage of students at a child's school who identify as nonwhite.

Analyses of mediation and interaction require sequential measurements of key variables (VanderWeele 2015). Figure 2 depicts the longitudinal measurement strategy we use to ensure appropriate temporal ordering of the confounders, exposure, mediator, and outcome.

Specifically, we first measure the baseline confounders (C) at the fall of kindergarten.[11] We then measure neighborhood context (A) the following spring. Next, we construct measures of school composition (Z) and school quality (M) that cover 1st grade. Finally, we use measures of academic achievement (Y) taken during the spring of 3rd grade as our focal outcome. Thus, our data are sequentially ordered as follows: $\{C, A, Z, M, Y\}$. Part B of the Online Supplement presents results from parallel analyses of academic achievement measured later during 5th and 8th grade, which are similar to those based on the 3rd grade assessments that we prioritize here.

Although our analyses focus narrowly on elementary schools during the early grades, this is the period for which the most accurate data on school quality are available. Moreover, this is also a critical stage in any child's formal education, as it aims to cultivate a set of foundational abilities that serve as important precursors for later learning.

7.3. Estimands

To investigate whether school quality explains the effect of neighborhood poverty on academic achievement, we decompose a measure for the overall impact of living in a disadvantaged neighborhood into components due to mediation versus interaction, which is accomplished using potential outcomes notation and the counterfactual framework (Rubin 1974; VanderWeele 2014; VanderWeele et al. 2014). Let Y_a denote a child's achievement level in 3rd grade had they previously been exposed to the level of neighborhood disadvantage given by a during kindergarten, possibly contrary to fact. Similarly, let M_a denote the quality of a child's school during 1st grade—that is, its level of effectiveness, resources, or disorder—under prior exposure

[11] Some baseline confounders could only be measured at the spring, rather than the fall, of kindergarten in the ECLS-K. These include family income, parental involvement, and maternal depression.

to the level of neighborhood disadvantage given by a.[12] Finally, let $M^R_{a|C}$ denote a level of school quality randomly selected from its distribution under neighborhood exposure status a conditional on baseline covariates C.

Given this notation, consider the following estimand:

$$RATE = E\left(Y_{a^* M^R_{a^*|C}} - Y_{a M^R_{a|C}}\right),$$

which is similar to an average total effect except that it is defined in terms of both a contrast between neighborhood contexts and a randomized intervention on school quality. Specifically, when $a^* > a$, this effect gives the expected difference in achievement if children were exposed to a more versus less disadvantaged neighborhood, with school quality randomly selected from its distribution under each of these alternative exposures. It is therefore referred to as a "randomized intervention analogue" of the average total effect (VanderWeele et al. 2014).

The $RATE$ can be decomposed into direct and indirect components as follows:

$$RATE = E\left(Y_{a^* M^R_{a|C}} - Y_{a M^R_{a|C}}\right) + E\left(Y_{a^* M^R_{a^*|C}} - Y_{a^* M^R_{a|C}}\right) = RNDE + RNIE.$$

The first term in this decomposition, $RNDE = E\left(Y_{a^* M^R_{a|C}} - Y_{a M^R_{a|C}}\right)$, is a randomized intervention analogue of a natural direct effect. In words, the $RNDE$ is the expected difference in achievement under exposure to a more versus less disadvantaged neighborhood if children were subsequently exposed to a level of school quality randomly selected from its distribution among those in less disadvantaged neighborhoods. It captures an effect of neighborhood context on achievement that is not due to mediation via the quality of elementary schools.

[12] For expositional simplicity, we use the term "school quality" to generically refer to effectiveness, recourses, or disorder throughout this section.

The second term in this decomposition, $RNIE = E\left(Y_{a^*M^R_{a^*|C}} - Y_{a^*M^R_{a|C}}\right)$, is a randomized intervention analogue of a natural indirect effect. It represents the expected difference in achievement if children were first exposed to a more disadvantaged neighborhood and then were subsequently exposed to a level of school quality randomly selected from its distribution in these more disadvantaged neighborhoods rather than from its distribution in less disadvantaged neighborhoods. The $RNIE$ captures an effect of neighborhood context on achievement that is due specifically to mediation via school quality.

The $RNDE$ can be further decomposed into a controlled direct effect and an interaction effect occurring in the absence of mediation:

$$RNDE = E(Y_{a^*m} - Y_{am}) + \left\{E\left(Y_{a^*M^R_{a|C}} - Y_{aM^R_{a|C}}\right) - E(Y_{a^*m} - Y_{am})\right\}$$

$$= CDE + RINT_{\text{ref}}.$$

The first term in this expression, $CDE = E(Y_{a^*m} - Y_{am})$, is a controlled direct effect. It represents the expected difference in achievement if children were exposed to a more versus less disadvantaged neighborhood and then were all exposed to elementary schools of the same quality m.

The second term, $RINT_{\text{ref}} = \left\{E\left(Y_{a^*M^R_{a|C}} - Y_{aM^R_{a|C}}\right) - E(Y_{a^*m} - Y_{am})\right\}$, is a reference interaction effect, which captures a component of the overall effect due to an interaction between neighborhood context and school quality that occurs absent any mediation. Specifically, it describes how the direct effect of living in a more versus less disadvantaged neighborhood differs depending on whether children are exposed to a level of school quality randomly selected from its distribution in less disadvantaged neighborhoods, $M^R_{a|C}$, as opposed to some fixed level, m. Because interactions are symmetrical, the $RINT_{\text{ref}}$ also describes how the effect of attending an elementary school with quality $M^R_{a|C}$ versus m differs depending on whether children live in more versus less disadvantaged neighborhoods. It captures a component of the overall effect due to interaction in the absence of mediation because it may be nonzero even if neighborhood context does not affect school quality.

Similarly, the $RNIE$ can be further decomposed into another effect due specifically to mediation and then an effect due to interaction occurring together with mediation:

$$RNIE = E\left(Y_{aM^R_{a^*|C}} - Y_{aM^R_{a|C}}\right) + \left\{E\left(Y_{a^*M^R_{a^*|C}} - Y_{aM^R_{a^*|C}}\right) - E\left(Y_{a^*M^R_{a|C}} - Y_{aM^R_{a|C}}\right)\right\}$$

$$= RPIE + RINT_{\text{med}}.$$

The first term in this expression, $RPIE = E\left(Y_{aM^R_{a^*|C}} - Y_{aM^R_{a|C}}\right)$, is a randomized intervention analogue of a pure indirect effect. It represents a component of the overall effect due only to mediation via school quality.

The second term, $RINT_{\text{med}} = E\left(Y_{a^*M^R_{a^*|C}} - Y_{aM^R_{a^*|C}}\right) - E\left(Y_{a^*M^R_{a|C}} - Y_{aM^R_{a|C}}\right)$, is a mediated interaction effect. It captures a component of the overall effect due to interaction between neighborhood context and school quality that occurs jointly with mediation. Specifically, it describes how the effect of living in a more versus less disadvantaged neighborhood differs depending on whether children are exposed to a level of school quality randomly selected from its distribution in more disadvantaged neighborhoods rather than from its distribution in less disadvantaged neighborhoods. Symmetrically, the $RINT_{\text{med}}$ also describes how the effect of exposure to a level of school quality randomly selected from its distribution in more versus less disadvantaged neighborhoods differs depending on the neighborhood environment in which a child lives. It captures a component of the overall effect due to interaction and mediation operating together because, in the absence of mediation, the distributions of $M^R_{a^*|C}$ and $M^R_{a|C}$ would be identical and thus the $RINT_{\text{med}}$ would equal zero.

To summarize, combining the expressions outlined previously yields the following additive decomposition:

$$RATE = RNDE + RNIE = CDE + RINT_{\text{ref}} + RPIE + RINT_{\text{med}},$$

where the CDE captures an effect of neighborhood context due to direct causation; the $RINT_{\text{ref}}$ captures an effect due to interaction but not mediation; the $RPIE$ captures an effect due to mediation; and the $RINT_{\text{med}}$ captures an effect due to both mediation and interaction operating jointly.

We focus on a decomposition defined in terms of randomized interventions on school quality because its components can be identified under more defensible assumptions than those required of other effect decompositions. In particular, unlike the components of alternative decompositions (e.g., VanderWeele 2014, 2015), all of the effects outlined previously can be identified in the presence of exposure-induced confounders. Nevertheless, identifying and estimating randomized intervention analogues of direct, indirect, and interaction effects still requires strong assumptions, as we explain in detail below.

7.4. Identification

The effects outlined previously can be identified from observed data under a set of so-called "ignorability" assumptions (VanderWeele 2014; VanderWeele et al. 2014), which can be formally expressed as follows:

$$Y_{am} \perp A|C; Y_{am} \perp M|C, A, Z; \text{ and } M_a \perp A|C.$$

In this notation, $\perp$ denotes statistical independence. The first of these assumptions states that the potential outcomes of the exposure and mediator, Y_{am}, must be independent of the observed exposure conditional on the baseline confounders. The second assumption states that the same potential outcomes must also be independent of the observed mediator conditional on the baseline confounders, prior exposure, and the exposure-induced confounders. Finally, the third assumption states that the potential outcomes for the mediator under prior exposure, M_a, must be independent of the observed exposure conditional on the baseline confounders.[13] These assumptions would all be satisfied if there were not any unobserved confounding of the exposure-outcome, mediator-outcome, or exposure-mediator relationships.

These are strong assumptions, and if they are not satisfied in this analysis, then estimates of the effects outlined previously may be biased. We attempt to mitigate confounding bias by adjusting for an extensive set of child and family covariates measured at baseline. In addition to these adjustments, we also control for baseline (i.e., fall of kindergarten) measures of academic achievement at both the child- and school-levels, which proxy for unobserved determinants of contextual selection and account for the possibility that past performance affects future residential or school choices. Finally, we adjust for post-exposure measures of school composition that may confound the effect of the mediator on the outcome. Analyses of mediation and interaction that adjust not only for individual characteristics at baseline but also for lagged

[13] Identifying the components of alternative decompositions—for example, one in which the average total effect, $E\left(Y_{a^*M_{a^*}} - Y_{aM_a}\right)$, is expressed as the sum of a natural direct and a natural indirect effect, $E\left(Y_{a^*M_a} - Y_{aM_a}\right) + E\left(Y_{a^*M_{a^*}} - Y_{a^*M_a}\right)$, without invoking the concept of a randomized intervention on the mediator—requires the additional assumption that $Y_{am} \perp M_{a^*}|C$ (VanderWeele 2014). This assumption is problematic, and we therefore avoid it, because an independence restriction on the joint distribution of Y_{am} and M_{a^*} is violated when there are exposure-induced confounders of the mediator-outcome relationship, whether these confounders are observed or not.

measures of the outcome and post-exposure covariates provide some of the strongest protection against confounding bias in observational research (Pearl 2000; VanderWeele 2015). Nevertheless, we also conduct a formal sensitivity analysis that evaluates whether our findings are robust to hypothetical patterns of unobserved confounding.

7.5. Estimation

The direct, indirect, and interaction effects of interest can be estimated from a set of regression models for the mediator, outcome, and exposure-induced confounders. The first model is for the conditional mean of school quality given neighborhood context and the baseline confounders. It can be formally expressed as follows:

$$E(M|C,A) = \theta_0 + \theta_1(C - \alpha_0) + \theta_2 A, \quad (1)$$

where $\alpha_0 = E(C)$ and thus $C - \alpha_0$ represents a transformation of the baseline confounders in which they are centered around their marginal means.

The second model is for the conditional mean of academic achievement given neighborhood context, school quality, the baseline confounders, and finally, measures of school composition, which may be exposure-induced confounders. It can be formally expressed as follows:

$$E(Y|C,A,Z,M) = \lambda_0 + \lambda_1(C - \alpha_0) +$$

$$\lambda_2 A + \lambda_3\big(Z - (\beta_0 + \beta_1 C + \beta_2 A)\big) + M(\lambda_4 + \lambda_5 A), \quad (2)$$

where $E(Z|C,A) = \beta_0 + \beta_1 C + \beta_2 A$ and thus $Z - (\beta_0 + \beta_1 C + \beta_2 A)$ represents a residual transformation of the exposure-induced confounders in which they are centered around their conditional means given prior exposure and the baseline confounders. This model is similar to a conventional linear regression except that it subsumes another model for $E(Z|C,A)$, which is used to residualize the exposure-induced confounders with respect to the observed past.

Under the ignorability assumptions outlined previously and under the assumption that our models for $E(M|C,A)$, $E(Z|C,A)$, and $E(Y|C,A,Z,M)$ are all correctly specified, the controlled direct effect is equal to

$$CDE = (\lambda_2 + \lambda_5 m)(a^* - a),$$

the reference interaction effect is equal to

$$RINT_{\text{ref}} = \lambda_5(\theta_0 + \theta_2 a - m)(a^* - a),$$

and the $RNDE$ is equal to the sum of these two expressions. Under the same set of assumptions, the pure indirect effect is equal to

$$RPIE = \theta_2(\lambda_4 + \lambda_5 a)(a^* - a),$$

the mediated interaction effect is equal to

$$RINT_{\text{med}} = \theta_2\lambda_5(a^* - a)^2,$$

and the $RNIE$ is equal to the sum of these two expressions. Lastly, the sum of the $RNDE$ and $RNIE$ gives the overall effect, or $RATE$. A derivation of these expressions is provided in Part C of the Online Supplement.

In the results section below, we focus on effects that contrast residence in a more disadvantaged neighborhood at the 80th percentile of the exposure distribution with residence in less disadvantaged neighborhood at the 20th percentile. In addition, we evaluate the controlled direct effect and reference interaction effect by setting each measure of school quality at its sample median.

We estimate these effects using the method of regression-with-residuals (RWR; Wodtke 2020; Wodtke and Almirall 2017; Wodtke et al. 2020; Wodtke and Zhou 2020; Zhou and

Wodtke 2019), which is implemented as follows. First, the model for $E(M|C, A)$ is estimated by least squares after centering the baseline confounders around their sample means. Second, the models for $E(Z|C, A)$ are estimated by least squares and used to compute residual terms for the exposure-induced confounders. Third, the model for $E(Y|C, A, Z, M)$ is estimated by regressing the outcome, Y, on $\{\tilde{C}, A, Z^{\perp}, M, AM\}$, where $\tilde{C} = C - \bar{C}$ represents the baseline confounders after centering them around their sample means and $Z^{\perp} = Z - \hat{E}(Z|C, A)$ denotes the residualized exposure-induced confounders. Finally, the estimated parameters from these models are used to construct the effects of interest with the formulas outlined previously.

The advantage of RWR over alternative estimation strategies is that it deals properly with exposure-induced confounders of the mediator-outcome relationship. In the presence of exposure-induced confounders, conventional regression and matching estimators that adjust for these variables naively are biased and inconsistent. This is because naively adjusting for confounders that are affected by prior exposure can engender bias due to over-control of intermediate pathways and endogenous selection (Elwert and Winship 2014; VanderWeele 2015). RWR avoids these biases by residualizing the exposure-induced confounders with respect to the observed past before including them in a regression model for the outcome. Adjusting for these residual terms sufficiently controls for mediator-outcome confounding while avoiding bias due to over-control or endogenous selection, as the residuals are orthogonal to prior exposure by design. In this way, RWR properly isolates the explanatory role of school quality from the potentially confounding influence of school composition.

Nevertheless, for completeness, we also conduct an analysis in which we shift away from the goal of isolating the explanatory role of school quality and instead examine whether school quality and school composition *jointly* explain neighborhood effects. To this end, we additionally compute versions of direct and indirect effects that capture mediation via both the quality and the composition of elementary schools. Under correctly specified models and provided that all of the confounding assumptions outlined previously hold for the set of mediators, $\{M, Z\}$, these effects can be constructed as follows:

$$RNDE^{\dagger} = RNDE - \beta_2\lambda_3(a^* - a),$$

which captures a direct effect not mediated by either school quality or school composition, and

$$RNIE^{\dagger} = RNIE + \beta_2\lambda_3(a^* - a),$$

which captures an indirect effect mediated through both school quality and school composition.

7.6. Measurement Error, Missing Data, and Variance Estimation

Our measure of school effectiveness suffers from known error. This is because it is computed from sample rather than population data at the school level and because it is based on achievement test scores that are themselves subject to measurement error. When a mediator is measured with random error, this can lead to attenuation bias in estimates of indirect effects and inflationary bias in estimates of direct effects. To correct for measurement error in school effectiveness, we implement a classical error-in-variables adjustment when fitting the outcome model (Draper and Smith 1998).[14] For this adjustment, we assume that the exposure and confounders are measured without error, that the mediator is measured with a reliability of $r_M =$

[14] The classical error-in-variables correction assumes that $E(Y|\mathbf{X}) = \mathbf{X}\boldsymbol{\lambda}$ and that $\widetilde{\mathbf{X}} = \mathbf{X} + \mathbf{U}$, where $\widetilde{\mathbf{X}} = \{\tilde{C}, A, Z^{\perp}, M, AM\}$ are the observed values of the predictors, $\mathbf{X}$ are the true values, and $\mathbf{U}$ are a set of independent and identically distributed random errors. In this situation, a consistent estimator for $\boldsymbol{\lambda}$ is $\left(\widetilde{\mathbf{X}}^{\mathbf{T}}\widetilde{\mathbf{X}} - \mathbf{C}\right)^{-\mathbf{1}}\widetilde{\mathbf{X}}^{\mathbf{T}}\mathbf{Y}$, where $\mathbf{C}$ is a diagonal matrix with elements equal to $N(1 - r_k)Var\left(\tilde{X}_k\right)$ and where N is the sample size, r_k is the reliability of the k^{th} predictor, $Var\left(\tilde{X}_k\right)$ is the total variance of the k^{th} predictor.

0.7, and that the exposure-by-mediator interaction term is measured with a reliability of $\frac{(r_A \times r_M)+\rho_{AM}^2}{1+\rho_{AM}^2}$, where $r_A = 1$ denotes the assumed reliability of the exposure and ρ_{AM} denotes the correlation between the exposure and mediator (Bohrnstedt and Marwell 1978). An assumed reliability of $r_M = 0.7$ for our measure of school effectiveness is consistent with estimates reported in prior research (e.g., von Hippel 2009). Moreover, experimentation with a range of plausible reliabilities generated substantively similar results, which are presented in Part D of the Online Supplement.

To adjust for the bias and inefficiency that may result from missing data, we simulate missing values for all variables using multiple imputation with 50 replications and then combine our estimates following Rubin (1987).[15] Overall, the proportion of missing information in this analysis is about 24 percent, which is due to a combination of panel attrition and item-specific nonresponse. Standard errors are computed using the cluster bootstrap with 500 replications in order to adjust for the clustering of children within schools. Finally, because it oversampled students of certain racial and ethnic groups, the ECLS-K is based on a sample design with unequal probabilities of selection. Nevertheless, we focus on unweighted estimates because they are very similar to results from a weighted analysis and are also more precise. When estimating causal effects from linear models, sampling weights are unnecessary and inefficient if the models sufficiently control for those factors that determine the unequal selection probabilities or if the sample design is otherwise ignorable (Solon et al. 2015; Winship and Radbill 1994). In this

[15] We also performed an analysis following von Hippel (2007) in which we multiply imputed all missing data but then dropped cases with missing values on the mediator or outcome prior to fitting Equations 1 and 2. Results from this analysis are similar to those we report here.

situation, unweighted estimates are preferred because they are both consistent and relatively more efficient than weighted estimates.[16]

7.7. Summary

The methods we employ to examine mediation and interaction extend conventional approaches (e.g., Alwin and Hauser 1975; Baron and Kenney 1986) in several ways. First, we delineate our estimands and identification assumptions precisely using counterfactual notation. Second, we introduce a decomposition that permits an assessment of mediation and interaction simultaneously, whereas conventional approaches typically assume away the latter to evaluate the former. Third, we resolve the problem of exposure-induced confounding, which is also typically assumed away—in most cases, naively. All of these extensions align our analytic approach more closely with theoretical models of contextual effects on academic achievement.

8. Results

8.1. Sample Characteristics

Table 1 presents descriptive statistics for math and reading test scores that have been standardized with respect to their mean and standard deviation at the fall of kindergarten. Several patterns are evident in these data, all of which are consistent with other recent studies of student

[16] Replication code is available at https://github.com/gtwodtke/nhood_mediation_schl_qual. Data from the ECLS-K can be obtained via licensing agreement with the U.S. Institute for Education Sciences. Instructions for accessing these data are provided at https://nces.ed.gov/pubsearch/licenses.asp. Data from the NCDB can be licensed from GeoLytics, Incorporated (https://geolytics.com/) for a modest fee. Other data on schools from the Common Core and Private School Universe Surveys are publicly available at https://nces.ed.gov/ccd/files.asp and https://nces.ed.gov/surveys/pss/pssdata.asp, respectively.

learning trajectories (e.g., von Hippel et al. 2018; von Hippel and Hamrock 2019). First, students learn at a rapid pace early on during elementary school, and they learn faster during the school year than during the summer. Second, the variance, or inequality, in math and reading abilities is substantial at the start of kindergarten but then shrinks over the course of students' elementary education. For example, by the spring of 3rd grade, the standard deviation of math test scores is about 19% smaller than it was at the fall of kindergarten. Finally, during kindergarten and 1st grade, inequality in student achievement appears to shrink primarily during the school year and to stagnate, or possibly even increase, over the summer. Taken together, these findings suggest that elementary schooling has an equalizing effect on reading and math abilities, while factors outside of school have disequalizing effects.

Table 2 presents descriptive statistics for child, neighborhood, and school characteristics. They indicate that sampled children attend – during 1st grade – schools in which about 38% of students receive a free lunch and about 40% are nonwhite, on average. In addition, these results also indicate that the average school raises its students' monthly learning rates by about 0.11 and 0.17 standard deviations in math and reading, respectively, compared to the rates that would prevail were students not in school. There is, however, considerable variation in school effectiveness around these averages, as indicated by the measure's sizeable standard deviation.

Table 3 presents descriptive statistics for the family covariates considered in this analysis. At the start of kindergarten, sampled children lived in households with an average income of about $49,000 and roughly 5 members. About 35% lived with parents whose highest level of education was a high school diploma or less, while about 31% had a parent with at least a bachelor's degree. A majority (67%) of sampled children had a mother who was married at the time of childbirth.

8.2. Neighborhood Context and School Quality

Figure 3 describes the relationship between our different indicators of school quality and neighborhood context. Specifically, it displays point estimates and confidence intervals from linear regressions of school effectiveness, resources, and disorder on neighborhood disadvantage, with all variables standardized to have zero mean and unit variance.[17] For measures of school effectiveness, the figure plots partial regression slopes that are adjusted for the average ability levels of students in each school at the fall of kindergarten to account for the possibility that children in poor neighborhoods may learn more from their elementary schooling, regardless of its quality, simply because they begin school with fewer academic skills. For measures of school resources and disorder, the figure displays unadjusted regression estimates.

Contrary to expectations, Figure 3 reveals rather modest relationships between neighborhood context and our different measures of school quality. For example, the upper panel of the figure shows that, conditional on school-average abilities at baseline, there is not a very strong relationship between neighborhood context and school effectiveness during 1st grade. Whether assessed in terms of contributions to math or reading abilities, the regression line is nearly flat and its slope is not appreciably different from zero. This indicates that, after adjusting for initial conditions at the start of kindergarten, children in more versus less disadvantaged neighborhoods attend elementary schools that are, on average, similarly effective at improving their students' academic skills during 1st grade. Consistent with these results, the bivariate relationship between neighborhood disadvantage and school resources is also fairly weak.

[17] Estimates from thin plate splines, which allow for complex forms of nonlinearity, were substantively similar.

Although the regression line has a negative slope, indicating that schools serving disadvantaged neighborhoods have fewer resources on average, this relationship is modest in substantive terms.

School disorder is the only measure that appears strongly linked with neighborhood context, where children from disadvantaged neighborhoods attend elementary schools with significantly more absenteeism and misbehavior than children from advantaged neighborhoods. These disruptions, however, seem not to significantly hinder a school's ability to provide effective instruction in reading and math—perhaps because elementary schools serving disadvantaged neighborhoods have developed strategies for mitigating their impact on student learning.

Figure 4 presents kernel density plots for our different measures of school quality across tertiles of neighborhood disadvantage, with all variables constructed as above. These density plots complement the regressions displayed in the previous figure by showing the full distribution of school effectiveness, resources, and disorder across subgroups of students living in more versus less disadvantaged neighborhoods. They reveal a substantial amount of overlap, especially for the distributions of school effectiveness and resources, indicating that children living in different neighborhood contexts attend elementary schools of broadly similar quality on these measures.

In sum, our descriptive analyses do not provide very much evidence that children in disadvantaged neighborhoods are frequently trapped in low-quality schools, while children in more advantaged neighborhoods disproportionately benefit from access to high-quality schools, as is commonly hypothesized in the literature on concentrated poverty and educational inequality. Rather, we find that children in disadvantaged neighborhoods attend elementary schools that are not terribly different from the schools attended by children in advantaged neighborhoods, at least with respect to their effectiveness and resources. Elementary schools serving children from disadvantaged neighborhoods are more disorderly, but this difference in climate does not appear to interfere with a school's ability to cultivate an effective instructional regime during 1st grade. These findings cast some initial doubt on the hypothesis that attendance at low-quality elementary schools mediates the effects of living in a disadvantaged neighborhood on academic achievement.

8.3. Effects of Neighborhood Context on Academic Achievement

Tables 4 and 5 present estimates for the effects of living in a disadvantaged neighborhood at the end of kindergarten on math and reading test scores, respectively, measured later during 3rd grade. Consistent with expectations and prior research, the total effect estimates in these tables suggest that exposure to a disadvantaged neighborhood has a considerable negative impact on academic achievement. Specifically, estimates of the $RATE$ indicate that earlier exposure to a more disadvantaged neighborhood at the 80th percentile of the treatment distribution, rather than a less disadvantaged neighborhood at the 20th percentile, reduces performance on 3rd grade math and reading assessments by about one-seventh of standard deviation. These effects are substantively large and statistically significant at stringent thresholds. To put them in perspective, note that they are roughly equivalent in magnitude to missing about one month of instruction during elementary school.

Contrary to expectations, however, the direct and indirect effect estimates provide little evidence that the overall impact of neighborhood poverty on academic achievement is mediated by school quality. For example, estimates of the $RNDE$ indicate that exposure to a more disadvantaged neighborhood at the 80th percentile of the treatment distribution, rather than a less disadvantaged neighborhood at the 20th percentile, would still reduce performance on math and reading assessments by about one-seventh of standard deviation, even after an intervention to shift the distribution of school effectiveness to that in less disadvantaged neighborhoods. Relatedly, estimates of the CDE indicate that living in a disadvantaged neighborhood would also reduce test scores by roughly the same margin even after an intervention to place students in schools at the 50th percentile of the effectiveness distribution. The direct effects are all substantively large, statistically significant at stringent thresholds, and similar to the total effect estimates discussed previously.

Conversely, estimates of the $RNIE$ indicate that, if children lived in more disadvantaged neighborhoods, an intervention to shift the school effectiveness distribution from that in less disadvantaged neighborhoods to that in more disadvantaged neighborhoods would barely change their test scores at all. Similarly, estimates of the $RPIE$ indicate that, if children lived in less disadvantaged neighborhoods, their test scores also wouldn't change much at all after an intervention to shift the distribution of school effectiveness from that in less disadvantaged neighborhoods to that in more disadvantaged neighborhoods. Both the $RNIE$ and $RPIE$ are substantively small and fail to reach conventional significance thresholds, despite being precisely estimated.

Also contrary to expectations, estimates of interaction effects provide little evidence that neighborhood context dampens or amplifies the effects of school effectiveness on achievement. Specifically, estimates for the $RINT_{\mathrm{ref}}$, which captures interaction in the absence of mediation, and for the $RINT_{\mathrm{med}}$, which captures interaction operating jointly with mediation, are all close to zero and fail to approach conventional thresholds for statistical significance. This suggests that living in a more disadvantaged neighborhood, rather than a less disadvantaged neighborhood, does not meaningfully alter the effects of attending a higher versus lower quality school during 1st grade on later student achievement.

These findings, moreover, are consistent across all of our different measures for school quality. Whether operationalized in terms of effectiveness, resources, or disorder, estimates from the ECLS-K provide little evidence that school quality mediates or interacts with the effects of concentrated poverty on academic achievement. For example, estimates of the $RNIE$ indicate that an intervention shifting the distribution of school resources from that in less disadvantaged neighborhoods to that in more disadvantaged neighborhoods would reduce reading test scores by only a tiny fraction of a standard deviation, while estimates of the $RINT_{\text{ref}}$ and $RINT_{\text{med}}$ suggest that any interaction effects between school resources and neighborhood disadvantage are also close to zero. Results based on our measure of school disorder are nearly identical.

All of the estimates discussed previously isolate the explanatory role of school quality from that of school composition. In the lower panels of Tables 4 and 5, we assess whether school quality and school composition jointly explain the effects of neighborhood context on academic achievement. In general, our results provide only weak evidence that neighborhood effects are mediated by the quality and composition of elementary schools when considered together. For example, estimates of the $RNDE^{\dagger}$ indicate that residence in a more disadvantaged neighborhood at the 80th percentile of the treatment distribution, rather than a less disadvantaged neighborhood at the 20th percentile, would reduce 3rd grade math scores by about 0.12 standard deviations, even after an intervention that would expose children to higher quality schools with lower levels of free lunch participation and with a more representative proportion of nonwhite students. This effect is both highly significant and similar to the total effects reported previously. In contrast, estimates of the $RNIE^{\dagger}$, which capture an indirect effect of neighborhood context that operates through both school quality and school composition, are small in substantive terms and fail to consistently reach stringent thresholds for statistical significance.

To help illuminate *why* differences in elementary schools do not appear to explain the effects of neighborhood disadvantage on academic achievement, Tables 6 through 8 present selected parameter estimates from our models of school quality, school composition, and academic achievement, which were used to construct the effect estimates outlined previously.

Table 6 presents estimates for the partial effect of neighborhood context from models of school quality and school composition during 1st grade. Recall that these models are linear functions of neighborhood disadvantage measured at the spring of kindergarten and a set of covariates, including baseline measures of achievement at both the child and school levels. The upper panel of the table presents estimates from models of school composition. They reveal a very strong relationship between the socioeconomic composition of neighborhoods and the elementary schools attended by resident children, as expected. According to these results, an increase of one standard deviation on the composite index of neighborhood disadvantage is estimated to increase subsequent exposure to low-income students at school (i.e., those who are eligible for a free lunch) by about 14 percentage points, or roughly one half of a standard deviation. This effect is substantively large and statistically significant at the most stringent thresholds. It is also comparable in magnitude to the partial effect of neighborhood disadvantage on exposure to nonwhite students at school, which is similarly pronounced.

The lower panel of Table 6 presents estimates for the partial effects of neighborhood context on our different measures of school quality. In contrast to effects on school composition, these results suggest a weaker and less consistent link between neighborhood disadvantage and the quality of elementary schools attended by resident children. Although there is some evidence that living in a disadvantaged neighborhood increases the likelihood that children attend schools with fewer resources and a more disorderly climate, these effects are muted in comparison to the relationship between neighborhood and school composition. Moreover, estimates from our models of school effectiveness indicate that living in a disadvantaged neighborhood does not impede access to elementary schools that provide considerable benefits with respect to student learning, their climate and resources notwithstanding. The absence of a strong link between neighborhood context and measures of school effectiveness, in particular, essentially precludes an important mediating role for the school environment in transmitting neighborhood effects on achievement test scores.

Tables 7 and 8 present selected parameter estimates from models of academic achievement at the end of 3rd grade, which are linear functions of school quality and composition during 1st grade, neighborhood disadvantage at the end of kindergarten, and a set of baseline covariates. Estimates from these models indicate that both neighborhood context and school effectiveness have substantively large and statistically significant effects – in the expected direction – on reading test scores. They also indicate, however, that these effects combine additively rather than multiplicatively, which precludes an explanatory role for school effectiveness arising from an interaction with neighborhood context. For math test scores, results are similar, except that school effectiveness has a smaller and statistically insignificant positive effect. This suggests that the benefits of attending a school that provides effective math instruction during the early years of a child’s primary education may fade out over time, whereas the harmful consequences of earlier exposure to a disadvantaged neighborhood are more lasting.

Estimates from our models of academic achievement also indicate that neither the resources, disorder level, nor composition of elementary schools have noteworthy effects on reading and math test scores measured at the end of 3rd grade, net of other factors. For all of these different characteristics of the school environment, their estimated partial effects on student achievement are close to zero, and they consistently fail to reach stringent thresholds for statistical significance.

Taken together, then, our results indicate that the characteristics of schools that are most closely linked with neighborhood context, like their composition and climate, are not that consequential for student achievement during the early elementary grades. Conversely, our estimates also indicate that those aspects of the school environment that are consequential for student achievement, such as factors contributing to effective reading instruction during 1st grade, are not that closely linked with neighborhood poverty. As a result, data from the ECLS-K provide little evidence that differences in elementary schools explain neighborhood effects on reading and math achievement.

8.4. Sensitivity Analysis

The validity of our inferences about the explanatory role of school quality depend on a number of strong assumptions about the absence of unobserved confounding. In particular, RWR estimates of direct and indirect effects are biased if there are any unobserved confounders of the exposure-outcome, mediator-outcome, or exposure-mediator relationships. In this section, we outline and implement a sensitivity analysis that examines whether our inferences are sensitive to hypothetical patterns of unobserved confounding (Wodtke and Zhou 2020). For simplicity, we prioritize results based on our measures of school effectiveness.

Consider the following set of linear structural equations for neighborhood disadvantage, school effectiveness, and achievement test scores:

$$A = \gamma_0 + \gamma_1(C - \alpha_0) + \varepsilon_A$$

$$M = \theta_0 + \theta_1(C - \alpha_0) + \theta_2 A + \varepsilon_M$$

$$Y = \lambda_0 + \lambda_1(C - \alpha_0) + \lambda_2 A + \lambda_3\big(Z - (\beta_0 + \beta_1 C + \beta_2 A)\big) + M(\lambda_4 + \lambda_5 A) + \varepsilon_Y,$$

where all variables are defined as previously. If there were no unobserved confounding of the exposure-outcome, mediator-outcome, or exposure-mediator relationships, then the error terms, $\{\varepsilon_A, \varepsilon_M, \varepsilon_Y\}$, would be pairwise independent.

If, however, the exposure-outcome relationship were confounded by unobserved factors, then ε_A and ε_Y would be correlated, and RWR estimates of the $RNDE$ would be biased. Specifically, if $\varepsilon_Y = \phi_{AY}\varepsilon_A + \psi_{AY}$ and $E(\psi_{AY}|C, A, Z, M) = 0$, the bias in estimates of the $RNDE$ due to unobserved exposure-outcome confounding is equal to

$$\text{Bias}_{AY}(RNDE) = \frac{\text{sd}(\psi_{AY})}{\text{sd}(\varepsilon_A)} \frac{\rho_{AY}}{\sqrt{1-\rho_{AY}^2}}(a^* - a),$$

where $\text{sd}(\varepsilon_A)$ can be estimated from a regression of A on C, $\text{sd}(\psi_{AY})$ can be estimated from our model for Y, and $\rho_{AY} = \text{corr}(\varepsilon_A, \varepsilon_Y)$ is the unknown correlation between errors. With this expression, we can construct and plot a set of bias-adjusted estimates for the $RNDE$ by evaluating the bias term across a range of values for ρ_{AY} and then by subtracting it from the RWR point and interval estimates.

Figure 5 plots bias-adjusted estimates of the $RNDE$ on reading and math test scores as a function of the error correlation, ρ_{AY}. A value of $\rho_{AY} = 0$ indicates that there is no unobserved exposure-outcome confounding and simply reproduces the estimates reported in Tables 4 and 5. A value of $\rho_{AY} < 0$ implies that families select into disadvantaged neighborhoods on the basis of unobserved factors that hinder the academic achievement of their children, net of observed covariates. These factors might include parental drug abuse or incarceration, for example. A value of $\rho_{AY} > 0$, by contrast, implies that families select into disadvantaged neighborhoods on the basis of unobserved factors that improve their children's academic achievement. We view this scenario as unlikely, and thus we only report bias-adjusted estimates for $\rho_{AY} \leq 0$.

Figure 5 indicates that bias-adjusted estimates of the $RNDE$ would reach zero under an error correlation of about -0.10. By way of reference, the partial correlation between parental years of schooling and reading test scores is about 0.10 in the ECLS-K, net of observed controls. Thus, the correlation between error terms in the exposure and outcome models would need to be comparable in absolute value to that between parental education and child achievement, conditional on measured covariates, in order to alter our conclusions about the $RNDE$.

Next, consider the scenario where the mediator-outcome relationship is confounded by unobserved factors. In this case, ε_M and ε_Y will be correlated, and RWR estimates of both the $RNDE$ and $RNIE$ will be biased. If $\varepsilon_Y = \phi_{MY}\varepsilon_M + \psi_{MY}$ and $E(\psi_{MY}|C, A, Z, M) = 0$, the bias in estimates of the $RNDE$ due to unobserved mediator-outcome confounding is equal to

$$\text{Bias}_{MY}(RNDE) = -\theta_2 \frac{\text{sd}(\psi_{MY})}{\text{sd}(\varepsilon_M)} \frac{\rho_{MY}}{\sqrt{1-\rho_{MY}^2}} (a^* - a),$$

and the bias in estimates of the $RNIE$ is equal to

$$\text{Bias}_{MY}(RNIE) = \theta_2 \frac{\text{sd}(\psi_{MY})}{\text{sd}(\varepsilon_M)} \frac{\rho_{MY}}{\sqrt{1-\rho_{MY}^2}} (a^* - a),$$

where $\text{sd}(\varepsilon_M)$ and θ_2 can be estimated from our model for school effectiveness, $\text{sd}(\psi_{MY})$ can be estimated from our model for test scores, and $\rho_{MY} = \text{corr}(\varepsilon_M, \varepsilon_Y)$ is the unknown error correlation. As before, we use these expressions to construct and plot a set of bias-adjusted estimates across a range of values for ρ_{MY} that allow us to assess whether our inferences about the $RNDE$ and $RNIE$ are sensitive to unobserved mediator-outcome confounding.

Figure 6 plots bias-adjusted estimates of the $RNDE$ and $RNIE$ as a function of the error correlation, ρ_{MY}. A value of $\rho_{MY} = 0$ indicates no unobserved mediator-outcome confounding and simply reproduces the estimates discussed previously. A value of $\rho_{MY} > 0$ implies that, net of observed covariates, families select into more effective schools on the basis of unobserved factors that improve the academic achievement of their children. Such factors might include parental valuations of academic learning and the behaviors that stem from these values, for example. A value of $\rho_{AY} < 0$, by contrast, implies that families select into more effective schools on the basis of unobserved factors that actually hinder their children's academic achievement. This might occur if deficient parents recognize their limitations and consequently seek out better schools for their children in order to compensate for their own personal shortcomings. We view both scenarios as at least minimally plausible, and thus we report bias-adjusted estimates for both positive and negative values of ρ_{MY}.

The upper panel of Figure 6 displays bias-adjusted estimates of the $RNDE$. For both reading and math achievement, these estimates are highly robust. Under any error correlation from -0.3 to 0.3, the bias-adjusted estimates indicate that exposure to a disadvantaged neighborhood has a negative direct effect that is substantively large and statistically significant at conventional thresholds. The lower panel of Figure 6 displays bias-adjusted estimates of the $RNIE$. These estimates only achieve statistical significance at extreme values of the error correlation, and even then they remain substantively small in magnitude. This suggests that our main conclusions about the explanatory role of school quality are robust to unobserved mediator-outcome confounding.

Finally, consider the scenario where the exposure-mediator relationship is confounded by unobserved factors. In this case, ε_A and ε_M will be correlated, and RWR estimates of both the

$RNDE$ and $RNIE$ will again be biased. Specifically, if $\varepsilon_M = \phi_{AM}\varepsilon_A + \psi_{AM}$ and $E(\psi_{AM}|C, A) = 0$, the bias in estimates of the $RNDE$ due to unobserved exposure-mediator confounding is equal to

$$\text{Bias}_{AM}(RNDE) = \frac{\text{sd}(\psi_{AM})}{\text{sd}(\varepsilon_A)} \frac{\rho_{AM}}{\sqrt{1-\rho_{AM}^2}} \lambda_5(a - \gamma_0)(a^* - a),$$

and the bias in estimates of the $RNIE$ is equal to

$$\text{Bias}_{AM}(RNIE) = \frac{\text{sd}(\psi_{AM})}{\text{sd}(\varepsilon_A)} \frac{\rho_{AM}}{\sqrt{1-\rho_{AM}^2}} (\lambda_4 + \lambda_5 a^*)(a^* - a),$$

where $\text{sd}(\varepsilon_A)$ can be estimated from a regression of A on C, $\text{sd}(\psi_{AM})$ can be estimated from our model for M, $\{\lambda_4, \lambda_5\}$ can be estimated by RWR applied to our model for Y, and $\rho_{AM} = \text{corr}(\varepsilon_A, \varepsilon_M)$ is the unknown error correlation. With these expressions, we examine whether our inferences about the $RNDE$ and $RNIE$ are sensitive to unobserved exposure-mediator confounding by constructing a set of bias-adjusted estimates under different values of ρ_{AM}.

Figures 7 plots bias-adjusted estimates of the $RNDE$ and $RNIE$, respectively, as a function of the error correlation, ρ_{AM}. A value of $\rho_{AM} = 0$ indicates no unobserved exposure-mediator confounding. Values of $\rho_{AM} < 0$ imply that families select into disadvantaged neighborhoods on the basis of unobserved factors that lead their children to attend less effective schools. And values of $\rho_{AY} > 0$ imply that families select into disadvantaged neighborhoods on the basis of unobserved factors that promote attendance at more effective schools, net of observed covariates. We view the last of these scenarios as implausible and therefore report bias-adjusted estimates only for values of $\rho_{MY} \leq 0$.

The upper panel of Figure 7 displays bias-adjusted estimates of the $RNDE$. These results suggest that our direct effect estimates are highly robust to unobserved exposure-mediator confounding. Specifically, they indicate that exposure to a disadvantaged neighborhood has a negative direct effect that is substantively large and statistically significant across the full range of error correlations considered here. The lower panel of Figure 7 displays bias-adjusted estimates of the $RNIE$. These estimates become positive and statistically significant at moderate values of the error correlation. This suggests that, if anything, differences in school effectiveness across neighborhoods may mitigate the harmful effects of living in a disadvantaged neighborhood. Nevertheless, the bias-adjusted estimates remain substantively small at all but extreme levels of the error correlation.

In sum, our inferences about neighborhood effects on academic achievement and the explanatory role of school quality remain largely unchanged under many different forms of confounding by unobserved factors.

8.5. Additional Robustness Checks

In addition to unobserved confounding, the validity of our inferences also depend on a number of strong assumptions about correct model specification and accurate measurement. First, if Equations 1 or 2 are incorrectly specified, then the effect estimates discussed previously may be biased. In Part E of the Online Supplement, we present results from an ancillary analysis in which we experiment with several more flexible specifications, including models that permit the effects of the exposure and mediator to vary across race, gender, parental education, and the rural-to-urban continuum. Effect estimates computed from these less restrictive specifications are very similar to those presented in Tables 4 and 5.

Second, faulty inferences may also arise if school quality has been inaccurately measured. With our approach to measuring this construct in terms of a school's effectiveness, systematic errors might arise because non-school determinants of achievement are less influential during the school year than during the summer or because the influence of schools on summer learning is nonzero. In Part F of the Online Supplement, we investigate whether our results are robust to this form of systematic measurement error by replicating our analysis using multiple alternative measures of school effectiveness. Specifically, we consider alternative measures that equate a school's effectiveness with (i) its 1st grade learning rate alone, which assumes non-school influences on students during the school year are absent entirely; (ii) the difference between its school-year learning rate and one half of the learning rate among its students during the summer, which assumes that non-school factors are half as influential when school is in session; (iii) the simple average of the 1st grade learning rate and the summer learning rate, which assumes that schools have very strong spillover effects on summer learning; (iv) a weighted average of the 1st grade learning rate and the summer learning rate, which assumes that schools have weaker but still nontrivial spillover effects during the summer; and (v) estimates of year-to-year value added from a conventional lagged dependent variable model, where test scores at the spring of 1st grade are modeled as a function of those from kindergarten and set of school random effects.

In addition, we replicate our analysis using alternative measures of school effectiveness constructed with data from the Early Childhood Longitudinal Study - Kindergarten Class of 2010, which included an additional wave of fall assessments and thus allows for computing school-year and summer learning rates through the end of 2nd grade. Using these data, we construct the two measures of school effectiveness that we view as most defensible, and then we average them over consecutive years to improve their reliability. Specifically, we measure a school's effectiveness as (vi) the difference between its school-year and summer learning rates, averaged over 1st and 2nd grade, and (vii) the difference between its school-year learning rate and one half of the learning rate observed during the previous summer, averaged over 1st and 2nd grade. Results based on all seven of these alternative measures are substantively similar to those discussed previously.

Finally, moving beyond analyses that consider each of our focal mediators separately, we also estimate direct effects of neighborhood context that jointly control for school effectiveness, resources, and disorder all in the same model. Results from this ancillary analysis indicate that the direct effects of neighborhood context remain substantively large, statistically significant, and comparable to the total effects discussed previously. They provide little evidence that our different measures of school quality, whether considered separately or jointly, explain the effects of neighborhood context on academic achievement.

9. Discussion

It is commonly hypothesized that neighborhood effects on educational outcomes are explained by differences in the schools that children attend (e.g., Jencks and Mayer 1990; Johnson 2012; Sanbonmatsu et al. 2006), but few prior studies investigate the role of school quality in transmitting the effects of concentrated poverty. In this study, we examine whether differences in the quality of elementary schools mediate or interact with neighborhood effects on academic achievement using novel counterfactual methods and a more defensible measurement strategy that captures multiple dimensions of school quality.

We find that living in a disadvantaged neighborhood reduces academic achievement. At the same time, however, we find little evidence that neighborhood effects are mediated by or interact with school quality, regardless of how this construct is operationalized. Differences in the quality of elementary schools do not seem to play an important mediating role because they are not very closely linked with the socioeconomic composition of neighborhoods, contrary to widely held assumptions in the literature on concentrated poverty. School quality also does not appear to interact with neighborhood context, as attending a higher versus lower quality elementary school has similar effects whether children live in an advantaged versus disadvantaged neighborhood. Finally, we find little evidence that neighborhood effects are jointly explained by differences in both school quality and school composition. This is because school composition is not very closely linked with student achievement, net of other factors, even though it is strongly related to neighborhood context.

These findings are difficult to reconcile with institutional resource theory, which contends that school quality is an especially important mediator of neighborhood effects on academic achievement (Jencks and Mayer 1990; Johnson 2012). They are also difficult to reconcile with either compound disadvantage or relative deprivation theories, which variously contend that the effects of school quality on achievement are dampened or amplified by living in an advantaged versus disadvantaged neighborhood (Crosnoe 2009; Jencks and Mayer 1990; Wodtke et al. 2016).

Rather, our findings suggest that neighborhood effects on academic achievement are most likely explained by other factors that are not directly linked to the quality, or the composition, of elementary schools. They suggest that the characteristics of elementary schools most closely linked with neighborhood context, such as the demographic composition of students, are not that consequential for student achievement, whereas those aspects of the school environment that are most consequential for student achievement, such as instructional effectiveness, are not that closely linked with neighborhood context.

These findings align with an emerging body of evidence suggesting that schools may serve as neutral or perhaps even equalizing institutions within the process by which academic disparities are generated and maintained (Downey 2020; Downey and Condron 2016; Raudenbush and Eschmann 2015). For example, they are consistent with prior research showing that socioeconomic gaps in achievement are largest before school even begins and then decline thereafter (von Hippel et al. 2018), with prior research documenting relatively weak associations between many different school-level characteristics and student test scores (Coleman et al. 1966; Lauen and Gaddis 2013), with research suggesting that both school inputs and school outputs do not vary as widely across neighborhood contexts as is often assumed (Downey et al. 2008, 2019; Owens and Candipan 2019), and with previous studies documenting a relatively weak link between measures of school composition and measures of school value added (Angrist et al. 2017; Deming 2014; Downey et al. 2019; Hanselman and Fiel 2017).

Our analysis extends prior research on the joint effects of neighborhoods and schools (e.g., Wodtke and Parbst 2017) by incorporating defensible measures of school quality and appropriately isolating their explanatory role from that of school composition with novel counterfactual methods. Through this approach, we reveal a complex set of relationships connecting neighborhood poverty, the school environment, and student outcomes that challenge existing theories of contextual effects and educational inequality in new ways: concentrated poverty is closely linked with the composition of schools, but not with their quality, however defined, and for this reason, its neighborhood effects on student achievement seem to operate independently of the school environment.

The apparently weak link between the socioeconomic composition of neighborhoods and the quality of elementary schools attended by resident children is counterintuitive and warrants scrutiny. Parents invest in their children at least in part by seeking out neighborhoods and schools that are anticipated to improve their educational outcomes (Hoxby 2003; Lareau 2003; Owens 2016). Because advantaged families have more to invest, a strong link between school quality and neighborhood composition should emerge endogenously via these family sorting processes (Durlauf 1996).

We conjecture that the close connection between neighborhood context and school quality predicted by most theoretical models of place-based inequality is not observed empirically in part because parents make decisions about how to invest in their children's education based on a highly diverse set of preferences and a highly imperfect information set. In particular, parents often lack access to accurate information about school quality, and they may select schools for their children on the basis of characteristics with relatively little impact on math and reading achievement (Abdulkadiroglu et al. 2014; Angrist et al. 2017; Deming 2014). It is also possible that school contributions to different aspects of child development are not highly correlated. In this situation, even if parents do make decisions with accurate information on school quality, they may prioritize characteristics of schools that contribute to non-academic dimensions of their children's development (Beuermann et al. 2020), thereby weakening the link between neighborhood and school contributions to academic skills. For example, parents of different class backgrounds may select schools based on the degree to which they impart cultural capital, respect for rules versus creative independence, or socioemotional skills, all of which can have a strong influence on later status attainment even though they may not be as closely linked with performance on achievement tests (Anyon 1980; Bourdieu and Passeron 1977; Heckman et al. 2014). A third possibility is that some parents may prioritize the consumption value of schooling over its investment value, and that these two different types of benefits are also not very closely associated, which would similarly attenuate the link between neighborhood and school effects on academic skills.

In addition to limited information and diverse preferences, broader structural constraints on the present organization of elementary education may restrict variation across primary schools and weaken the widely hypothesized connection between parental resources and the degree to which the schools that families select for their young children actually improve academic outcomes. Beyond the bureaucratic standardization of funding, staffing, and physical plant requirements (Guppy and Davies 2006), we speculate that the relatively uniform curricular focus of early elementary grades is important for constraining variation in school effectiveness across neighborhoods. Because kindergarten and 1st grade typically focus on a fairly standard set of academic skills, variation among instructional regimes at this level is likely more limited, and the scope for student selection into a stratified curriculum more constrained, than is the case later on in secondary school. These constraints may also weaken the link between neighborhood composition and school quality at the elementary level.

But even if elementary schools are not to blame for neighborhood-based disparities in academic achievement, they can still be part of the solution. Many studies show how different types of school reforms can dramatically improve performance among disadvantaged students and narrow achievement gaps (e.g., Chenoweth 2009; Hassrick et al. 2017). Caution is needed, however, when singling out schools serving poor communities for criticism, overhaul, and sometimes even outright closure, as often occurs in public discourse on school reform. Our results suggest that the elementary schools serving children from poor communities are, on average, educating their 1st grade students as effectively as the schools serving advantaged communities, even though they have somewhat fewer resources and a more disruptive climate.

Consequently, closing or overhauling schools in poor neighborhoods may not be the best means for mitigating neighborhood effects on academic achievement. Many of these schools are valuable community resources that have noteworthy positive impacts on their students, despite public stereotypes to the contrary. Our results suggest that policies focused on renewed investment in these schools, as opposed to overhauling or closing them, may be more effective at improving educational outcomes in poor neighborhoods. These conclusions also resonate with recent evidence showing that school closures stemming from the COVID-19 pandemic have had disparate consequences by social class, where students from low-income communities have suffered the largest setbacks to their learning after the abrupt transition away from conventional forms of instruction (Goldstein 2020).

An important methodological implication of this study is that the link between neighborhood context and school quality is highly sensitive to the choice of metric used to evaluate schools. In sharp contrast to our findings from the ECLS-K, prior studies that rely on proxy measures with poor construct validity, such as the demographic composition or average ability levels of students (e.g., Dobbie and Fryer 2011; Wodtke and Parbst 2017), indicate that poor neighborhoods are overwhelmingly served by "low-quality" schools. The discrepancy between these results and ours underscores the importance of operationalizing school quality in a defensible manner that more closely corresponds with the benefits that schools provide to students.

This study also introduced novel methods for decomposing effects into components due to mediation versus interaction, for estimating these components in the presence of exposure-induced confounding, and for assessing the sensitivity of estimates to the presence of unobserved confounding. Social scientists have become increasingly interested not only in establishing the existence of causal effects but also in explaining how they arise (e.g., Hallsten and Pfeffer 2017; Schneider and Harknett 2019). The decomposition outlined in the present study should therefore find wide application, wherever there is interest in understanding the process by which a cause produces its effects. Similarly, exposure-induced confounding is ubiquitous in the social sciences (VanderWeele 2015), as causal effects are typically transmitted through a confluence of interrelated mechanisms. And it is difficult to properly isolate these different mechanisms using experimental research designs that rely on random assignment to achieve identification (Imai, Tingley, and Yamamoto 2013). Thus, the method of RWR, especially when paired with a formal sensitivity analysis, is also widely relevant.

For example, these methods could be used to disentangle the effects of neighborhood poverty from other characteristics of the residential area, besides the school environment, that are thought to mediate them. Like school composition, neighborhood composition may influence student outcomes through multiple different pathways, including many different types of neighborhood resources (e.g., community associations, childcare centers, and public amenities) or indicators of neighborhood disorder (e.g., crime, infrastructure in disrepair, and a lack of interpersonal trust). Our framework for separating the effects of population composition from other dimensions of some focal context could be usefully applied to different features of the neighborhood environment.

Although this study has important implications for theory, policy, and methods, it is not without limitations. The first is that our measures of school quality, despite their many advantages, do not capture every aspect of the school environment that might affect student achievement or differ across more versus less disadvantaged neighborhoods. For example, our measure of school resources may not capture "hidden funding gaps" between schools within the same district (Hall and Ushomirsky 2010). Output-based measures like school effectiveness should, in theory, absorb the influence of all unobserved mediating factors at the school level. But because output-based measures are, in practice, typically based on a non-exhaustive subset of outputs, they may also fail to capture some important pathways. It remains possible, then, that unmeasured dimensions of the school environment, such as teacher quality, curricular resources, or in-school violence, play an important role in transmitting neighborhood effects on other dimensions of academic achievement.

A second limitation is our focus on achievement test scores both for evaluating a school's quality via its effectiveness and for measuring student outcomes. School effectiveness is itself a multidimensional construct that involves more than just academic skills, and some of these dimensions may be more or less closely linked with neighborhood context and student outcomes. For example, elementary schools may differ in the degree to which they impart so-called "non-cognitive skills," such as conscientiousness, perseverance, and sociability, and these skills may be especially important for successfully navigating crucial academic transitions, like graduating from high school (Heckman et al. 2014). By focusing only on achievement test scores, our study may obscure the role of elementary schools in explaining neighborhood effects on other outcomes. An important direction for future research will be to examine a broader set of student outputs both for measuring school effectiveness and for evaluating student success.

A third limitation of this study is our focus on point-in-time effects during elementary school, when it's possible, or perhaps even likely, that the causal processes of interest may become more pronounced when exposures are measured over a longer time horizon or at later developmental stages. Our analysis is designed to emulate a hypothetical intervention that would change a child's neighborhood environment at the end of kindergarten, and their school environment at the beginning of 1^{st} grade, while leaving all factors prior to the start of kindergarten undisturbed. It does not capture the influence of a child's residential context from birth to school entry, when socioeconomic gaps in achievement are initially formed. Nor does it capture the effects of changes to a child's educational environment later during middle or high school, when differences across instructional regimes become more pronounced. It remains possible, then, that school quality is more important for explaining neighborhood effects on educational outcomes during adolescence. Future research should therefore examine the cumulative or time-dependent effects of neighborhood and school environments over the entire early life course. This may demand new and more complex forms of data collection, as few existing studies track the neighborhoods, schools, and development of children from birth through late adolescence.

Finally, this study is limited by its reliance on data collected between 1998 and 2007, given that many districts across the U.S. have recently undergone changes that affect the schooling options available to residents. The recent and rapid expansion of charter schools and intra-district open enrollment policies may have altered the relationship between neighborhood composition and school quality among contemporary cohorts of students. We focused on data from the ECLS-K class of 1998 because it allows for the longest possible follow-up period – through the end of 8th grade – and thus for an assessment of whether the effects of early contextual exposures fade out over time. Although we conceptually replicated our key findings with more recent data from the ECLS-K class of 2010 (see Part F of the Online Supplement), the influence of school choice expansion on the causal processes of interest awaits a rigorous empirical assessment.

These limitations notwithstanding, our results provide considerable evidence that children growing up in disadvantaged neighborhoods perform worse academically than they would have living elsewhere not because of differences in the quality of their elementary schools but rather because of other unmeasured causal mechanisms. This suggests that unpacking the "black box" through which neighborhood effects are transmitted during childhood will likely require a renewed focus on alternative social processes, including exposure to crime and violence, environmental health hazards, and differences in peer subcultures, among a variety of other possibilities.

References

Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82:137-196.

Abdulkadiroglu, Atila, Parag Pathak, Jonathan Schellenberg, Christopher R. Walters. 2020. "Do Parents Value School Effectiveness?" *American Economic Review* 110:1502-39.

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

Alwin, Duane F. and Robert M. Hauser. 1975. "The Decomposition of Effects in Path Analysis." *American Sociological Review* 40:37-47.

Anderson, Elijah. 1999. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. New York: Norton.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, Christopher R. Walters. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." *The Quarterly Journal of Economics* 132:871-919.

Anyon, Jean. 1980. "Social Class and the Hidden Curriculum of Work." *Journal of Education* 162:67-92.

Arum, Richard. 2000. "Schools and Communities: Ecological and Institutional Dimensions." *Annual Review of Sociology* 26:395-418.

Arum, Richard. 2005. *Judging School Discipline: The Crisis of Moral Authority*. Cambridge, MA: Harvard University Press.

Auld, M. Christopher and Nirmal Sidhu. 2005. "Schooling, Cognitive Ability and Health." *Health Economics* 14:1019-34.

Bailey, Drew, Greg J. Duncan, Candice L. Odgers, and Winnie Yu. 2015. "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions." *Journal of Research on Educational Effectiveness* 10:7-39.

Baron, Reuben M. and David A. Kenny. 1986. "The Moderator-mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51:1173-1182.

Beuermann, Diether, C. Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo. 2020. "What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output." NBER Working Paper 25342.

Billingham, Chase M. and Matthew O. Hunt. 2016. "School Racial Composition and Parental Choice." *Sociology of Education* 89:99-117.

Bohrnstedt, George W. and Gerald Marwell. 1978. "The Reliability of Products of Two Random Variables." *Sociological Methodology* 9:254-273.

Borman, Geoffrey D. and N. Maritza Dowling. 2008. "Teacher Attrition and Retention: A Meta-analytic and Narrative Review of the Research." *Review of Educational Research* 78:367-409.

Bourdieu, Pierre and Jean-Claude Passeron. 1977. *Reproduction in Education, Society and Culture*. London: Sage.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, Matthew Ronfeldt, and James Wykoff. 2011. "The Effect of School Neighborhoods on Teacher Retention Decisions." Pp. 377-396 in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by G. Duncan and R. Murnane. New York: Russell Sage.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2005. "Explaining the Short Careers of High-achieving Teachers in Schools with Low-performing Students." *The American Economic Review* 95:166-171.

Caldas, Stephen J. 1993. "Reexamination of Input and Process Factor Effects on Public School Achievement." *Journal of Educational Research* 86:206-214.

Caldwell, Bettye M. and Robert H. Bradley. 1984. *Home Observation for Measurement of the Environment (Home) - Revised Edition*. Little Rock, AR: University of Arkansas.

Card, David and Jesse Rothstein. 2007. "Racial Segregation and the Black-White Test Score Gap." *Journal of Public Economics* 91:2158-2184.

Carlson, Deven and Josua M. Cowen. 2014. "Student Neighborhoods, Schools, and Test Score Growth: Evidence from Milwaukee, Wisconsin." *Sociology of Education* 88:38-55.

Chenoweth, Karin. 2009. *How It's Being Done: Urgent Lessons from Unexpected Schools.* Cambridge, MA: Harvard Education Press.

Chetty, Raj, and John N. Friedman, and Jonah E. Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-added Estimates." *American Economic Review* 104:2593-2632.

Chetty, Raj, and John N. Friedman, and Jonah E. Rockoff. 2014b. "Measuring the Impacts of Teachers II: Teacher Value-added and Student Outcomes in Adulthood." *American Economic Review* 104:2633-2679.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106:855-902.

Clampet-Lundquist, Susan and Douglas S. Massey. 2008. "Neighborhood Effects of Economic Self-sufficiency: A Reconsideration of the Moving to Opportunity Experiment." *American Journal of Sociology* 114:107-143.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Cook, Thomas D., Melissa R. Herman, Meredith Phillips and Richard A. Settersten. 2002. "Some Ways in Which Neighborhoods, Nuclear Families, Friendship Groups, and Schools Jointly Affect Changes in Early Adolescent Development." *Child Development* 73:1283-1309.

Corcoran, Sean, William N. Evans, Jennifer Godwin, Sheila E. Murray, and Robert M. Schwab. 2004. “The Changining Distribution of Education Finance: 1972-1997.” Pp. 433-465 in *Social Inequality*. New York: Russell Sage Foundation.

Crosnoe, Robert. 2009. “Low-Income Students and the Socioeconomic Composition of Public High Schools.” *American Sociological Review* 74:709-730.

Crowder, Kyle and Liam Downey. 2010. “Interneighborhood Migration, Race, and Environmental Hazards: Modeling Microlevel Processes of Environmental Inequality.” *American Journal of Sociology* 115:1110-1149.

Cruse, Craig and David Powers. 2006. “Estimating School District Poverty with Free and Reduced-price Lunch Data.” U.S. Census Bureau-Small Area Estimates Branch Research Report. http://www.census.gov/hhes/www/saipe/asapaper/CrusePowers2006asa.pdf.

Davis, James A. 1966. “The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men.” *American Journal of Sociology* 72:17-31.

Davies, Scott and Neil Guppy. 2010. *The Schooled Society*. New York: Oxford University Press.

DeLuca, Stefanie and Peter Rosenblatt. 2010. “Does Moving to Better Neighborhoods Lead to Better Schooling Opportunities?” *Teachers College Record* 112:1143-1491.

Deming, David J. 2014. “Using School Choice Lotteries to Test Measures of School Effectiveness.” *American Economic Review* 104:406-411.

Devine, John. 1996. *Maximum Security: The Culture of Violence in Inner-city Schools*. Chicago, IL: University of Chicago Pres

Dobbie, Will and Roland G. Fryer Jr. 2011. “Are High-quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children’s Zone.” *American Economic Journal: Applied Economics* 3:158-187.

Dobbie, Will and Roland G. Fryer. 2014. “The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools.” *American Economic Journal: Applied Economics* 6:58-75.

Downey, Douglas B. 2020. *How Schools Really Matter: How We Think About Schools and Inequality and Why It’s Mostly Wrong*. Chicago: University of Chicago Press.

Downey, Douglas B. and Dennis J. Condron. 2016. “Fifty Years Since the Coleman Report: Rethinking the Relationship Between Schools and Inequality.” *Sociology of Education* 89:207-220.

Downey, Douglas B., David M. Quinn, and Melissa Alcaraz. 2019. “The Distribution of School Quality: Do Schools Serving Mostly White and High-SES Children Produce the Most Learning?” *Sociology of Education* 92:386-403.

Downey, Douglas B., Paul T. von Hippel, and Melanie Hughes. 2008. “Are ‘Failing’ Schools Really Failing? Using Seasonal Comparison to Evaluate School Effectiveness.” *Sociology of Education* 81:242-270.

Draper, Norman R. and Harry Smith. 1998. *Applied Regression Analysis*. New York: Wiley.

Duffett, Ann, Steve Farkas, and Tom Loveless. 2008. *High-achieving Students in the Era of No Child Left Behind*. Washington, DC: Thomas B. Fordham Institute.

Durluaf, Steven N. 1996. “A Theory of Persistent Income Inequality.” *Journal of Economic Growth* 1:75-93.

Elwert, Felix and Christopher Winship. 2014. “Endogenous Selection Bias: The Problem of

Conditioning on a Collider Variable." *Annual Review of Sociology* 40:31-53.

Esposito, Cynthia. 1999. "Learning in Urban Blight: School Climate and its Effect on the School Performance of Urban, Minority, Low-income Children." *School Psychology Review* 28:365-377.

Ferryman, Kadija S., Xavier de Souza Briggs, Susan J. Popkin and Maria Rendon. 2008. "Do Better Neighborhoods for MTO Families Mean Better Schools?" *Metropolitan and Housing Communities* 3:1-12.

Figlio, David N. 2007. "Boys Named Sue: Disruptive Children and their Peers." *Education Finance and Policy* 2:376-394.

Galster, George C. 2012. "The Mechanism(s) of Neighbourhood Effects: Theory, Evidence, and Policy Implications." Pp. 23-56 in *Neighbourhood Effects Research: New Perspectives*, edited by M. van Ham, D. Manley, N. Bailey, L. Simpson and D. Maclennan. New York: Springer.

Goldsmith, Pat R. 2009. "Schools or Neighborhoods or Both? Race and Ethnic Segregation and Educational Attainment." *Social Forces* 87:1913-1942.

Goldstein, Dana. 2020. "Research Shows Students Falling Months Behind During Virus Disruptions." *The New York Times.* Retrived from https://www.nytimes.com/2020/06/05/us/coronavirus-education-lost-learning.html.

Gottfried, Michael A. 2019. "Chronic Absenteeism in the Classroom Context: Effects on Achievement." *Urban Education* 54:3-34.

Greenman, Emily, Katerina Bodovski, and Katherine Reed. 2011. "Neighborhood Characteristics, Parental Practices, and Children's Math Achievement in Elementary School." *Social Science Research* 40:1434-1444.

Hall, Daria and Natasha Ushomirsky. 2010. "Close the Hideen Funding Gaps in Our Schools." Washington, DC: The Education Trust. Retrieved from https://files.eric.ed.gov/fulltext/ED511871.pdf.

Hallsten, Martin and Fabian T. Pfeffer. 2017. "Grand Advantage: Family Wealth and Grandchildren's Educational Acheivement in Sweden." *American Sociological Review* 82:328-360.

Hanselman, Paul, and Jeremy J. Fiel. 2016. "School Opportunity Hoarding? Racial Segregation and Access to High Growth Schools." *Social Forces* 95:1077-1104.

Hanushek, Eric A. 2003. "The Failure of Input-based Schooling Policies." *The Economic Journal* 113:F64-F98.

Hanushek, Eric A. 2006. "School Resources." Pp. 865-908 in *Handbook of the Economics of Education (Vol. 2)* edited by E. A. Hanushek and Finis Welch. Amsterdam: North Holland.

Hassrick, Elizabeth M., Stephen W. Raudenbush, and Lisa Rosen. 2017. *The Ambitious Elementary School: Its Conception, Design, and Implications for Educational Equality.* Chicago: University of Chicago Press.

Harding, David J. 2003. "Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping out and Teenage Pregnancy." *American Journal of Sociology* 109:676-719.

Harding, David J. 2009. "Violence, Older Peers, and the Socialization of Adolescent Boys in Disadvantaged Neighborhoods." *American Sociological Review* 74:445-464.

Harding, David J., Lisa Gennetian, Christopher Winship, Lisa Sanbonmatsu, and Jeffrey R. Kling. 2011. "Unpacking Neighborhood Influences on Education Outcomes: Setting the

Stage for Future Research" Pp. 277-296 in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* edited by Greg J. Duncan and Richard Murnane. New York: Russell Sage.

Heuer, Ruth and Stephanie Stullich. 2011. "Comparability of State and Local Expenditures among Schools within Districts: A Report from the Study of School-Level Expenditures." Washington, DC: U.S. Department of Education.

Hoxby, Caroline. 2003. *The Economics of School Choice*. Chicago: University of Chicago Press.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. 2013. "Experimental Designs for Identifying Causal Mechanisms." *Journal of the Royal Statistical Society (Series A)* 176:5-51.

Isenberg, Eric, Jeffrey Max, Philip Gleason, Liz Potamites, Robert Santillano, Heinrich Hock, and Michael Hansen. 2013. *Access to Effective Teaching for Disadvantaged Students*. Washington, DC: Institute for Education Sciences.

Jackson Kirabo, C., Rucker C. Johnson, and Claudia Persico. 2016. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *The Quarterly Journal of Economics* 131:157-218.

Jencks, Christopher and Susan E. Mayer. 1990. "The Social Consequences of Growing up in a Poor Neighborhood." Pp. 111-186 in *Inner-City Poverty in the United States*, edited by L. E. Lynn and M. G. H. McGreary. Washington, DC: National Academy Press.

Johnson, Odis Jr. 2012. "A Systematic Review of Neighborhood and Institutional Relationships Related to Education." *Education and Urban Society* 44:477-511.

Kahlenberg, Richard D. 2001. *All Together Now: Creating Middle Class Schools through Public School Choice*. Washington, DC: Brookings.

Koedel, Cory and Julian Betts. 2010. "Value Added to What? How a Ceiling in the Testing Instrument Influences Value-added Estimation." *Education Finance and Policy* 5:54-81.

Labaree, David F. 2012. *Someone Has to Fail.* Cambridge, MA: Harvard University Press.

Ladd, Helen and Susanna Loeb. 2013. "The Challenges of Measuring School Quality: Implications for Educational Equity." Pp. 22-55 in *Education, Justice, and Democracy*, edited by D. Allen and R. Reich. Chicago: University of Chicago Press.

Lauen, Douglas Lee and S. Michael Gaddis. 2013. "Exposure to Classroom Poverty and Test Score Achievement: Contextual Effects of Selection?" *American Journal of Sociology* 118:943-979.

Lareau, Annette. 2003. *Unequal Childhoods: Class, Race, and Family Life*. Berkeley, CA: Univiversity of California Press.

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

Murnane, Richard J. and Frank Levy. 2006. *Teaching the New Basic Skills: Principles for Educating Children to Thrive in a Changing Economy*. New York: Free Press.

Nolan, Kathleen. 2011. *Police in the Hallways: Discipline in an Urban High School.* Minneapolis, MN: University of Minnesota Press.

Orr, Larry, Judith D. Feins, Robin Jacob, Erik Beecroft, Lisa Sanbonmatsu, Lawrence F. Katz, Jeffrey B. Liebman, and Jeffrey R. Kling. 2003. *Moving to Opportunity for Fair Housing Demonstration Program: Interim Impacts Evaluation*. Washington, DC: 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* 83:287-311.

Owens, Ann. 2016. “Inequality in Children’s Contexts: Income Segregation of Households With and Without Children.” *American Sociological Review* 81: 549-574.

Owens, Ann and Jennifer Candipan. 2019. “Social and Spatial Inequalities of Educational Opportunity: A Portrait of Schools Serving High- and Low-income Neigborhoods in U.S. Metropolitan Areas.” *Urban Studies* 56:3178-3197.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference (2nd Edition)*. New York: Cambridge University Press.

Pollock, Judith M., Donald A. Rock, and Michael J. Weiss. 2005. *Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 Pyschometric Report for the Third Grade*. Washington, DC: National Center for Education Statistics.

Radloff, Lenore S. 1977. “The CES-D Scale: A Self-report Depression Scale for Research in the General Population.” *Applied Psychological Measurement* 1:385-401.

Raudenbush, Stephen W. and Robert D. Eschmann. 2015. “Does Schooling Increase or Reduce Social Inequality?” *Annual Review of Sociology* 41:443-470.

Raudenbush, Stephen W. 2004. *Schooling, Statistics, and Poverty: Can We Measure School Improvement?* Princeton, NJ: Educational Testing Service.

Rendon, Maria G. 2014. “Drop Out and ‘Disconnected’ Young Adults: Examining the Impact of Neighborhood and School Contexts.” *The Urban Review* 46:169-196.

Rosenbaum, James E. 1995. “Changing the Geography of Opporuntity by Expanding Residential Choice: Lessons from the Gautreaux Program.” *Housing Policy Debate* 6:231-269.

Rosenfeld, Lindsay, Rima Rudd, Ginger L. Chew, Karen Emmons and Dolores Acevedo-Garcia. 2010. “Are Neighborhood-Level Characteristics Associated with Indoor Allergens in the Household?” *Journal of Asthma* 47:66-75.

Rothstein, Richard. 2004. *Class and Schools.* New York: Teachers College Press.

Rubin, Donald B. 1974. “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.” *Journal of Educational Psychology* 66:688-701.

Sampson, Robert J. 2001. “How Do Communities Undergird or Undermine Human Development? Relevant Contexts and Social Mechanisms.” Pp. 3-30 in *Does It Take a Village? Community Effects on Children, Adolescents, and Families*, edited by A. Booth and N. Crouter. Mahwah, N.J.: Erlbaum.

Sampson, Robert J. 2008. “Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure.” *American Journal of Sociology* 114:189-231.

Sampson, Robert J. 2012. *Great American City*. Chicago, IL: University of Chicago Press.

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* 41:649-691.

Sanbonmatsu, Lisa, Jens Ludwig, Lawrence Katz, Lisa A. Gennetian, Greg J. Duncan, Ronald C. Kessler, Emma Adam, Thomas W. McDade and Stacy T. Lindau. 2011. *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation*. Washington, DC: U.S. Department of Housing and Urban Development.

Saporito, Salvatore and Deenesh Sohoni. 2007. “Maping Educational Inequality: Concentrations of Poverty among Poor and Minority Students in Public Schools.” *Social Forces* 85:1227-1253.

Schneider, Daniel and Kristen Harknett. 2019. “Consequences of Routine Work-Schedule Instability for Worker Health and Well-Being.” *American Sociological Review* 84:82-114.

Sharkey, Patrick. 2010. “The Acute Effect of Local Homocides on Children's Cognitive Performance.” *Proceedings of the National Academy of Sciences* 107:11733-11738.

Sharkey, Patrick and Jacob W. Faber. 2014. “Where, When, Why, and For Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects.” *Annual Review of Sociology* 40:559-579.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. “What Are We Weighting For?” *Journal of Human Resources* 50:301-316.

Steinberg, Laurence. 1997. *Beyond the Classroom: Why School Reform Has Failed and What Parents Need to Do*. New York: Simon & Schuster.

VanderWeele, Tyler J. 2014. “A Unification of Mediation and Interaction.” *Epidemiology* 25:749-761.

VanderWeele, Tyler J. 2015. *Explanation in Causal Inference*. New York: Oxford University Press.

VanderWeele, Tyler J., Stijn Vansteelandt, and James M. Robins. 2014. “”Effect Decomposition in the Presence of an Exposure-induced Mediator-outcome Confounder.” *Epidemiology* 25:300-306.

von Hippel, Paul T. 2007. “Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data.” *Sociological Methodology* 37:83-117.

von Hippel, Paul T. 2009. “Achievement, Learning, and Seasonal Impact as Measures of School Effectiveness: It’s Better to be Valid than Reliable.” *School Effectiveness and School Improvement* 20:187-213.

von Hippel, Paul T. and Caitlin Hamrock. 2019. “Do Test Score Gaps Grow Before, During, or Between School Years? Measurement Artifacts and What We Can Know in Spite of Them.” *Sociological Science* 6:43-80.

von Hippel, Paul T., Joseph Workman, and Douglas B. Downey. 2018. “Inequality in Reading and Math Skills Forms Mainly before Kindergarten: A Replication, and Partial Correction, of ‘Are Schools the Great Equalizer?’” *Sociology of Education* 91:323-357.

Wang, Ming-Te and Jessica Degol. 2016. “School Climate: A Review of the Construct, Measurement, and Impact on Student Outcomes.”*Educational Pyschology Review* 28:315-352.

Willms, J. Douglas. 2010. “School Composition and Contextual Effects on Student Outcomes.” *Teachers College Record* 112:1008-1038.

Wilson, William J. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago: University of Chicago Press.

Winship, Christopher, and Larry Radbill. 1994. “Sampling Weights and Regression Analysis.” *Sociological Methods & Research* 23:230-257.

Wodtke, Geoffrey T. 2020. “Regression-based Adjustment for Time-varying Confounders.” *Sociological Methods and Research* 49:906-946.

Wodtke, Geoffrey T. and Daniel Almirall. 2017. “Estimating Moderated Causal Effects with Time-varying Treatments and Time-varying Moderators: Structural Nested Mean Models and Regression-with-residuals.” *Sociological Methodology* 47:212-245.

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* 76:713-736.

Wodtke, Geoffrey T., Felix Elwert, and David J. Harding. 2016. "Neighborhood Effect Heterogeneity by Family Income and Developmental Period." *American Journal of Sociology* 121:1168-1222.

Wodtke, Geoffrey T. and Matthew Parbst. 2017. "Neighborhoods, Schools, and Academic Achievement: A Formal Mediation Analysis of Contextual Effects on Reading and Mathematics Abilities." *Demography* 54:1653-1676.

Wodtke, Geoffrey T., Zahide Alaca, and Xiang Zhou. 2020. "Regression-with-residuals Estimation of Marginal Effects: A Method of Adjusting for Treatment-induced Confounders that are also Effect Modifiers." *Journal of the Royal Statistical Society: Series A* 183:311-332.

Wodtke, Geoffrey T. and Xiang Zhou. 2020. "Effect Decomposition in the Presence of Exposure-induced Confounding: A Regression-with-residuals Approach." *Epidemiology* 31:369-375.

Zhou, Xiang and Geoffrey T. Wodtke. 2019. "A Regression-with-residuals Method for Estimating Controlled Direct Effects." *Political Analysis* 27:360-369.

Figures

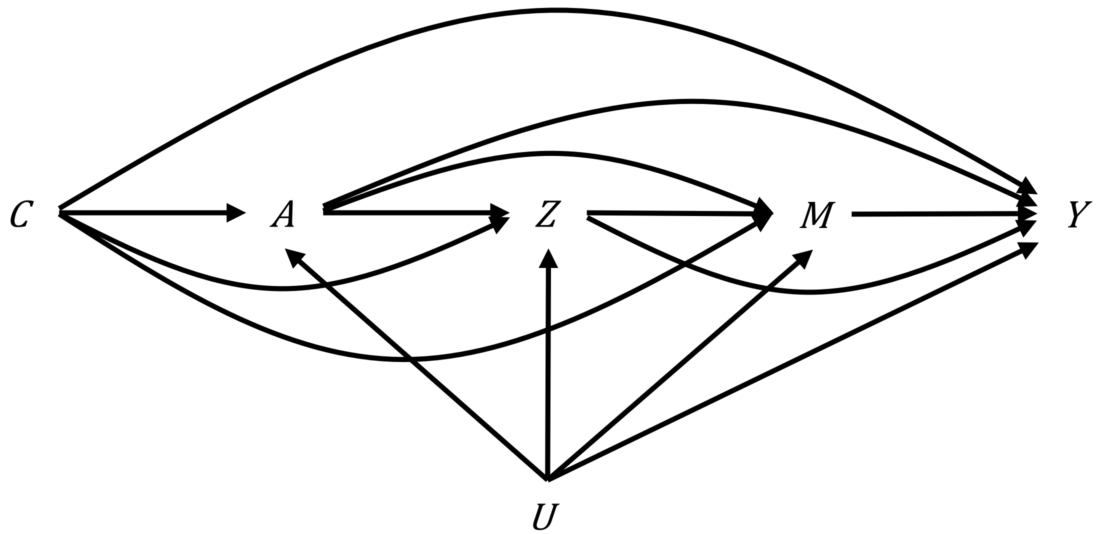

Figure 1. Hypothesized causal relationships between the baseline confounders (C), neighborhood context (A), school composition (Z), school quality (M), achievement test scores (Y), and unobserved factors (U).

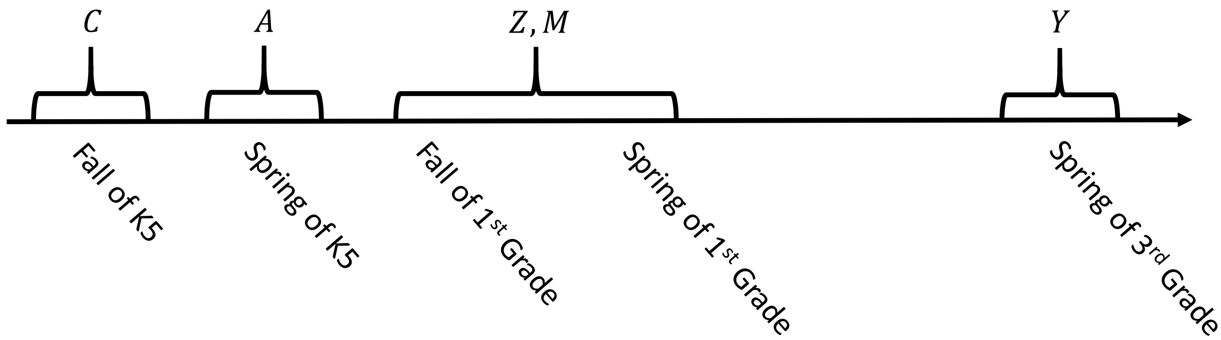

Figure 2. Longitudinal measurement strategy to ensure appropriate temporal ordering of baseline confounders (C), neighborhood context (A), school composition (Z), school quality (M), and achievement test scores (Y).

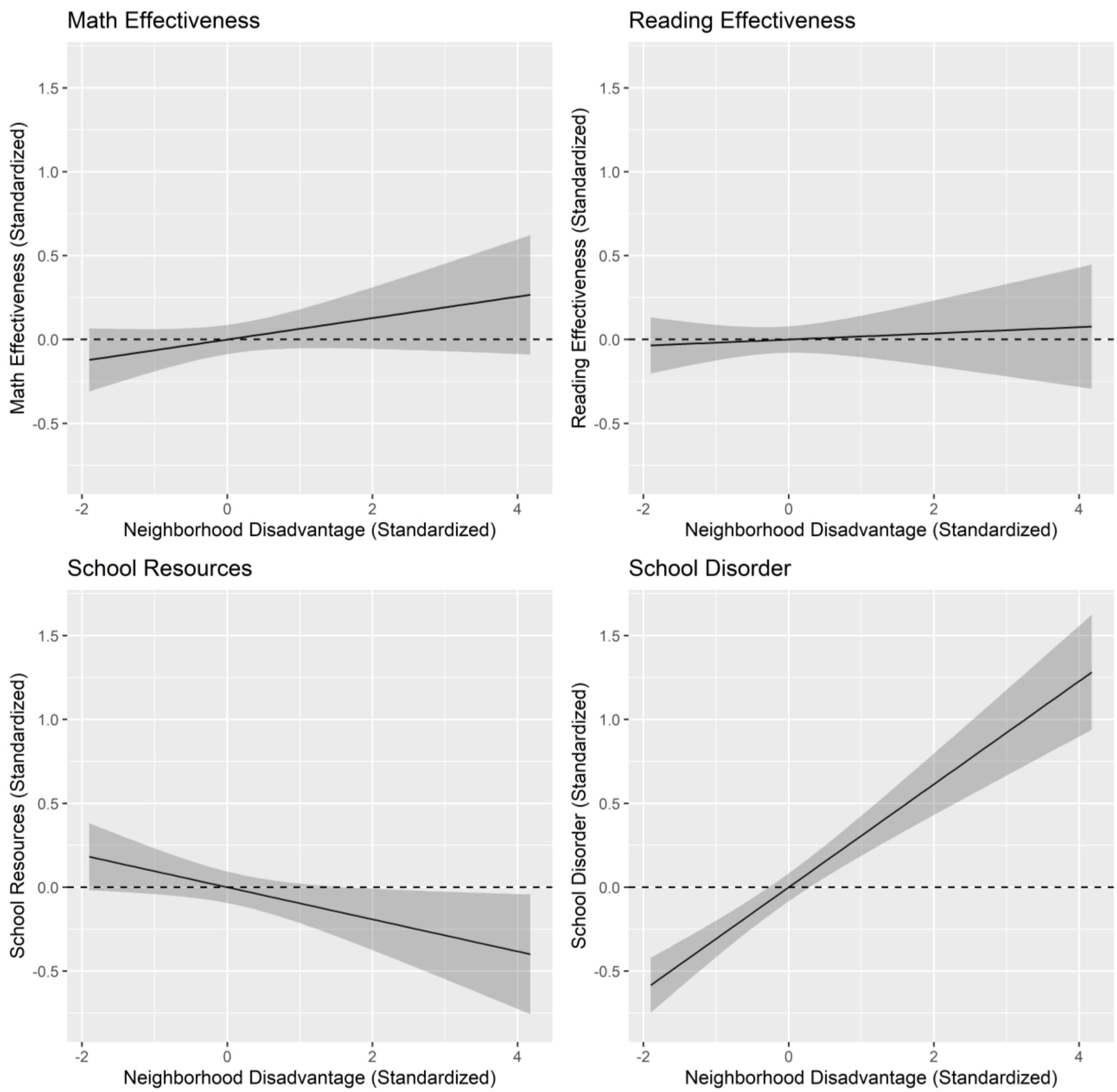

Figure 3. The bivariate relationship between school quality and neighborhood disadvantage, ECLS-K Class of 1998-99.

Notes: Estimates are combined across MI datasets.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-4" and "Student record abstract form (kindergarten, 1st grade)"; GeoLytics Neighborhood Change Database, 2013.

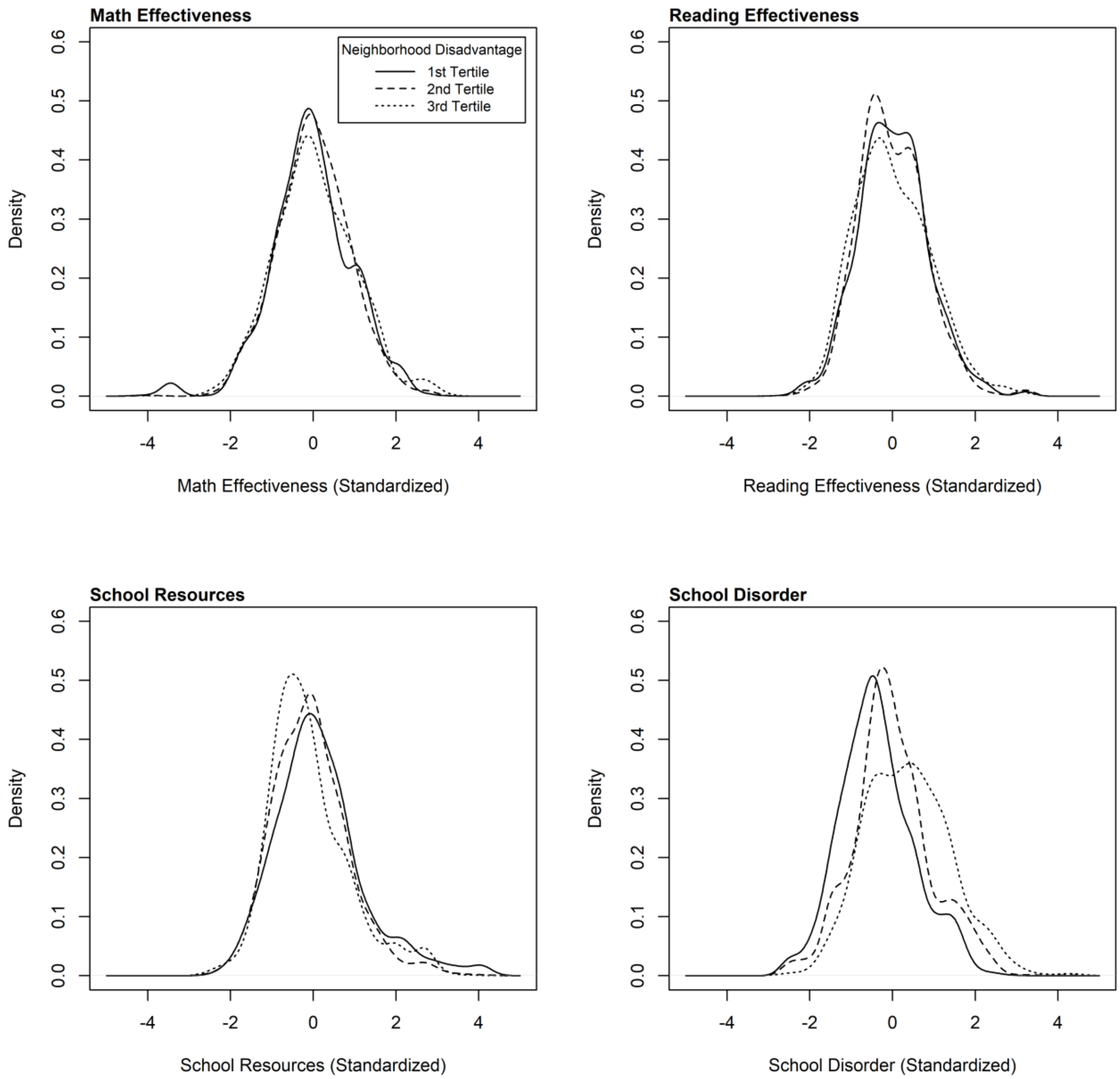

Figure 4. Kernel density plots of school quality across tertiles of neighborhood disadvantage, ECLS-K Class of 1998-99.

Notes: Estimates are combined across MI datasets.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-4" and "Student record abstract form (kindergarten, 1st grade)"; GeoLytics Neighborhood Change Database, 2013.

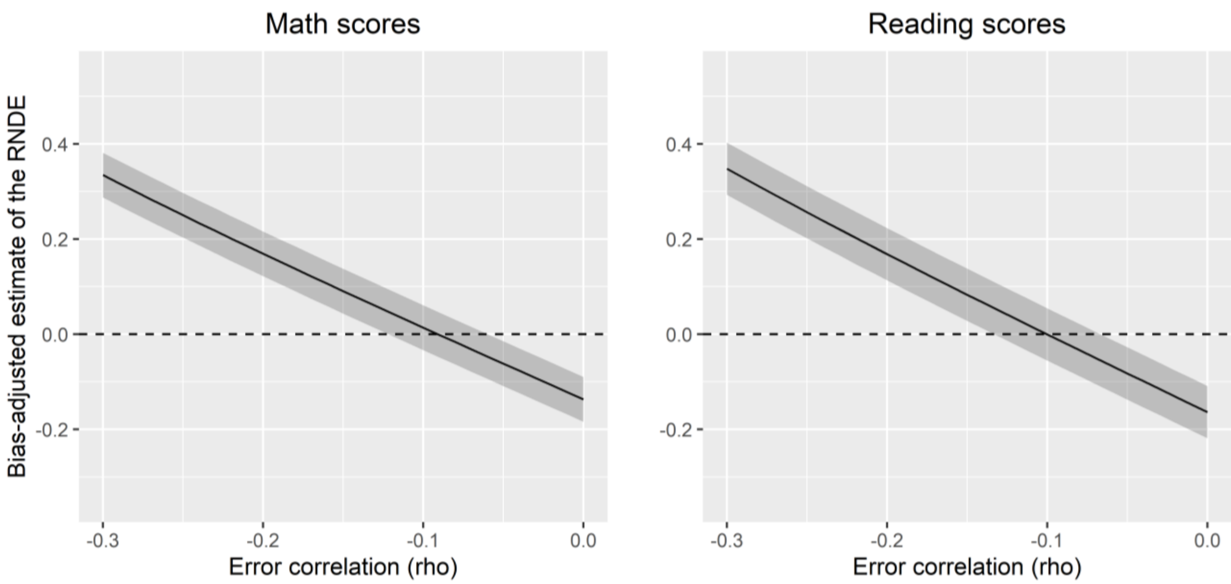

Figure 5. Bias-adjusted Estimates of the $RNDE$ as Functions of the Error Correlation $\rho_{AY} = \mathrm{corr}(\varepsilon_A, \varepsilon_Y)$.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

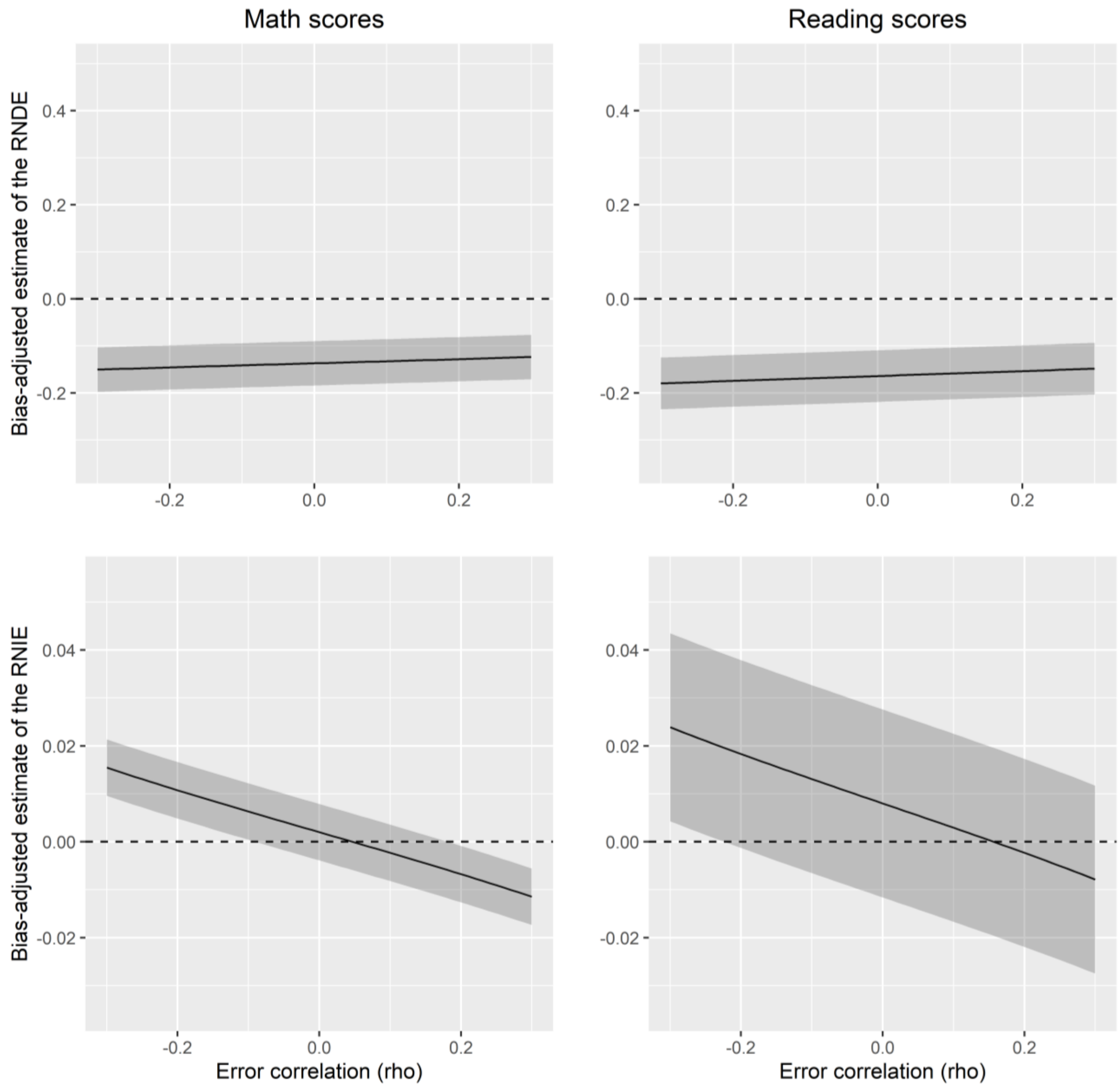

Figure 6. Bias-adjusted Estimates of the $RNDE$ and $RNIE$ as Functions of the Error Correlation $\rho_{MY} = \mathrm{corr}(\varepsilon_M, \varepsilon_Y)$.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

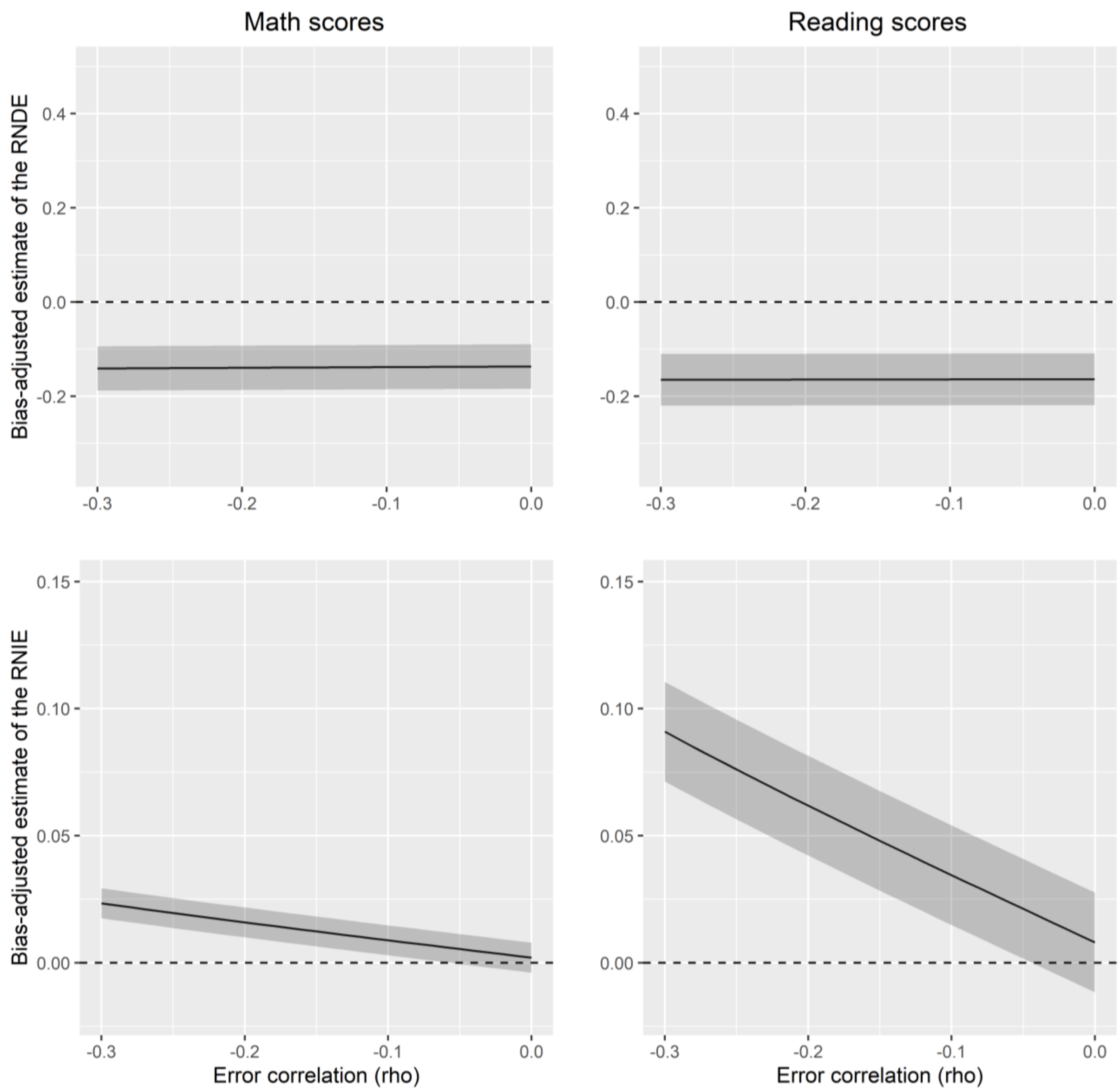

Figure 7. Bias-adjusted Estimates of the $RNDE$ and $RNIE$ as Functions of the Error Correlation $\rho_{AM} = \mathrm{corr}(\varepsilon_A, \varepsilon_M)$.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Tables

Table 1. Child test scores, ECLS-K Class of 1998-99

Variable	Mean	SD
Math test scores		
Fall of kindergarten	0.00	1.00
Spring of kindergarten	1.00	0.97
Fall of 1st grade	1.49	0.97
Spring of 1st Grade	2.52	0.88
Spring of 3rd Grade	3.87	0.81
Spring of 5th Grade	4.66	0.85
Spring of 8th Grade	5.33	0.93
Reading test scores		
Fall of kindergarten	0.00	1.00
Spring of kindergarten	1.11	0.99
Fall of 1st grade	1.53	1.00
Spring of 1st Grade	2.70	0.90
Spring of 3rd Grade	3.99	0.62
Spring of 5th Grade	4.48	0.58
Spring of 8th Grade	4.96	0.75

Notes: Estimates are combined across MI datasets.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-7".

Table 2. Child, neighborhood, and school characteristics, ECLS-K Class of 1998-99

Variable	Mean	SD
Contextual measures		
Neighborhood disadvantage (kindergarten)	0.00	1.00
School poverty (1st grade)	37.64	27.31
School proportion non-white (1st grade)	40.24	36.45
School effectiveness (math, 1st grade)	0.11	0.05
School effectiveness (reading, 1st grade)	0.17	0.05
School resources (1st grade)	0.00	1.00
School disorder (1st grade)	0.00	1.00
Child measures		
Gender		
Male	0.51	
Female	0.49	
Race		
White (non-Hispanic)	0.55	
Black or African American (non-Hispanic)	0.15	
Hispanic	0.17	
Asian	0.05	
Other	0.07	
Birth weight		
Low (<88 ounces)	0.08	

Notes: Estimates are combined across MI datasets.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-4," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Wave 1 Parent Interview"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table 3. Family characteristics, ECLS-K Class of 1998-99

Variable	Mean	SD
Cognitive stimulation scale	0.01	0.48
Mother's age at birth	27.52	6.32
Parental practices scale	0.00	0.38
Parental mental health scale	17.59	5.51
Parental income ($1000s)	49.08	36.90
Household size	4.54	1.44
Parental education		
Less than high school diploma	0.10	
High school diploma or equivalent	0.25	
Vocational/technical degree	0.05	
Some college	0.27	
Bachelor's degree	0.17	
Graduate degree	0.14	
Mother married at birth	0.67	
Father's employment status		
35 hours or more per week	0.86	
Less than 35 hours per week	0.04	
Other	0.10	
Mother's employment status		
35 hours or more per week	0.45	
Less than 35 hours per week	0.22	
Other	0.33	

Notes: Estimates are combined across MI datasets.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Parent Interview, Waves 1 and 2".

Table 4. Decomposition of the Overall Effect of Neighborhood Context on 3rd Grade Math Test Scores into Direct, Indirect, and Interaction Effects, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.135	(0.024)	<0.001	-0.135	(0.024)	<0.001	-0.133	(0.024)	<0.001
$RNDE$	-0.137	(0.024)	<0.001	-0.137	(0.024)	<0.001	-0.129	(0.024)	<0.001
CDE	-0.137	(0.024)	<0.001	-0.135	(0.025)	<0.001	-0.129	(0.024)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.953	-0.003	(0.003)	0.431	0.000	(0.001)	1.000
$RNIE$	0.002	(0.003)	0.532	0.003	(0.003)	0.355	-0.004	(0.003)	0.217
$RPIE$	0.001	(0.003)	0.657	0.000	(0.003)	0.908	-0.003	(0.004)	0.402
$RINT_{\text{med}}$	0.001	(0.003)	0.780	0.002	(0.003)	0.463	-0.001	(0.004)	0.766
$RNDE^{\dagger}$	-0.122	(0.031)	<0.001	-0.125	(0.031)	<0.001	-0.121	(0.031)	<0.001
$RNIE^{\dagger}$	-0.013	(0.015)	0.369	-0.010	(0.015)	0.508	-0.013	(0.015)	0.408

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table 5. Decomposition of the Overall Effect of Neighborhood Context on 3rd Grade Reading Test Scores into Direct, Indirect, and Interaction Effects, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.156	(0.029)	<0.001	-0.155	(0.029)	<0.001	-0.157	(0.029)	<0.001
$RNDE$	-0.164	(0.028)	<0.001	-0.154	(0.029)	<0.001	-0.158	(0.029)	<0.001
CDE	-0.164	(0.028)	<0.001	-0.152	(0.029)	<0.001	-0.158	(0.029)	<0.001
$RINT_{\text{ref}}$	0.000	(0.003)	0.936	-0.002	(0.004)	0.600	0.000	(0.001)	0.928
$RNIE$	0.008	(0.010)	0.430	-0.001	(0.003)	0.827	0.001	(0.004)	0.787
$RPIE$	0.008	(0.010)	0.428	-0.003	(0.003)	0.405	-0.002	(0.004)	0.674
$RINT_{\text{med}}$	0.000	(0.004)	0.932	0.002	(0.004)	0.627	0.003	(0.004)	0.542
$RNDE^{\dagger}$	-0.120	(0.035)	0.001	-0.114	(0.035)	0.001	-0.118	(0.036)	0.001
$RNIE^{\dagger}$	-0.036	(0.020)	0.069	-0.041	(0.018)	0.026	-0.039	(0.019)	0.036

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table 6. Selected Coefficients from Models of School Effectiveness, Resources, Disorder, and Composition, ECLS-K Class of 1998-99

Outcome	Partial effect of neighborhood disadvantage		
	Est.	(SE)	P-value
School composition (exposure-induced confounders)			
School free lunch participation	14.495	(0.584)	<0.001
School proportion nonwhite	9.826	(0.821)	<0.001
School quality (mediators)			
School effectiveness – math	0.037	(0.042)	0.368
School effectiveness – reading	0.036	(0.042)	0.390
School resources	-0.102	(0.043)	0.019
School disorder	0.126	(0.039)	0.001

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal coefficient is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table 7. Selected Coefficients from Models of 3rd Grade Math Test Scores, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
Neighborhood disadvantage									
A	-0.085	(0.015)	<0.001	-0.084	(0.015)	<0.001	-0.080	(0.015)	<0.001
School quality									
M	0.026	(0.022)	0.231	-0.008	(0.012)	0.502	-0.018	(0.014)	0.185
$A \times M$	0.008	(0.016)	0.628	-0.008	(0.010)	0.402	-0.003	(0.011)	0.762
School free lunch participation									
$Z^{\perp}$	-0.001	(0.001)	0.484	-0.001	(0.001)	0.617	0.000	(0.001)	0.689
School proportion non-white									
$Z^{\perp}$	0.000	(0.001)	1.000	0.000	(0.001)	1.000	0.000	(0.001)	1.000

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal coefficient is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table 8. Selected Coefficients from Models of 3rd Grade Reading Test Scores, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
Neighborhood disadvantage									
A	-0.101	(0.017)	<0.001	-0.095	(0.018)	<0.001	-0.097	(0.018)	<0.001
School quality									
M	0.138	(0.032)	<0.001	0.010	(0.014)	0.480	-0.001	(0.015)	0.926
$A \times M$	0.003	(0.024)	0.895	-0.007	(0.012)	0.597	0.008	(0.012)	0.511
School free lunch participation									
$Z^{\perp}$	-0.001	(0.001)	0.484	0.000	(0.001)	0.689	0.000	(0.001)	0.689
School proportion non-white									
$Z^{\perp}$	-0.002	(0.001)	0.005	-0.002	(0.001)	0.002	-0.002	(0.001)	0.002

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal coefficient is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

ONLINE SUPPLEMENT

Part A: Measuring School Effectiveness

In this appendix, we explain our approach to measuring school effectiveness. We operationalize school effectiveness as the difference between a school's average learning rate among its 1st grade students during the school year and the average learning rate among those same students during the previous summer. If all students in the ECLS-K were tested on the first and last days of both kindergarten and 1st grade, then school-year versus summer learning rates could be estimated directly by subtracting successive test scores. The ECLS-K, however, visited schools to administer assessments on a staggered schedule. As a result, students at different schools may have been tested anywhere from one to three months from the beginning or end of the school year as part of the spring and fall assessments. To adjust for the differential timing of these tests, we follow Downey et al. (2008, 2019) and model test scores as a linear function of the amount of time that each child had spent in kindergarten, on summer break, and in 1st grade at the time each test was administered.

Specifically, we model test scores measured at time t for child i in school j, which are here denoted by SCR_{tij}, as follows:

$$SCR_{tij} = (\gamma_0 + \mu_{0j} + \tau_{0ij}) + KND_{tij}(\gamma_1 + \mu_{1j} + \tau_{1ij}) + SUM_{tij}(\gamma_2 + \mu_{2j} + \tau_{2ij}) + FST_{tij}(\gamma_3 + \mu_{3j} + \tau_{3ij}) + \varepsilon_{tij},$$

where there are $t = 1, \dots, 4$ testing occasions between the start of kindergarten and the end of 1st grade and where KND_{tij}, SUM_{tij}, and FST_{tij} respectively denote the amount of time in months that a child had spent in kindergarten, on summer break, and in 1st grade prior to each testing occasion. In this model, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ is a vector of fixed effects that capture the achievement level and learning rates during kindergarten, summer, and 1st grade averaged across all schools; $\boldsymbol{\mu}_j = (\mu_{0j}, \mu_{1j}, \mu_{2j}, \mu_{3j})$ is a vector of random effects that capture each school's departure from the overall average achievement level and learning rates; and $\boldsymbol{\tau}_{ij} = (\tau_{0ij}, \tau_{1ij}. \tau_{2ij}, \tau_{3ij})$ is another vector of random effects that capture each child's deviation from their school's average achievement level and learning rates. We assume that $\boldsymbol{\mu}_j$ and $\boldsymbol{\tau}_{ij}$ are uncorrelated and that both follow multivariate normal distributions with zero means and unrestricted covariance matrices. The disturbance term in this model, ε_{tij}, represents random measurement error, whose variance at each time t is constrained to equal the total variance of the test scores multiplied by one minus their reliability.

We fit this model by the method of maximum likelihood to data from our analytic sample of children in the ECLS-K after imposing several additional restrictions. Specifically, we exclude children who do not have valid school identifiers in waves 1 to 4, who attended a school with a year-round academic calendar or that required attendance at a summer school program, or who transferred schools during either school year. After these restrictions, the median number of students per school is 15, with an interquartile range of 12 to 18. With maximum likelihood estimates (MLEs) of the fixed effects and variance components, we then compute best linear unbiased predictions (BLUPs) of the school-level random effects. Finally, for each school j, we compute its quality as $(\hat{\gamma}_3 + \tilde{\mu}_{3j}) - (\hat{\gamma}_2 + \tilde{\mu}_{2j})$, where "hats" denote MLEs and "tildes" denote BLUPs. In this expression, $(\hat{\gamma}_3 + \tilde{\mu}_{3j})$ is the predicted learning rate among students in school j during 1st grade, and $(\hat{\gamma}_2 + \tilde{\mu}_{2j})$ is the predicted learning rate among the same students over the previous summer. Under the assumptions outlined previously, the difference between them isolates the degree to which a school increases its students' learning rates above those that would prevail had its students not attended school. It thereby reflects a school's effectiveness more accurately than other measures that confound the influence of school- and non-school factors or that have only tenuous connections to student achievement.

Part B: Parallel Analyses of 5^{th} and 8^{th} Grade Achievement Test Scores

In this appendix, we present results from a parallel analysis of neighborhood effects on achievement test scores measured during the spring of 5^{th} grade and the spring of 8^{th} grade. Tables B.1 and B.2 present estimated effects on 5^{th} grade achievement. Tables B.3 and B.4 present estimated effects on 8^{th} grade achievement. These estimates are very similar to those presented in the main text that focus on achievement measured at the spring of 3^{rd} grade. This suggests that living in a disadvantaged neighborhood during kindergarten has lasting effects on achievement through the end of middle school. It also suggests that these effects, like those on 3^{rd} grade achievement, are not explained by differences in school quality measured earlier during 1^{st} grade.

Table B.1. Decomposition of the Overall Effect of Neighborhood Context on 5th Grade Math Test Scores into Direct, Indirect, and Interaction Effects, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.164	(0.027)	<0.001	-0.163	(0.027)	<0.001	-0.161	(0.027)	<0.001
$RNDE$	-0.166	(0.027)	<0.001	-0.162	(0.027)	<0.001	-0.157	(0.028)	<0.001
CDE	-0.166	(0.027)	<0.001	-0.160	(0.028)	<0.001	-0.157	(0.028)	<0.001
$RINT_{\mathrm{ref}}$	0.000	(0.003)	0.936	-0.002	(0.003)	0.659	0.000	(0.001)	0.920
$RNIE$	0.002	(0.003)	0.583	-0.001	(0.003)	0.617	-0.004	(0.004)	0.265
$RPIE$	0.000	(0.003)	1.000	-0.003	(0.003)	0.352	-0.002	(0.004)	0.608
$RINT_{\mathrm{med}}$	0.002	(0.004)	0.607	0.001	(0.003)	0.675	-0.002	(0.004)	0.626
$RNDE^{\dagger}$	-0.170	(0.033)	<0.001	-0.169	(0.033)	<0.001	-0.167	(0.033)	<0.001
$RNIE^{\dagger}$	0.006	(0.015)	0.696	0.006	(0.015)	0.708	0.006	(0.015)	0.696

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-6,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table B.2. Decomposition of the Overall Effect of Neighborhood Context on 5th Grade Reading Test Scores into Direct, Indirect, and Interaction Effects, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.156	(0.028)	<0.001	-0.153	(0.029)	<0.001	-0.156	(0.029)	<0.001
$RNDE$	-0.162	(0.028)	<0.001	-0.153	(0.029)	<0.001	-0.155	(0.029)	<0.001
CDE	-0.162	(0.028)	<0.001	-0.149	(0.029)	<0.001	-0.155	(0.029)	<0.001
$RINT_{\text{ref}}$	0.001	(0.003)	0.842	-0.003	(0.004)	0.464	0.000	(0.001)	0.934
$RNIE$	0.006	(0.008)	0.444	-0.001	(0.003)	0.827	-0.001	(0.004)	0.710
$RPIE$	0.005	(0.007)	0.466	-0.003	(0.004)	0.346	-0.005	(0.004)	0.252
$RINT_{\text{med}}$	0.001	(0.004)	0.819	0.003	(0.004)	0.494	0.003	(0.004)	0.395
$RNDE^{\dagger}$	-0.132	(0.035)	<0.001	-0.126	(0.035)	<0.001	-0.131	(0.035)	<0.001
$RNIE^{\dagger}$	-0.024	(0.018)	0.169	-0.028	(0.017)	0.105	-0.025	(0.018)	0.148

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-6,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table B.3. Decomposition of the Overall Effect of Neighborhood Context on 8th Grade Math Test Scores into Direct, Indirect, and Interaction Effects, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.163	(0.032)	<0.001	-0.162	(0.032)	<0.001	-0.162	(0.033)	<0.001
$RNDE$	-0.163	(0.032)	<0.001	-0.160	(0.032)	<0.001	-0.157	(0.033)	<0.001
CDE	-0.163	(0.032)	<0.001	-0.160	(0.033)	<0.001	-0.157	(0.033)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.931	0.000	(0.004)	1.000	0.000	(0.001)	1.000
$RNIE$	0.001	(0.003)	0.795	-0.002	(0.003)	0.519	-0.005	(0.004)	0.238
$RPIE$	-0.001	(0.003)	0.790	-0.002	(0.003)	0.505	-0.005	(0.005)	0.259
$RINT_{\text{med}}$	0.002	(0.003)	0.639	0.000	(0.003)	1.000	0.001	(0.004)	0.871
$RNDE^{\dagger}$	-0.126	(0.039)	0.001	-0.125	(0.039)	0.001	-0.127	(0.039)	0.001
$RNIE^{\dagger}$	-0.036	(0.016)	0.025	-0.037	(0.016)	0.021	-0.035	(0.016)	0.027

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-7," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table B.4. Decomposition of the Overall Effect of Neighborhood Context on 8th Grade Reading Test Scores into Direct, Indirect, and Interaction Effects, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.159	(0.030)	<0.001	-0.152	(0.029)	<0.001	-0.155	(0.030)	<0.001
$RNDE$	-0.159	(0.030)	<0.001	-0.150	(0.029)	<0.001	-0.154	(0.030)	<0.001
CDE	-0.161	(0.030)	<0.001	-0.146	(0.029)	<0.001	-0.154	(0.030)	<0.001
$RINT_{\text{ref}}$	0.001	(0.003)	0.644	-0.004	(0.004)	0.282	0.000	(0.001)	0.939
$RNIE$	0.001	(0.003)	0.836	-0.002	(0.003)	0.571	-0.002	(0.004)	0.666
$RPIE$	-0.002	(0.004)	0.733	-0.005	(0.004)	0.163	-0.005	(0.005)	0.261
$RINT_{\text{med}}$	0.002	(0.004)	0.600	0.004	(0.004)	0.356	0.004	(0.004)	0.377
$RNDE^{\dagger}$	-0.096	(0.037)	0.010	-0.089	(0.037)	0.016	-0.096	(0.038)	0.012
$RNIE^{\dagger}$	-0.063	(0.016)	<0.001	-0.063	(0.017)	<0.001	-0.059	(0.017)	<0.001

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-7,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Part C: Derivation of Parametric Expressions for the CDE, $RINT_{\text{ref}}$, $RPIE$, and $RINT_{\text{med}}$

In this appendix, we derive parametric expressions for the direct, indirect, and interaction effects of interest. If $Y(a,m) \perp A|C; Y(a,m) \perp M|C,A,Z;$ and $M(a) \perp A|C$, VanderWeele et al. (2014) show that the $RNDE$ and $RNIE$ can be expressed in terms of the observed data as follows:

$$RNDE = E\left(Y\big(a^*, M^R(a|C)\big) - Y\big(a, M^R(a|C)\big)\right) = \textstyle\sum_c \sum_m \sum_z \big(E(Y|c,a^*,z,m)P(z|c,a^*) -$$

$$E(Y|c,a,z,m)P(z|c,a)\big)\, P(m|c,a)\, P(c) \text{ and}$$

$$RNIE = E\left(Y\big(a^*, M^R(a^*|C)\big) - Y\big(a^*, M^R(a|C)\big)\right) = \textstyle\sum_c \sum_m \sum_z \big(P(m|c,a^*) -$$

$$P(m|c,a)\big)E(Y|c,a^*,z,m)P(z|c,a^*)P(c).$$

If, in addition, the conditional mean of M given $\{C,A\}$ is equal to

$$E(M|C,A) = \theta_0 + \theta_1(C - \alpha_0) + \theta_2 A,$$

and the conditional mean Y given $\{C,A,Z,M\}$ is equal to

$$E(Y|C,A,Z,M) = \lambda_0 + \lambda_1(C-\alpha_0) + \lambda_2 A + \lambda_3\big(Z - (\beta_0 + \beta_1 C + \beta_2 A)\big) + M(\lambda_4 + \lambda_5 A),$$

where $E(C) = \alpha_0$ and $E(Z|C,A) = \beta_0 + \beta_1 C + \beta_2 A$, then

$$RNDE = \sum_c \sum_m \sum_z \big(E(Y|c,a^*,z,m)P(z|c,a^*) - E(Y|c,a,z,m)P(z|c,a)\big)\, P(m|c,a)\, P(c)$$

$$= \sum_c \sum_m \sum_z \Big(\big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a^* + \lambda_3\big(z - E(Z|c,a^*)\big)$$

$$+ m(\lambda_4 + \lambda_5 a^*)\big)\, P(z|c,a^*)$$

$$- \big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a + \lambda_3\big(z - E(Z|c,a)\big)$$

$$+ m(\lambda_4 + \lambda_5 a)\big)\, P(z|c,a)\Big)\, P(m|c,a)\, P(c)$$

$$= \sum_c \sum_m \Big(\big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a^* + \lambda_3\big(E(Z|c,a^*) - E(Z|c,a^*)\big)$$

$$+ m(\lambda_4 + \lambda_5 a^*)\big)$$

$$- \big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a + \lambda_3\big(E(Z|c,a) - E(Z|c,a)\big)$$

$$+ m(\lambda_4 + \lambda_5 a)\big)\Big)\, P(m|c,a)\, P(c)$$

$$= \sum_c \sum_m \Big(\big(\lambda_2 a^* + m(\lambda_4 + \lambda_5 a^*)\big) - \big(\lambda_2 a + m(\lambda_4 + \lambda_5 a)\big)\Big)\, P(m|c,a)\, P(c)$$

$$= \sum_c \Big(\big(\lambda_2 a^* + E(M|c,a)(\lambda_4 + \lambda_5 a^*)\big) - \big(\lambda_2 a + E(M|c,a)(\lambda_4 + \lambda_5 a)\big)\Big)\, P(c)$$

$$= \sum_c \Big(\Big(\lambda_2 a^* + \big(\theta_0 + \theta_1\big(c - E(C)\big) + \theta_2 a\big)(\lambda_4 + \lambda_5 a^*)\Big)$$

$$- \Big(\lambda_2 a + \big(\theta_0 + \theta_1\big(c - E(C)\big) + \theta_2 a\big)(\lambda_4 + \lambda_5 a)\Big)\Big)\, P(c)$$

$$= \Big(\Big(\lambda_2 a^* + \big(\theta_0 + \theta_1\big(E(C) - E(C)\big) + \theta_2 a\big)(\lambda_4 + \lambda_5 a^*)\Big)$$

$$- \Big(\lambda_2 a + \big(\theta_0 + \theta_1\big(E(C) - E(C)\big) + \theta_2 a\big)(\lambda_4 + \lambda_5 a)\Big)\Big)$$

$$= \big(\lambda_2 + \lambda_5(\theta_0 + \theta_2 a)\big)(a^* - a)$$

and

$$RNIE = \sum_c \sum_m \sum_z \big(P(m|c,a^*) - P(m|c,a)\big) E(Y|c,a^*,z,m) P(z|c,a^*) P(c)$$

$$= \sum_c \sum_m \sum_z \big(P(m|c,a^*) - P(m|c,a)\big) \Big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a^*$$

$$+ \lambda_3 \Big(z - E\big(Z|c,a^*\big)\Big) + m(\lambda_4 + \lambda_5 a^*)\Big) P(z|c,a^*) P(c)$$

$$= \sum_c \sum_m \big(P(m|c,a^*) - P(m|c,a)\big) \Big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a^*$$

$$+ \lambda_3 \Big(E\big(Z|c,a^*\big) - E\big(Z|c,a^*\big)\Big) + m(\lambda_4 + \lambda_5 a^*)\Big) P(c)$$

$$= \sum_c \Big(\big(E(M|c,a^*)(\lambda_4 + \lambda_5 a^*)\big) - \big(E(M|c,a)(\lambda_4 + \lambda_5 a^*)\big)\Big) P(c)$$

$$= \sum_c \Big(\Big(\big(\theta_0 + \theta_1\big(c - E(C)\big) + \theta_2 a^*\big)(\lambda_4 + \lambda_5 a^*)\Big)$$

$$- \Big(\big(\theta_0 + \theta_1\big(c - E(C)\big) + \theta_2 a\big)(\lambda_4 + \lambda_5 a^*)\Big)\Big) P(c)$$

$$= \Big(\big(\theta_0 + \theta_1\big(E(C) - E(C)\big) + \theta_2 a^*\big)(\lambda_4 + \lambda_5 a^*)\Big)$$

$$- \Big(\big(\theta_0 + \theta_1\big(E(C) - E(C)\big) + \theta_2 a\big)(\lambda_4 + \lambda_5 a^*)\Big)$$

$$= \theta_2(\lambda_4 + \lambda_5 a^*)(a^* - a).$$

Under the same ignorability assumptions defined previously, the CDE can be expressed in terms of the observed as

$$CDE = E\big(Y(a^*,m) - Y(a,m)\big) = \sum_c \sum_z \big(E(Y|c,a^*,z,m)P(z|c,a^*) -$$

$$E(Y|c,a,z,m)P(z|c,a)\big) P(c),$$

and given correct models for the outcome, mediator, and exposure-induced confounders, it is equal to

$$CDE = \sum_c \sum_z \big(E(Y|c, a^*, z, m)P(z|c, a^*) - E(Y|c, a, z, m)P(z|c, a)\big)\, P(c)$$

$$= \sum_c \sum_z \Big(\big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a^* + \lambda_3\big(z - E(Z|c, a^*)\big)$$

$$+ m(\lambda_4 + \lambda_5 a^*)\big)\, P(z|c, a^*)$$

$$- \big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a + \lambda_3\big(z - E(Z|c, a)\big)$$

$$+ m(\lambda_4 + \lambda_5 a)\big)\, P(z|c, a)\Big)\, P(c)$$

$$= \sum_c \Big(\big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a^* + \lambda_3\big(E(Z|c, a^*) - E(Z|c, a^*)\big)$$

$$+ m(\lambda_4 + \lambda_5 a^*)\big)$$

$$- \big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a + \lambda_3\big(E(Z|c, a) - E(Z|c, a)\big)$$

$$+ m(\lambda_4 + \lambda_5 a)\big)\Big)\, P(c) = \sum_c \big((\lambda_2 a^* + \lambda_5 m a^*) - (\lambda_2 a + \lambda_5 a m)\big)P(c)$$

$$= (\lambda_2 + \lambda_5 m)(a^* - a).$$

By extension, the reference interaction effect is equal to

$$RINT_{\mathrm{ref}} = RNDE - CDE$$

$$= \Big(\big(\lambda_2 + \lambda_5(\theta_0 + \theta_2 a)\big)(a^* - a)\Big) - \big((\lambda_2 + \lambda_5 m)(a^* - a)\big)$$

$$= \lambda_5(\theta_0 + \theta_2 a - m)(a^* - a).$$

Similarly, VanderWeele (2014) shows that the pure indirect effect can be expressed in terms of the observed data as

$$RPIE = E\Big(Y\big(a, M^R(a^*|C)\big) - Y\big(a, M^R(a|C)\big)\Big) = \sum_c \sum_m \sum_z \big(P(m|c, a^*) - P(m|c, a)\big)E(Y|c, a, z, m)P(z|c, a)P(c),$$

which, under the models outlined previously, is equal to

$$RPIE = \sum_c \sum_m \sum_z \big(P(m|c,a^*) - P(m|c,a)\big)E(Y|c,a,z,m)P(z|c,a)P(c)$$

$$= \sum_c \sum_m \sum_z \big(P(m|c,a^*) - P(m|c,a)\big)\Big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a$$

$$+ \lambda_3\big(z - E(Z|c,a)\big) + m(\lambda_4 + \lambda_5 a)\Big)P(z|c,a)P(c)$$

$$= \sum_c \sum_m \big(P(m|c,a^*) - P(m|c,a)\big)\Big(\lambda_0 + \lambda_1\big(c - E(C)\big) + \lambda_2 a$$

$$+ \lambda_3\big(E(Z|c,a) - E(Z|c,a)\big) + m(\lambda_4 + \lambda_5 a)\Big)P(c)$$

$$= \sum_c \big(E(M|c,a^*)(\lambda_4 + \lambda_5 a)\big) - \big(E(M|c,a)(\lambda_4 + \lambda_5 a)\big)P(c)$$

$$= \sum_c \Big(\big(\theta_0 + \theta_1\big(c - E(C)\big) + \theta_2 a^*\big)(\lambda_4 + \lambda_5 a)\Big)$$

$$- \Big(\big(\theta_0 + \theta_1\big(c - E(C)\big) + \theta_2 a\big)(\lambda_4 + \lambda_5 a)\Big)P(c)$$

$$= \Big(\big(\theta_0 + \theta_1\big(E(C) - E(C)\big) + \theta_2 a^*\big)(\lambda_4 + \lambda_5 a)\Big)$$

$$- \Big(\big(\theta_0 + \theta_1\big(E(C) - E(C)\big) + \theta_2 a\big)(\lambda_4 + \lambda_5 a)\Big) = \theta_2(\lambda_4 + \lambda_5 a)(a^* - a).$$

And by extension, the mediated interaction effect is equal to

$$RINT_{\text{med}} = RNIE - RPIE$$

$$= \big(\theta_2(\lambda_4 + \lambda_5 a^*)(a^* - a)\big) - \big(\theta_2(\lambda_4 + \lambda_5 a)(a^* - a)\big)$$

$$= \theta_2 \lambda_5 (a^* - a)^2.$$

Part D: Effect Estimates under Alternative Reliabilities for School Quality

In this appendix, we present effect estimates across a range of assumed reliabilities for our measure of school effectiveness when implementing the error-in-variables correction. In the main text, we implemented this correction assuming a reliability of 0.7, which is consistent with estimates reported in prior research (von Hippel 2009). Tables D.1 and D.2 report effect estimates from models that assume a reliability of 0.6 and 0.8, respectively. Across the entire range of reliabilities considered here, results from the ECLS-K are substantively similar. In general, they indicate that neighborhood disadvantage negatively affects academic achievement and that school effectiveness does not appreciably mediate or interact with these effects.

Table D.1. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores Estimated from Models that Assume a Reliability of 0.8 for School Quality, ECLS-K Class of 1998-99

	Math Test Scores			Reading Test Scores		
Estimand	Est.	(SE)	P-value	Est.	(SE)	P-value
RATE	-0.135	(0.024)	<0.001	-0.156	(0.029)	<0.001
RNDE	-0.137	(0.024)	<0.001	-0.163	(0.028)	<0.001
CDE	-0.137	(0.024)	<0.001	-0.163	(0.028)	<0.001
$RINT_{\text{ref}}$	0.000	(0.001)	0.943	0.000	(0.002)	0.881
RNIE	0.002	(0.003)	0.529	0.007	(0.008)	0.425
RPIE	0.001	(0.002)	0.650	0.006	(0.008)	0.427
$RINT_{\text{med}}$	0.001	(0.002)	0.775	0.000	(0.003)	0.890

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table D.2. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores Estimated from Models that Assume a Reliability of 0.6 for School Quality, ECLS-K Class of 1998-99

	Math Test Scores			Reading Test Scores		
Estimand	Est.	(SE)	P-value	Est.	(SE)	P-value
RATE	-0.135	(0.024)	<0.001	-0.155	(0.030)	<0.001
RNDE	-0.138	(0.024)	<0.001	-0.165	(0.029)	<0.001
CDE	-0.138	(0.024)	<0.001	-0.165	(0.030)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.960	0.000	(0.003)	1.000
RNIE	0.002	(0.004)	0.538	0.011	(0.014)	0.438
RPIE	0.002	(0.003)	0.650	0.011	(0.014)	0.446
$RINT_{\text{med}}$	0.001	(0.003)	0.764	0.000	(0.005)	0.983

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Part E: Sensitivity of Effect Estimates to Alternative Model Specifications

The models on which we focus in the main text constrain the effects of the exposure and mediator to be invariant across levels of the confounders. If these constraints are inappropriate and the effects of interest are not in fact invariant, then the estimates we report in the main text may suffer from bias due to model misspecification. In this appendix, we present effect estimates from models for school quality and student achievement that permit effect heterogeneity by race, gender, parental education, and across the rural-to-urban continuum.

Specifically, we present effect estimates from models of school quality with form

$$E(M|C,A) = \theta_0 + \theta_1\delta(C) + \theta_2 A + \theta_3\delta(C^*)A$$

and from models of the outcome with form

$$E(Y|C,A,Z,M) = \lambda_0 + \lambda_1\delta(C)$$

$$+\lambda_2 A + \lambda_3\delta(Z) + M(\lambda_4 + \lambda_5 A) + \delta(C^*)\big(\lambda_6 A + M(\lambda_7 + \lambda_8 A)\big),$$

where $\delta(C) = C - E(C)$, $\delta(Z) = Z - E(Z|C,A)$ and $\delta(C^*)$ denotes selected elements of $\delta(C)$. In the first model, the interaction term $\theta_3\delta(C^*)A$ allows the effect of treatment on the mediator to differ across levels of C^*. In the second model, the interaction term $\delta(C^*)\big(\lambda_6 A + M(\lambda_7 + \lambda_8 A)\big)$ allows the effects of the exposure and mediator on the outcome to differ across levels of C^*. A convenient property of these interaction terms is that they are equal to zero when averaged over C^* (Wodtke et al. 2020). This implies that the direct, indirect, and interaction effects of interest can be constructed using exactly the same parametric expressions as provided in the main text, even though the models on which they are based no longer constrain these effects to be invariant across C^*.

Tables E.1 to E.6 present results from models that permit the effects of interest to differ by race, gender, and parental education, respectively. These estimates are very similar to those reported in the main text, which suggests that our key findings are robust to effect heterogeneity across key demographic subgroups.

Tables E.7 and E.8 present results from models that additionally adjust for rural versus urban residence using the U.S. Department of Agriculture's rural-to-urban continuum codes (RUCCs) and then allow the coefficients of interest to differ across this covariate. RUCCs classify counties by population size, their degree of urbanization, and their incorporation in or adjacency to metropolitan areas. Specifically, counties in metropolitan areas are coded 1 to 3 depending on whether they have more than 1 million residents, between 250,000 and 1 million residents, or less than 250,000 residents, respectively, while counties outside of metropolitan areas are coded 4 through 9, with categories ranging from "urban population of 20,000 residents or more that is adjacent to a metropolitan area" to "completely rural population or urban population with less than 2,500 residents that is not adjacent to a metropolitan area." Effect estimates from models that allow for heterogeneity across the urban-to-rural continuum are also substantively similar to those we prioritize in the main text.

Table E.1. Effects of Neighborhood Context on 3rd Grade Math Test Scores Estimated from Models that Permit Heterogeneity by Race, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.141	(0.025)	<0.001	-0.139	(0.025)	<0.001	-0.138	(0.025)	<0.001
$RNDE$	-0.144	(0.025)	<0.001	-0.142	(0.025)	<0.001	-0.133	(0.025)	<0.001
CDE	-0.144	(0.025)	<0.001	-0.139	(0.025)	<0.001	-0.133	(0.025)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.916	-0.003	(0.004)	0.389	0.000	(0.001)	1.000
$RNIE$	0.003	(0.004)	0.412	0.003	(0.003)	0.348	-0.004	(0.004)	0.222
$RPIE$	0.002	(0.004)	0.637	0.000	(0.003)	1.000	-0.003	(0.004)	0.418
$RINT_{\text{med}}$	0.002	(0.003)	0.659	0.003	(0.004)	0.389	-0.001	(0.004)	0.720

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table E.2. Effects of Neighborhood Context on 3rd Grade Reading Test Scores Estimated from Models that Permit Heterogeneity by Race, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.154	(0.029)	<0.001	-0.158	(0.030)	<0.001	-0.159	(0.030)	<0.001
$RNDE$	-0.165	(0.029)	<0.001	-0.158	(0.030)	<0.001	-0.160	(0.030)	<0.001
CDE	-0.165	(0.030)	<0.001	-0.156	(0.030)	<0.001	-0.160	(0.030)	<0.001
$RINT_{\text{ref}}$	0.000	(0.003)	1.000	-0.002	(0.004)	0.617	0.000	(0.001)	1.000
$RNIE$	0.010	(0.011)	0.327	0.000	(0.004)	0.920	0.001	(0.004)	0.818
$RPIE$	0.010	(0.011)	0.322	-0.003	(0.003)	0.445	-0.001	(0.004)	0.732
$RINT_{\text{med}}$	0.000	(0.004)	1.000	0.002	(0.004)	0.633	0.002	(0.004)	0.617

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table E.3. Effects of Neighborhood Context on 3rd Grade Math Test Scores Estimated from Models that Permit Heterogeneity by Gender, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.134	(0.024)	<0.001	-0.134	(0.024)	<0.001	-0.133	(0.024)	<0.001
$RNDE$	-0.136	(0.024)	<0.001	-0.137	(0.024)	<0.001	-0.129	(0.024)	<0.001
CDE	-0.137	(0.024)	<0.001	-0.134	(0.024)	<0.001	-0.129	(0.024)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.953	-0.003	(0.003)	0.445	0.000	(0.001)	1.000
$RNIE$	0.002	(0.003)	0.532	0.002	(0.003)	0.374	-0.004	(0.003)	0.217
$RPIE$	0.001	(0.003)	0.657	0.000	(0.003)	0.908	-0.003	(0.004)	0.418
$RINT_{\text{med}}$	0.001	(0.003)	0.780	0.002	(0.003)	0.463	-0.001	(0.004)	0.752

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table E.4. Effects of Neighborhood Context on 3rd Grade Reading Test Scores Estimated from Models that Permit Heterogeneity by Gender, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.155	(0.029)	<0.001	-0.154	(0.029)	<0.001	-0.157	(0.029)	<0.001
$RNDE$	-0.164	(0.028)	<0.001	-0.154	(0.029)	<0.001	-0.158	(0.029)	<0.001
CDE	-0.164	(0.028)	<0.001	-0.151	(0.029)	<0.001	-0.158	(0.029)	<0.001
$RINT_{\mathrm{ref}}$	0.000	(0.003)	0.936	-0.002	(0.004)	0.626	0.000	(0.001)	0.928
$RNIE$	0.008	(0.011)	0.429	-0.001	(0.003)	0.803	0.001	(0.004)	0.787
$RPIE$	0.008	(0.010)	0.428	-0.003	(0.003)	0.405	-0.002	(0.004)	0.682
$RINT_{\mathrm{med}}$	0.000	(0.004)	0.934	0.002	(0.004)	0.646	0.003	(0.004)	0.526

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table E.5. Effects of Neighborhood Context on 3rd Grade Math Test Scores Estimated from Models that Permit Heterogeneity by Parental Education, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.138	(0.025)	<0.001	-0.137	(0.024)	<0.001	-0.133	(0.024)	<0.001
$RNDE$	-0.140	(0.025)	<0.001	-0.139	(0.024)	<0.001	-0.129	(0.025)	<0.001
CDE	-0.140	(0.025)	<0.001	-0.137	(0.025)	<0.001	-0.129	(0.025)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.928	-0.002	(0.003)	0.659	0.000	(0.001)	0.920
$RNIE$	0.002	(0.004)	0.573	0.002	(0.003)	0.594	-0.004	(0.004)	0.330
$RPIE$	0.001	(0.004)	0.775	0.000	(0.003)	0.947	-0.002	(0.004)	0.600
$RINT_{\text{med}}$	0.001	(0.004)	0.753	0.001	(0.003)	0.681	-0.002	(0.004)	0.703

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table E.6. Effects of Neighborhood Context on 3rd Grade Reading Test Scores Estimated from Models that Permit Heterogeneity by Parental Education, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.148	(0.029)	<0.001	-0.152	(0.029)	<0.001	-0.152	(0.029)	<0.001
$RNDE$	-0.157	(0.028)	<0.001	-0.151	(0.029)	<0.001	-0.152	(0.029)	<0.001
CDE	-0.157	(0.029)	<0.001	-0.150	(0.029)	<0.001	-0.152	(0.029)	<0.001
$RINT_{\mathrm{ref}}$	0.000	(0.003)	0.972	-0.001	(0.004)	0.822	0.000	(0.001)	0.928
$RNIE$	0.009	(0.012)	0.423	-0.001	(0.004)	0.864	0.000	(0.004)	0.943
$RPIE$	0.010	(0.012)	0.416	-0.002	(0.003)	0.650	-0.001	(0.004)	0.884
$RINT_{\mathrm{med}}$	0.000	(0.004)	0.959	0.001	(0.004)	0.826	0.001	(0.005)	0.845

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table E.7. Effects of Neighborhood Context on 3rd Grade Math Test Scores Estimated from Models that Permit Heterogeneity across the Rural to Urban Continuum, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.148	(0.026)	<0.001	-0.155	(0.028)	<0.001	-0.145	(0.025)	<0.001
$RNDE$	-0.149	(0.026)	<0.001	-0.158	(0.029)	<0.001	-0.141	(0.026)	<0.001
CDE	-0.149	(0.026)	<0.001	-0.153	(0.027)	<0.001	-0.141	(0.026)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.912	-0.005	(0.005)	0.299	0.000	(0.001)	1.000
$RNIE$	0.001	(0.003)	0.590	0.003	(0.003)	0.317	-0.004	(0.004)	0.317
$RPIE$	0.002	(0.004)	0.537	-0.001	(0.004)	0.710	-0.003	(0.004)	0.710
$RINT_{\text{med}}$	-0.001	(0.003)	0.809	0.005	(0.005)	0.317	-0.001	(0.004)	0.910

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, “Direct Child Assessments in waves 1-5,” “Student record abstract form (kindergarten and 1st grade),” “School Administrator Questionnaire in First Grade,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000; U.S. Department of Agriculture, Economic Research Service, The 2013 Rural-Urban Continuum Codes.

Table E.8. Effects of Neighborhood Context on 3rd Grade Reading Test Scores Estimated from Models that Permit Heterogeneity across the Rural to Urban Continuum, ECLS-K Class of 1998-99

Estimand	School effectiveness			School resources			School disorder		
	Est.	(SE)	P-value	Est.	(SE)	P-value	Est.	(SE)	P-value
$RATE$	-0.176	(0.031)	<0.001	-0.192	(0.032)	<0.001	-0.179	(0.030)	<0.001
$RNDE$	-0.183	(0.030)	<0.001	-0.194	(0.033)	<0.001	-0.178	(0.030)	<0.001
CDE	-0.183	(0.030)	<0.001	-0.187	(0.031)	<0.001	-0.178	(0.030)	<0.001
$RINT_{\mathrm{ref}}$	0.000	(0.003)	0.939	-0.008	(0.006)	0.188	0.000	(0.001)	1.000
$RNIE$	0.007	(0.008)	0.368	0.002	(0.004)	0.528	-0.001	(0.004)	0.792
$RPIE$	0.008	(0.009)	0.393	-0.004	(0.003)	0.228	-0.002	(0.004)	0.686
$RINT_{\mathrm{med}}$	0.000	(0.004)	0.922	0.006	(0.005)	0.218	0.001	(0.005)	0.868

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000; U.S. Department of Agriculture, Economic Research Service, The 2013 Rural-Urban Continuum Codes.

Part F: Sensitivity of Effect Estimates to Alternative Measures of School Effectiveness

In the main text, we operationalize school effectiveness as the difference between a school's average learning rate among its 1st grade students during the school year and the average learning rate among those same students during the previous summer. This measure captures a school's "value added" with respect to its students' reading and math skills under the following two assumptions: first, the influence of non-school factors on achievement must operate similarly during the school year and the summer, and second, schools must not affect summer learning. If either of these assumptions are violated, then our measure would suffer from systematic error, possibly leading to invalid inferences about the explanatory role of school quality.

One approach to evaluating the sensitivity of our results to potential violations of these assumptions is to reanalyze the data with measures of school effectiveness that subtract only a fraction of the summer learning rate from the school-year learning rate, which adjusts for the possibility that non-school contributions to learning differ during school year versus the summer (Downey et al. 2008). The proper weight to give summer learning is unknown, but it must lie somewhere between one, which is the weight given to it in our analysis from the main text, and zero. In this appendix, we replicate our analysis using, first, a measure that gives the summer learning rate a weight of zero and thus equates a school's effectiveness with its school-year learning rate alone, and second, a measure that equates a school's effectiveness with the difference between its school-year learning rate and one-half the learning rate among its students during the summer. The first of these measures would capture a school's value added if non-school factors have no influence on achievement during the school year, while the second measure would be valid if non-school influences on student learning are half as strong during the school year as compared to the summer. Results based on these two alternative measures are presented in Tables F.1 and F.2. They are very similar to those presented in the main text, regardless of the weight given to the summer learning rate when subtracting it from the school-year learning rate.

Next, we consider the possibility that schools have spillover effects on summer learning by replicating our analysis with two additional measures of school effectiveness. Specifically, we replicate our analysis using a simple average of the school-year learning rate and the summer learning rate—that is, the full year-to-year growth rate from the end of kindergarten to the end of 1^{st} grade as predicted by our seasonal learning models. This measure would capture a school's value added if non-school influences on student learning during the school year were minimal and if schools were responsible for nearly all of the learning achieved by students over the summer. We also replicate our analysis using a weighted average of the school-year learning rate and the summer learning rate, where the summer learning rate is given only one-third as much weight as the school-year learning rate. This measure would accurately capture school value added if non-school influences on student learning during the school year were minimal and if schools were only responsible for a small fraction of the learning achieved by students over the summer. Results based on these two measures of school effectiveness are presented in Tables F.3 and F.4. They, too, are similar to those we present in the main text.

In addition to measures of school effectiveness computed from seasonal learning models (e.g., Downey et al. 2008, 2019), we also measure this construct using more conventional estimates of value added from lagged dependent variable models (e.g., McCaffrey et al. 2004; Raudenbush and Bryk 2002), where test scores at the spring of 1st grade are modeled as a function of those from kindergarten and set of school random effects. Specifically, we consider models with the following form:

$$SCR_{4ij} = \nu_0 + \nu_1 SCR_{2ij} + \nu_2 SCR_{1ij} + \nu_2 X + \zeta_j + \varepsilon_{4ij},$$

where SCR_{4ij} denotes the test score of child i in school j at wave $t = 4$ (spring of 1st grade), SCR_{2ij} and SCR_{1ij} denote scores from assessments administered earlier at waves $t = 2$ and $t = 1$ (the spring and fall of kindergarten), X denotes a vector of basic controls for race, gender, and parental education, $\mathbf{\nu} = (\nu_0, \nu_1, \nu_2, \nu_3)$ is a vector of fixed effects, and finally, ζ_j is a school random effect assumed to be normally distributed with mean zero and unrestricted variance. In this model, the random effect would capture school j's influence on student achievement during 1st grade if the model is correctly specified and selection into different schools were ignorable conditional on prior student achievement and demographic controls. We fit these models by the method of maximum likelihood and then compute BLUPs for the school random effects, which we use as yet another alternative measure for school effectiveness. Results from a parallel analysis of neighborhood effects based on this measure are presented in Table F.5. Consistent with the results we prioritize in the main text, estimates in Table F.5 also provide little evidence that differences in elementary school quality mediate or interact with the effects of neighborhood poverty on academic achievement.

Lastly, we replicate our analysis on data from the Early Childhood Longitudinal Study - Kindergarten Class of 2010 (ECLS-K 2010). The ECLS-K 2010 is very similar in design to the ECLS-K 1998, but it included an additional wave of fall assessments and thus allows for estimating school-year and summer learning rates through the end of 2nd grade. With these data, we compute effect estimates based on the two measures of school effectiveness that we view as most defensible, now averaged over consecutive years to improve their reliability. Specifically, we report neighborhood effect estimates based on, first, the difference between the school-year and summer learning rates averaged over 1st and 2nd grade, and second, the difference between the school-year learning rate and one half of the summer learning rate averaged over 1st and 2nd grade.

We estimate these learning rates from seasonal growth models analogous to those outlined in Part A of the Online Supplement but now modified to account for the additional wave of fall assessments in the ECLS-K 2010. With these data, we model test scores measured at time t for child i in school j, denoted by SCR_{tij}, as follows:

$$SCR_{tij} = (\gamma_0 + \mu_{0j} + \tau_{0ij}) + KND_{tij}(\gamma_1 + \mu_{1j} + \tau_{1ij}) + SUM1_{tij}(\gamma_2 + \mu_{2j} + \tau_{2ij}) + FST_{tij}(\gamma_3 + \mu_{3j} + \tau_{3ij}) + SUM2_{tij}(\gamma_4 + \mu_{4j} + \tau_{4ij}) + SND_{tij}(\gamma_5 + \mu_{5j} + \tau_{5ij}) + \varepsilon_{tij},$$

where there are $t = 1, \dots, 6$ testing occasions between the start of kindergarten and the end of 2nd grade and where KND_{tij}, $SUM1_{tij}$, FST_{tij}, $SUM2_{tij}$, and SND_{tij} respectively denote the amount of time in months that a child had spent in kindergarten, the summer after kindergarten, in 1st grade, the summer after 1st grade, and in 2nd grade prior to each testing occasion. In this model, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)$ is a vector of fixed effects that capture the achievement level and seasonal learning rates across all schools; $\boldsymbol{\mu}_j = (\mu_{0j}, \mu_{1j}, \mu_{2j}, \mu_{3j}, \mu_{4j}, \mu_{5j})$ is a vector of random effects that capture each school's departure from the overall average achievement level and learning rates; and $\boldsymbol{\tau}_{ij} = (\tau_{0ij}, \tau_{1ij}. \tau_{2ij}, \tau_{3ij}, \tau_{4ij}, \tau_{5ij})$ is another vector of random effects that capture each child's deviation from their school's average achievement level and learning rates.

As before, we assume that $\boldsymbol{\mu}_j$ and $\boldsymbol{\tau}_{ij}$ are uncorrelated and that both follow multivariate normal distributions with zero means and unrestricted covariance matrices, and that the disturbance term in this model, ε_{tij}, represents random measurement error, whose variance is constrained to equal the total variance of the test scores multiplied by one minus their reliability. We then fit this model by the method of maximum likelihood after excluding sampled children without valid school identifiers, who attended a year-round school, or who transferred schools.

For each school j, we compute its effectiveness first as $\frac{1}{2}\left[\left(\hat{\gamma}_3 + \tilde{\mu}_{3j}\right) - \left(\hat{\gamma}_2 + \tilde{\mu}_{2j}\right) + \left(\hat{\gamma}_5 + \tilde{\mu}_{5j}\right) - \left(\hat{\gamma}_4 + \tilde{\mu}_{4j}\right)\right]$, and second as $\frac{1}{2}\left[\left(\hat{\gamma}_3 + \tilde{\mu}_{3j}\right) - \left(\hat{\gamma}_2 + \tilde{\mu}_{2j}\right)/2 + \left(\hat{\gamma}_5 + \tilde{\mu}_{5j}\right) - \left(\hat{\gamma}_4 + \tilde{\mu}_{4j}\right)/2\right]$, where "hats" denote MLEs and "tildes" denote BLUPs for the random effects.

Tables F.6 and F.7 present neighborhood effect estimates based on these two measures of school effectiveness from the ECLS-K 2010. They are computed using the random subsample of $n = 6{,}110$ children in $k = 290$ schools at baseline in the ECLS-K 2010 who received both fall and spring assessments during kindergarten, 1st grade, and 2nd grade. When computing these estimates, all covariates and models were constructed to mirror those from our analysis of the ECLS-K 1998 as closely as possible.

Estimates from the ECLS-K 2010 are substantively similar to those reported in the main text. The total effect estimates in Tables F.6 and F.7 suggest that exposure to a disadvantaged neighborhood at the 80th percentile of the treatment distribution, rather than an advantaged neighborhood at the 20th percentile, reduces performance on 3rd grade math and reading assessments by about one-twelfth and by about one-quarter of standard deviation, respectively. Estimates of the direct effects are all comparable in magnitude to the total effects, while estimates of the indirect and interaction effects are all close to zero and nearly all fail to reach conventional thresholds for statistical significance. For reading test scores, estimates of the $RNIE$ are marginally significant, but at less than one-fiftieth of a standard deviation and only about 5% of the estimated total effect (e.g., $\text{proportion mediated} = RNIE/RATE = -0.013/-0.242 = 0.054$), their substantive magnitude is trivial.

To summarize, across all of these different analyses and measurement strategies, there is little evidence that differences in elementary school quality play an important explanatory role in transmitting neighborhood effects on reading and math achievement.

References

Downey, Douglas B., David M. Quinn, and Melissa Alcaraz. 2019. "The Distribution of School Quality: Do Schools Serving Mostly White and High-SES Children Produce the Most Learning?" *Sociology of Education* 92:386-403.

Downey, Douglas B., Paul T. von Hippel, and Melanie Hughes. 2008. "Are 'Failing' Schools Really Failing? Using Seasonal Comparison to Evaluate School Effectiveness." *Sociology of Education* 81:242-270.

McCaffrey, Daniel F., J. R. Lockwood, Daniel Koretz, Thomas A. Louis, and Laura Hamilton. 2004. "Models for Value-added Modeling of Teacher Effects." *Journal of Educational and Behavioral Statistics* 29:67-101.

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

Table F.1. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores based on Measures of School Effectiveness Equal to a School's 1st Grade Learning Rate, ECLS-K Class of 1998-99

	Math Test Scores			Reading Test Scores		
Estimand	Est.	SE	P-value	Est.	SE	P-value
RATE	-0.135	(0.024)	<0.001	-0.154	(0.029)	<0.001
RNDE	-0.135	(0.024)	<0.001	-0.149	(0.028)	<0.001
CDE	-0.135	(0.024)	<0.001	-0.149	(0.028)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.894	0.000	(0.002)	0.886
RNIE	0.000	(0.006)	0.986	-0.004	(0.009)	0.637
RPIE	0.000	(0.006)	1.000	-0.005	(0.010)	0.628
$RINT_{\text{med}}$	0.000	(0.002)	1.000	0.000	(0.003)	0.873

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table F.2. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores based on Measures of School Effectiveness Equal to the Difference between a School's 1st Grade Learning Rate and One-half the Learning Rate of its Students during the Previous Summer, ECLS-K Class of 1998-99

	Math Test Scores			Reading Test Scores		
Estimand	Est.	SE	P-value	Est.	SE	P-value
RATE	-0.135	(0.024)	<0.001	-0.153	(0.029)	<0.001
RNDE	-0.138	(0.024)	<0.001	-0.159	(0.028)	<0.001
CDE	-0.138	(0.024)	<0.001	-0.158	(0.028)	<0.001
$RINT_{\text{ref}}$	0.000	(0.002)	0.860	-0.001	(0.003)	0.868
RNIE	0.002	(0.004)	0.573	0.005	(0.012)	0.658
RPIE	0.002	(0.003)	0.638	0.006	(0.013)	0.651
$RINT_{\text{med}}$	0.001	(0.002)	0.794	0.000	(0.003)	0.923

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table F.3. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores based on Measures of School Effectiveness Equal to the Simple Average of a School's 1st Grade and Summer Learning Rates, ECLS-K Class of 1998-99

	Math Test Scores			Reading Test Scores		
Estimand	Est.	SE	P-value	Est.	SE	P-value
RATE	-0.136	(0.025)	<0.001	-0.159	(0.029)	<0.001
RNDE	-0.133	(0.025)	<0.001	-0.157	(0.029)	<0.001
CDE	-0.133	(0.025)	<0.001	-0.153	(0.029)	<0.001
$RINT_{\text{ref}}$	0.001	(0.003)	0.858	-0.004	(0.004)	0.375
RNIE	-0.003	(0.005)	0.483	-0.002	(0.003)	0.659
RPIE	-0.002	(0.003)	0.568	-0.006	(0.004)	0.201
$RINT_{\text{med}}$	-0.002	(0.004)	0.627	0.004	(0.005)	0.385

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table F.4. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores based on Measures of School Effectiveness Equal to a Weighted Average of a School's 1st Grade and Summer Learning Rates, ECLS-K Class of 1998-99

	Math Test Scores			Reading Test Scores		
Estimand	Est.	SE	P-value	Est.	SE	P-value
RATE	-0.137	(0.024)	<0.001	-0.155	(0.029)	<0.001
RNDE	-0.133	(0.025)	<0.001	-0.149	(0.029)	<0.001
CDE	-0.134	(0.025)	<0.001	-0.148	(0.029)	<0.001
$RINT_{\text{ref}}$	0.001	(0.003)	0.617	-0.001	(0.003)	0.679
RNIE	-0.005	(0.007)	0.474	-0.006	(0.005)	0.274
RPIE	-0.003	(0.005)	0.509	-0.007	(0.006)	0.233
$RINT_{\text{med}}$	-0.002	(0.003)	0.629	0.002	(0.003)	0.595

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table F.5. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores based on Measures of School Effectiveness Equal to Conventional Estimates of School Value Added from Lagged Dependent Variable Models, ECLS-K Class of 1998-99

	Math Test Scores			Reading Test Scores		
Estimand	Est.	SE	P-value	Est.	SE	P-value
RATE	-0.137	(0.025)	<0.001	-0.152	(0.029)	<0.001
RNDE	-0.125	(0.024)	<0.001	-0.158	(0.029)	<0.001
CDE	-0.124	(0.024)	<0.001	-0.158	(0.029)	<0.001
$RINT_{\text{ref}}$	-0.002	(0.003)	0.606	0.000	(0.003)	0.943
RNIE	-0.011	(0.011)	0.322	0.006	(0.012)	0.620
RPIE	-0.014	(0.015)	0.351	0.005	(0.010)	0.617
$RINT_{\text{med}}$	0.003	(0.005)	0.551	0.001	(0.003)	0.790

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 1998, "Direct Child Assessments in waves 1-5," "Student record abstract form (kindergarten and 1st grade)," "School Administrator Questionnaire in First Grade," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013; U.S. Department of Education, Institute of Education Sciences, Common Core of Data, 1999-2000.

Table F.6. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores based on Measures of School Effectiveness Equal to the Difference between the School Year and Summer Learning Rates Averaged over 1st and 2nd Grade, ECLS-K Class of 2010-11

	Math Test Scores			Reading Test Scores		
Estimand	Est.	SE	P-value	Est.	SE	P-value
RATE	-0.085	(0.033)	0.009	-0.242	(0.036)	<0.001
RNDE	-0.082	(0.032)	0.011	-0.228	(0.037)	<0.001
CDE	-0.084	(0.033)	0.010	-0.229	(0.037)	<0.001
$RINT_{\text{ref}}$	0.002	(0.004)	0.668	0.002	(0.004)	0.708
RNIE	-0.003	(0.004)	0.402	-0.015	(0.007)	0.035
RPIE	-0.002	(0.004)	0.607	-0.012	(0.009)	0.176
$RINT_{\text{med}}$	-0.001	(0.003)	0.702	-0.003	(0.008)	0.684

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 2010, “Direct Child Assessments, Waves 1-7,” “School Administrator Questionnaire,” and “Parent Interview, Waves 1 and 2”; GeoLytics Neighborhood Change Database, 2013.

Table F.7. Effects of Neighborhood Context on 3rd Grade Achievement Test Scores based on Measures of School Effectiveness Equal to the Difference between the School Year Learning Rate and One Half of the Summer Learning Rate Averaged over 1st and 2nd Grade, ECLS-K Class of 2010-11

	Math Test Scores			Reading Test Scores		
Estimand	Est.	SE	P-value	Est.	SE	P-value
RATE	-0.085	(0.032)	0.009	-0.242	(0.036)	<0.001
RNDE	-0.080	(0.032)	0.013	-0.229	(0.037)	<0.001
CDE	-0.082	(0.032)	0.011	-0.231	(0.037)	<0.001
$RINT_{\text{ref}}$	0.002	(0.004)	0.555	0.002	(0.004)	0.666
RNIE	-0.005	(0.004)	0.256	-0.013	(0.007)	0.044
RPIE	-0.003	(0.004)	0.516	-0.010	(0.008)	0.187
$RINT_{\text{med}}$	-0.002	(0.004)	0.558	-0.003	(0.007)	0.673

Notes: Estimates are combined across MI datasets. SEs are computed using the cluster bootstrap. P-values come from two-sided Wald tests of the null hypothesis that the focal estimand is equal to zero.

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Early Childhood Longitudinal Survey - Kindergarten Class of 2010, "Direct Child Assessments, Waves 1-7," "School Administrator Questionnaire," and "Parent Interview, Waves 1 and 2"; GeoLytics Neighborhood Change Database, 2013.