The Spillover Effects of Grade-Retained Classmates: Evidence from Urban Elementary Schools

Author(s): Michael A. Gottfried

Source: *American Journal of Education*, Vol. 119, No. 3 (May 2013), pp. 405-444

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/10.1086/669851

The Spillover Effects of Grade-Retained Classmates: Evidence from Urban Elementary Schools

MICHAEL A. GOTTFRIED
Loyola Marymount University

Numerous studies have examined how grade retention can impede the academic success of those retained students. One uncharted line of research is the spillover effect that retained students may exert on their classmates. The purpose of this study is to evaluate this relationship for students in urban elementary school classrooms. To do so, this study analyzes a longitudinal data set comprising entire populations of five elementary school cohorts within the School District of Philadelphia. Because individual student records can be linked to teacher and classroom data as well as to school, grade, and year identifiers, this study employs a series of multilevel fixed-effects models to address estimation issues regarding omitted variable bias. All results indicate that the effects of having a greater number of grade-retained peers are detrimental to the standardized achievement outcomes of nonretained classmates. Data-driven policy implications are discussed.

About a half century of research and policy making on academic peer effects has rested on the fundamental premise that exposure to stronger peers can improve outcomes of their classmates. It was the dissemination of the seminal *Equality of Educational Opportunity* report, or the "Coleman Report" (Coleman et al. 1966), that sparked this interest in how peers can influence the outcomes of other students. In the Coleman Report, the authors found academic advantages for economically disadvantaged black students attending schools with middle-class students. Since then, most research has also focused on peer effects primarily by assessing direct measures of academic ability (e.g., Henderson et al. 1978; Summers and Wolfe 1977; Zimmer and Toma 2000) or socioeconomic status (e.g., Caldas and Bankston 1997; Link and Mulligan 1991; Willms 1986). Recent attention, however, has turned toward evaluating academic peer effects through additional channels, such as having classmates who are fre-

Electronically published March 12, 2013
American Journal of Education 119 (May 2013)

0195-6744/2013/11903-0004$10.00

quently absent (e.g., Gottfried 2011), having classmates with special needs (e.g., Fletcher 2010), or having classmates with behavior problems (Figlio 2007). In the direction of this recent research agenda to determine new ways in which peers can influence the academic outcomes of their classmates, this study evaluates the effect of having classroom peers who have been grade retained.

A significant body of research has focused on the outcomes for grade-retained students. Despite the continuation of the implementation of grade-retention practices in schools, most research concludes that retention is ineffective as an academic policy and is hence not supported as a school practice (Alexander et al. 1997; Holmes 1989; Jackson 1975; Jimerson 1999, 2001a, 2001b; Jimerson and Kaufman 2003; Karweit 1999). Although a few recent studies may have shown some benefits of retention (e.g., Hughes et al. 2010), particularly for smaller subsets of students (e.g., Greene and Winters 2007; Lorence and Dworkin 2006), research on the whole finds a lack of any systematic benefits of grade retention. This is true academically (Jimerson et al. 2002; Reynolds 1992) and developmentally (Anderson et al. 2005; Morrison et al. 1997; Shepard and Smith 1990).

Academically, empirical research widely upholds that grade-retained students have lower achievement levels than if they had been continuously promoted (Alexander et al. 2003; McCoy and Reynolds 1999; Meisels and Liaw 1993; Reynolds 1992). Indeed, because of advances in methodology, recent research has supported the conclusion that the negative relationship between being grade retained and subsequent lower achievement is causal (Hong and Raudenbush 2005; Roderick and Nagaoka 2005). In this way, grade retention is seen as directly detrimental to student success by placing retained students even lower on the distribution of academic performance in postretained years. Furthermore, Pagani et al. (2001) found the effects of retention to have both short- and long-term ramifications on achievement outcomes. In fact, early grade retention has been supported as one of the most powerful predictors of becoming a high school dropout (Jimerson et al. 2002), thereby extending the negative academic risk outcomes beyond measures of achievement and across time.

Developmentally, the field also widely upholds the finding that being retained has socioemotional consequences, compared to if the same students were continuously promoted (Holmes and Matthews 1984; Jimerson 2001a, 2001b; Shepard and Smith 1990; see Hong and Yu [2008] for an exception). For instance, Mantzicopoulos and Morrison (1992) found an increase in at-

---

Michael A. Gottfried, PhD, is an assistant professor in the School of Education at Loyola Marymount University. He studies the economics of education, urban education, and K–12 policy.

tention problems in the repeated year of kindergarten for those who had been retained versus a control group of kindergartners who had been promoted to first grade. Retained students have also been shown to have lower self-esteem and higher rates of school absences relative to their continuously promoted counterparts (Jackson 1975; Jimerson 2001a, 2001b; Shepard 1989). In addition, Pagani et al. (2001) found that those students who had been grade retained tended to have higher cases of anxiety, inattentiveness, and classroom disruption. Indeed, these effects had both short- and long-term consequences, and these outcomes were exacerbated for boys, although significantly present in both genders.

Hence, there appear to be multiple mechanisms by which grade-retained students can influence the outcomes of other members of the classroom. First, if being grade retained leads to worsened academic performance compared to having been continuously promoted, then grade-retained students may negatively affect other peers through having lower achievement levels. That is, if teachers respond to the particular educational needs of grade-retained students by allocating regular class time, then nonretained students may potentially be adversely affected as classroom instruction is slowed for everyone. Increasingly large numbers of grade-retained peers in the classroom would suggest that even greater portions of instruction would be dedicated to remediation, thereby further slowing educational advancement for other classmates.

Second, as previously described, research has suggested an increase in adverse socioemotional consequences from being grade retained; thus, there may be spillover effects in the classroom through the increased disruption, inattentiveness, and truancy of grade-retained students. For instance, as mentioned, grade retention may lead to a higher propensity to miss school. High absence rates have been shown to cause students to feel a greater sense of alienation from their classmates, teachers, and schools and may thus disrupt classroom instruction through their negative interactions and social disengagement (Ekstrom et al. 1986; Finn 1989; Johnson 2005; Newmann 1981). Therefore, if being retained generates negative social outcomes, such as inattentiveness and truancy, and if these behaviors in turn produce further problems in school, such as alienation, then retained students can generate multiple noninstructional, or behavioral, disruptions in classrooms. This can also affect the learning process for nonretained peers, as teachers again must divert their instructional resources elsewhere rather than to teaching.

Lazear (2001) theorized about this classroom spillover mechanism by proposing that instruction in the classroom can be characterized as a public good, in which some students may exert congestion effects onto the teacher's instructional time. As such, there are negative effects on achievement generated when one student's actions impede the learning for other classmates. According to this model, a grade-retained student, who diminishes regular instruction as a result of worsened academic or behavioral outcomes from having been retained, utilizes teaching time in ways that nonretained students may not find useful. In essence, retention produces both an individual effect through a decrease in retained students' own learning outcomes, as previous research has documented, and also a peer effect by compromising instruction and thereby adversely affecting the educational outcomes for other, nonretained students in the class.

Although this is the first empirical study to examine the effects of retained students on their classmates, research on mainstreaming practices of special education students and English language learners (ELLs) also provides insight into the way in which students with unique learning needs may negatively influence other, more "typical" members of the same classroom. Students with special needs have been documented as exhibiting higher incidences of behavioral issues than their classmates (Daniel and King 1997; Morgan-D'Atrio et al. 1996), are suspended at twice the rates of their classmates (Ergenbright 2010), and take a disproportionate amount of teachers' time and attention with respect to classroom management (Downing et al. 1997; Greene et al. 2002). Under these circumstances, students with special needs can negatively affect the outcomes of classmates without disabilities through direct mechanisms (such as by inducing disorderly behaviors from their classmates through their own disruptive actions) or through indirect mechanisms (such as by redirecting teachers' attention to managing students with special needs, thereby leaving teachers with less time to foster the achievement of other classmates; e.g., Lazear 2001).

Given the unique learning needs of ELL students, it has also been documented that ELL students may slow the progression of the entire classroom (Schmidt 2000). In doing so, they often utilize a disproportionate amount of a teacher's time and attention, thereby skewing the teacher's distribution of classroom management (Hayworth 2009; Karabenick and Noda 2004). Under these circumstances, ELL students might negatively affect the outcomes by inducing disengagement from their classmates through a disruption of regular-paced classroom instruction (Karabenick and Noda 2004). Additionally, ELL students may be redirecting instruction to themselves, thereby leaving teachers with less time to foster the achievement of others (Schmidt 2000).

Indeed, quantitative research in both special education and ELL has supported the presence of these negative mechanisms. As for special education, Friesen et al. (2010) and Hanushek et al. (2002) examine the spillover effects of students with special needs on achievement outcomes of their classmates. Both studies focus their analyses on the effects of grade mates rather than the effects of classmates. Nonetheless, what their studies do suggest is that there are significant effects on the achievement outcomes of classmates sharing educational settings with students with special needs. Fletcher (2009, 2010) focuses more precisely on the peer effects of having classmates (as opposed to grade mates) with disabilities. The author finds in both studies that having peers with special needs decreases reading and math achievement for classmates without special needs in elementary school. As for ELL students, Cho (2012) focuses on the academic achievement effects of having ELL classmates. The author finds that having ELL peers decreases reading (but not math) achievement during both kindergarten and first grade. Several contextual factors moderate these effects, including classroom ability grouping, gender, and family income.

Hence, the case for assessing the effects of students with unique learning needs on the outcomes of other classmates has been well established in other areas. Given the individual risk factors associated with having been grade retained and the purported mechanisms by which grade-retained students may exert an influence on their nonretained classmates, this study utilizes quasi-experimental methods to assess the achievement effects that the number of grade-retained students has on other students in the classroom. Although numerous studies have evaluated the effects that grade retention has on those retained students per se, the literature has been quite limited on discerning the classroom-level effects of grade retention on nonretained classmates' achievement. That is, the extant body of research has not fully considered a spillover effect of retention.

Lavy et al. (2008) conducted research that comes closest to evaluating this effect. They examined the proportion of delayed peers in a cohort in Israel (i.e., peers who were older than would be expected for their cohort) and found a negative effect on student performance. The authors did use the term "repeaters" to describe these older students but admitted that those students may not have been grade retained, as the majority of the repeaters had simply delayed entry into first grade. Their findings, then, might instead approximate the effect of having older students in a classroom but might be more effectively interpreted as the effect of age-related peer effects rather than pure retention effects. What the results do suggest, however, is that a peer effect may arise from being exposed to classmates of differing ability levels, which is corroborated by recent research in the field on peer achievement (e.g., Lefgren 2004; Neidell and Waldfogel 2008; Zabel 2008; Zimmer and Toma 2000).

Given the research on both the individual effects of being retained and the potential peer effects, this study makes several meaningful contributions through three key research questions.

1. *Are there peer effects of grade-retained classmates on the achievement outcomes of nonretained students in the same classroom? Are these results generalizable across multiple measures of achievement?*

Prior studies have examined the relationship between having been retained and postretention achievement and developmental outcomes. However, none has evaluated these issues in terms of the spillover effect that retention may have on other students in the classroom, specifically in terms of standardized testing achievement. Thus, this study investigates this issue by assessing if a relationship exists between exposure to grade-retained peers in the classroom and the achievement outcomes for those nonretained students in the same room. Given the widespread findings of a negative effect of grade retention on the retained student, it is hypothesized in this study that there will be a negative spillover effect on the achievement outcomes for other students in the same classroom.

2. *How do gender, race, and socioeconomic status moderate the spillover effect of grade-retained classmates?*

As previously mentioned, Pagani et al. (2001) found differential results of retention based on gender. Hence, this study examines if the peer effect of grade-retained students is also moderated by gender. Additionally, few studies have examined the effects of retention for minority and high-poverty youths, or urban students at all (Reynolds 1992; Willson and Hughes 2006). However, research has consistently upheld the fact that academic and developmental problems are exacerbated for minority youths in urban school systems, particularly when it comes to retention issues (Natriello et al. 1990). Those studies that have researched the effects of retention in urban school districts have focused on the effects pertaining to the retained students themselves (e.g., McCoy and Reynolds 1999; Reynolds 1992; Willson and Hughes 2006). None, however, have examined the effects of retention on nonretained classmates. Thus, this study also contributes to this field by focusing on a population of students for which there is little research on the effects of grade retention. It does so by documenting the classroom contexts for full cohorts of elementary school children as they progress through school in a large, high-poverty urban district. Given past findings of differential effects by gender, race, and socioeconomic status, it is hypothesized in this study that there will be differential effects of grade-retained classmates by demographic characteristics.

Finally, given that this is an empirical study, there is a question of methodology. Many studies in the retention literature have methodological limitations, such as small sample sizes, limited generalizability, a lack of statistical controls for various student and classroom inputs, or a complete lack of classroom identification information (thereby making it a difficult if not impossible task to identify actual classmates). In this regard, it is crucial to address the following research question:

3. *Are the findings robust to multiple methodological approaches?*

The data set employed in this study is longitudinal and nonselective and contains classroom identification information for each student in every academic year. With these data, it is possible to link students to classrooms, teachers, and schools as well as other covariates, such as demographic information and neighborhood characteristics. Therefore, having these comprehensive, multilevel, and longitudinal data from elementary school students in a large urban district allows for the precise identification of classroom peers for every student in every academic year. Hence, it is possible to fully utilize quasi-experimental variation in classroom peer groupings over time to address issues pertaining to the nonrandom assignment of classroom peers. This study does so by using multiple levels of fixed effects and controlling for unobservable influences that may otherwise be confounding the estimates of having grade-retained classmates. Given that these more stringent approaches account for unobservable factors, it is hypothesized that the effect of grade-retained peers will be tempered with the implementation of increasingly rigorous methods.

With these guiding questions, this study contributes both to the research on peer effects and to the research on grade retention by evaluating the classroom effect of having a larger number of grade-retained classroom peers on achievement outcomes in elementary school. Moreover, research has suggested that retention disproportionately affects racial minority students and those living in poverty (Alexander et al. 2003; Corman 2003; House 1999). Thus, it is the youths in America's urban schools who face the highest risk of failure from retention policies, which may in turn have ramifications for other students in the classroom who might have already been at risk for educational decline. The benefits of relying on the findings in this study are thus quite clear: by identifying significant factors in the schooling experiences of urban elementary students, it is possible to develop policy and support interventions for students early in school, before they enter into secondary education, where the risks of educational failure and developmental decline drastically worsen.

Method

Data

Compilation of the data set.—To effectively evaluate the spillover effects of grade-retained students on the achievement outcomes of their nonretained classmates, this study compiled a unique data set of students, residential neighborhoods, teachers, and classroom observations. Student and teacher data were obtained from the School District of Philadelphia via the district's Office of Student Records and through the district's Personnel Office. Neighborhood data were obtained from the 2000 Census flat files at the census block level. Information was collected on students' home addresses, including street number, name, and zip code. The merging of neighborhood data with the student-level database was achieved by geocoding each address to its longitude and latitude and by assigning each student to a census block group.

Overall, the data were available for five cohorts of students. The first three cohorts were first observed as the kindergarten, first-, and second-grade classes in the 1994–95 academic year. The fourth cohort was composed of the kindergarten classes of 1995–96, and the fifth cohort was composed of the kindergarten classes of the 1996–97 school year. Each cohort was then observed through the end of the 1999–2000 school year. Inclusive of both reading and math standardized tests, the analytical sample consists of all nonretained student observations for a total of $N = 26{,}936$ observations within all 175 public, neighborhood schools that contain elementary grades (either K–5 or K–8).[1]

The analytical sample is restricted to nonretained students in order to evaluate the spillover effect of grade-retained peers. Additionally, the sample is restricted to third- and fourth-grade observations because students were included in the analyses only if data exist on their current and lagged standardized achievement tests, in reading or math or both. Since students in this data set have standardized testing information for second, third, and fourth grades, only third- and fourth-grade observations could be used in the value-added (lagged model) specifications described in the next section. Furthermore, in order to be included in the sample, data also had to exist for other measures, including gender, race, academic indicators (some include lagged information), classroom and teacher characteristics, assignment information (room, grade, and school identifiers), and neighborhood information.[2]

Table 1 provides details on the dependent and independent variables employed in this study. The dependent variables are the normal curve equivalent scores (NCE) for the Stanford Achievement Test (SAT9). The NCEs are the generally preferred measurement for methodological reasons, as they have statistical properties that allow for evaluating achievement over time (Balfanz and Byrnes 2006). NCEs range in value from 1 to 99.

Student data.—For every student in a given academic year, the data set contains demographic information concerning personal characteristics, including gender and race. Additional student identifiers include special education status, ELL free lunch status,[3] and whether or not the student has any behavioral issues, as determined by his or her behavior grade from the previous academic year.

Data at the student level of analysis also contain residential neighborhood information. The empirical model contains four attributes that describe the census block on which the student resides, including the percentage of a student's census block that is white, the percentage of a student's block at or below poverty, the household vacancy rate for the block, and the block's median household income. Note that in the absence of other direct measures

TABLE 1

Descriptive Statistics of the Nonretained Sample

	Mean	SD
Output: academic achievement:		
SAT9 reading score	42.89	14.91
SAT9 math score	57.34	19.26
Key variable:		
Number of retained classmates	2.58	2.08
Student inputs: demographics:		
Male	.52	.50
Black	.68	.47
Latino	.10	.30
Asian	.04	.21
Other	.00	.04
Student inputs: academic indicators:		
1-year lagged SAT9 reading score	39.02	15.45
1 year lagged SAT9 math score	57.82	18.68
Special education	.03	.17
English language learner	.03	.18
Free lunch	.50	.50
Behavior problem	.07	.25
Days absent	11.81	11.20
Student inputs: census block characteristics:		
Census block white	.30	.33
Census block at/below poverty	.14	.09
Census block vacant	.13	.09
Median block income (in dollars)	28,329.74	12,181.92
Classroom inputs: teacher descriptors:		
Male	.03	.17
Black	.11	.32
Latino	.00	.07
Asian	.00	.03
Master's degree	.31	.17
Classroom inputs: nonabsence variables:		
Class size	28.78	3.09
Mean lagged reading score	24.87	10.42
Mean lagged math score	37.00	12.72
N	26,936	

of family data, free lunch status and neighborhood information often serve in empirical models as proxies for family background (Hanushek et al. 2003), as they are based on direct observation of family and neighborhood characteristics (e.g., household and census block incomes).

Teacher data.—Table 1 also presents teacher descriptive information. Data on teachers are sourced both from student records and from the district's Personnel Office. A student record provides the name of the teacher assigned to a student's classroom in a given academic year. In addition, a detailed teacher data set was obtained from the district's Personnel Office. From these, two sets of variables were incorporated into the data set. First, for each teacher, basic characteristics include race and gender. Second, a binary variable indicates whether a teacher had a master's degree, based on the record that provides detail on which graduate school the teacher had attended.

Classroom and peer data.—Students can be grouped unambiguously into classrooms because school and classroom assignment information is included in the student database. In contrast, the teacher data set does not include school or classroom assignment. Teachers are matched to their classrooms by identifying their names, as they appear on personnel records, to the teachers' names as they appear in the student data set.[4] The name that appears on the report card is not always the full name of the teacher, and thus the matching algorithm is conservative in requiring that teacher first and last names be present to have a successful student-classroom-teacher match.

Once students have been matched to teachers and classrooms, it is possible to construct classroom-level measures. The key variable in this study is the peer effect of retained classmates, that is, the number of grade-retained students in the classroom. Constructing this measure began first by identifying those individual students in the room who had been retained at some point in grades K–3 (if the room was a third-grade classroom) or grades K–4 (if the room was a fourth-grade classroom). If a student had the same grade level on his or her record twice (or more),[5] the indicator for retention was coded as a 1 in the current year, indicating that the student had been retained, either in the current grade or in a previous grade.[6] On the other hand, a value of 0 for this covariate indicates that a student was promoted continuously throughout his or her tenure over the sample time period. The analytical sample consists only of those students who have a 0 for retention, thereby composing the nonretained classmate sample on which the effect of retained peers was tested.

Individual-level retention data can be aggregated up to classroom-level covariates, thereby creating the measure of the peer effect of grade-retained students in this study. This measure is constructed as classroom head counts, that is, the number of students in the classroom who have been grade retained at some point prior to the current school year. Note that the data show that there is very little within-grade, within-school variation on class size, and thus constructing this measure as a percentage would yield no additional information, particularly given that the empirical model will control for grade level.[7]

As a first step to empirically account for the possibility that large numbers of retained students may be placed in specially sized or academically tracked rooms, both class size and peer ability (constructed as the mean classroom lagged test scores in reading or math depending on the regression outcome) are included in the analyses. More rigorous ways of accounting for the placement of students is discussed in the methods section. The average class size is approximately 28 students. Mean lagged test scores are based on the 1-year lagged testing outcomes for students in the classroom. Hence, student *i*'s lagged test outcome is not included in the average class score. Thus, each student will experience a slightly different average class score in the data. In an ancillary analysis, a squared term of class mean ability was tested to capture nonlinear in addition to linear effects of classroom test performance. However, the results were not significantly different with the inclusion of a nonlinear term.

Table 2 presents the correlations and their significances between classroom counts of students who had been retained and other classroom covariates. The bottom portion of the table suggests that in the sample of classrooms there is a low correlation between class size and number of retained students, or mean class testing ability and number of retained students. Thus, classrooms with higher head counts of retained students do not appear to be systematically related to particular observable characteristics of classrooms, whether large or small, high achieving or otherwise.[8]

The results from table 2 also demonstrate very small correlations between other characteristics and the number of retained students in a classroom. For instance, there is almost no correlation between the gender and race of the nonretained sample of students and the number of retained students in their classrooms. Similarly sized low correlation coefficients of approximately 0 are found throughout the table, indicating that the academic, residential neighborhood, and teacher characteristics do not systematically relate to the number of retained peers that nonretained students have in their classrooms.

Analytical Approach

Constructing a baseline model.—To examine the educational spillover effects of retained classmates in urban elementary school classrooms, this study employs the education production function developed by education economists and sociologists. To produce an academic outcome, it is possible to utilize an education production function to model the relationship between schooling inputs and output measures of achievement. Building on the conceptual foundations provided by Coleman et al. (1966) and the empirical research developed by Hanushek (1979), Henderson et al. (1978), and Summers and Wolfe (1977), the particular model selected in this study evaluates the outcome of standardized test scores in both reading and math as a function of various

TABLE 2

Correlations between Inputs and the Number of Retained Peers in the Classroom for Nonretained Students

	Correlation
Student inputs: demographics:	
Male	.01
White	−.01
Black	.02*
Latino	.05***
Asian	.00
Student inputs: academic indicators:	
Current year's SAT9 reading score	−.03***
Current year's SAT9 math score	−.02*
1-year lagged SAT9 reading score	.04***
1-year lagged SAT9 math score	−.01
Special education	−.01
English language learner	.01+
Free lunch	.02**
Behavior problem	.00
Attendance	−.03***
Student inputs: census block characteristics:	
Census block white	.00
Census block at/below poverty	.02*
Census block vacant	.00
Median block	−.02*
Classroom inputs: teacher descriptors:	
Male	.00
White	−.03***
Black	−.05***
Latino	−.02*
Asian	−.02*
Master's degree	.00
Classroom inputs: nonabsence variables:	
Class size	.12***
Mean lagged reading score	.00
Mean lagged math score	−.05***

+ $p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

educational inputs relating to students, residential neighborhoods, classrooms, families, schools, grades, and years.

Specifically, the analysis of the education production function begins with a contemporaneous model of achievement. That is, at any time period t, an educational output, such as test scores, is a function of the cumulative, concurrent influences of individual student characteristics including ability and demographic characteristics, classroom environments, family factors, and school resources. Similarly, achievement in all other time periods, such as $t-1$, is a function of the cumulative, concurrent inputs from the same period, $t-1$. Conceptually, the model to be estimated for a student in a given school year t is

$$A_{it} = f(R_{it}, I_{it}, F_{it}, N_{it}, C_{it}, T_{it}, S_{it}), \tag{1}$$

where A is a standardized measure of achievement at year t for student i, R is the spillover effect of having classroom peers who have been retained in year t, I is a series of individual student characteristics, F is a vector of family background influences, N includes variables relating to the student's residential neighborhood census block on which he or she resides, C is a set of classroom inputs, T are teacher characteristics, and S are school characteristics.

Hypothetically, this model of educational production could be estimated for each school year in a student's educational history. As a result, it is theoretically possible to build a model of achievement in year t that captures both current and previous inputs to education. This model is known as the historical education production function. Incorporating the inputs of past years, however, requires an enormous, if not impossible, amount of data. A specification that mitigates some of the data requirements and is commonly used in the education production function literature considers the change in achievement between two years. That is, by taking the difference between year t's historical model of education production (which contains all inputs to education from the first year of schooling through year t) and year $t-1$'s historical model of education production (which contains all inputs to education from the first year of schooling through year $t-1$), all that remains in the model is current achievement, a 1-year lagged measure of prior achievement, and current inputs to education. This is known as the value-added specification (Hanushek et al. 2003; Hanushek et al. 2002) and is expressed as follows:

$$A_{it} = f(R_{it}, I_{it}, F_{it}, N_{it}, C_{it}, T_{it}, S_{it}, A_{i(t-1)}). \tag{2}$$

In this formulation, the inputs are limited strictly to measures from year $t-1$ through year t. As in the contemporaneous model of education production, educational output is represented by a student's standardized test performance in reading or math in the current time period t. Also, the value-added specification contains the same vector of inputs for contemporaneous measures of peer effects, individual student characteristics, classroom environments, family factors, and school resources. The distinction of this value-added model, then, arises from the fact that in this specification, the effects of all prior inputs are assumed to be captured in the lagged achievement measure from period $t-1$. As a consequence, biases created by omitted variables in the past result in biases in the estimated coefficient only on lagged achievement. Thus, current achievement is not confounded with omitted characteristics that persisted in prior periods of schooling (Hanushek et al. 2003).

From this conceptual model of educational attainment, the study utilizes a linear empirical specification of the education production function:

$$\begin{aligned} A_{ijgkt} = {} & \beta_0 + \beta_1 R_{ijgkt} + \beta_2 I_{it} + \beta_3 F_{it} + \beta_4 N_{it} + \beta_5 C_{jgkt} + \beta_6 T_{jgkt} \\ & + \beta_7 A_{ijgk(t-1)} + \varepsilon_{igjkt}, \end{aligned} \tag{3}$$

where A is standardized test performance for nonretained student i in classroom j in grade g in school k in year t as the dependent variable on the left-hand side of the equation and in year $t-1$ as the lagged measure of achievement on the right-hand side of the equation. In this linear model, the 1-year lagged achievement score can be assumed to proxy for individual fixed effects (Hanushek et al. 2002). For example, if unobserved student ability is intransient over time, then this model accurately estimates the relationship between inputs and achievement.

Empirically, the sets of independent variables, described by the education production function, would be estimated as follows. As a vector of key independent variables, R is the spillover effect of retained peers (measured by the number of retained classmates) that nonretained student i experiences in classroom j in grade g in school k in year t. At the student level, other sets of independent variables include I, a vector of a student's characteristics in year t; F, a function of family inputs for student i in time period t; and N, student residential neighborhood census block characteristics for student i in time period t. At the classroom level, the model assigns the following inputs: C are classroom-specific characteristics (e.g., class size and peer ability) for classroom j in grade g in school k in time t, and T are teacher effects (e.g., gender, race, and degree) in classroom j in grade g in school k in time period t.

Finally, the error term ε includes all unobserved determinants of achievement. Empirically, this component is estimated with Huber/White/sandwich robust standard errors (Huber 1967; White 1980), adjusted for classroom clustering. It is in this error term that the multilevel structure of the data is taken into account. Because students are nested in schools by classroom and hence share common but unobservable characteristics and experiences, clustering student data by classroom provides for a corrected error term given this nonindependence of individual-level observations. As a result, all coefficient estimates become more robust, as they have been corrected for this multilevel nature of the data (Primo et al. 2007).[9]

Accounting for unobserved heterogeneity.—One issue that arises with the empirical specification as described thus far is that there may be unobserved school factors that are correlated with both the number of retained peers in a classroom and achievement. For example, year after year schools may attract particularly effective teachers, and thus the school might have less of a need to retain students; however, this would also imply that students, on average, might perform better on tests from having more effective teachers. Through attracting effective teachers, the unobserved covariates of this school environment would be related to both the number of retained peers and the outcome. Alternatively, more involved parents may choose to send their children to schools where the probability of being placed in a room with retained students is extremely small. However, a school's higher aggregate level of parental involvement may also be influencing student achievement. As a third example, schools that have higher standards for academic achievement might be more likely to retain children in a grade, but these schools might also be making other investments in the school environment to enhance student achievement. In this latter case, one might underestimate the negative influence of grade-retained classmates. Indeed, these underlying factors are unobserved to the researcher, even with the large set of control variables included in the linear model. In these three cases, the estimation of the spillover effects of retained peers on achievement outcomes would be biased.

As a result, a second specification in this study includes school-level fixed effects:

$$\begin{aligned} A_{ijgkt} = {} & \beta_0 + \beta_1 R_{ijgkt} + \beta_2 I_{it} + \beta_3 F_{it} + \beta_4 N_{it} + \beta_5 C_{jgkt} + \beta_6 T_{jgkt} \\ & + \beta_7 A_{ijgk(t-1)} + \delta_k + \varepsilon_{igjkt}, \end{aligned} \tag{4}$$

where δ_k are school fixed effects. Technically, the term δ_k is a set of binary variables that indicates if a student had attended a particular school (for each school variable in the data set, 1 indicates yes and 0 indicates no). This set of school indicator variables leaves out one school as the reference group (this process is analogous to creating indicator variables for race, where one racial category is left out as the reference group).

School fixed effects δ_k control for the influences of schools by capturing unobserved systematic differences across each unique school. By, in essence, holding constant those time-invariant school-specific characteristics, such as curriculum, school neighborhood, educational investments, organization, hiring practices, aggregate parental involvement, and retention policies, the model is accounting for school-level variance, and hence, the principal source of variation used to identify the peer effect occurs across classrooms within each school. In other words, by controlling for educational effectiveness and policies at the school level and by implementing classroom-level clustering, school fixed effects allow for a focus on within-school and between-classroom differences.

The approach of school fixed effects is compelling in this analysis. Because the data in this study are multilevel, there is within-school variation at the classroom level in those measures on the number of retained peers. Unlike this study, many empirical models have had to rely on grade or school averages, and thus peer variables would lack within-school variation (i.e., schoolwide average retention rate). As a result, it would not be possible to assess the within-school spillover effect of retained students, as there would be no variation when variables are aggregated. Analytically, under these circumstances the fixed effects would absorb variation in classroom variables, since they would be the same for every student in the school. However, given the multilevel structure of the data set employed in this study, it is possible to estimate the spillover effect of retained peers for each classroom within a school.

Although school fixed effects do account for time-invariant school-level characteristics, there may also be variation by grade. For instance, the retention rate is higher in grade 4 than in grade 3 because it is measured cumulatively. Thus, the rate may be picking up some (of any) difference across grades. Additionally, there may be unobserved policies at a given school that are specific to a grade of students that affect both retention and achievement. For example, highly effective teachers with excellent classroom management skills may be assigned to fourth grade versus third grade at a particular school for unobserved reasons. More effective teachers in a given grade might diminish the need for retention in that grade (i.e., lowering the number of grade-retained classmates); these same teachers, however, may also improve achievement outcomes for that same set of students. Additionally, school administrators may have grade-specific standards for achievement and thus might be more likely to retain children in a particular grade, but these schools might also be making other educational investments in that same grade to boost achievement outcomes.

These scenarios might influence the estimate of the number of retained peers and achievement outcomes. Hence, a second modification considers school-by-grade fixed effects in order to account for within-school, within-grade variation:

$$\begin{aligned} A_{ijgkt} = {} & \beta_0 + \beta_1 R_{ijgkt} + \beta_2 I_{it} + \beta_3 F_{it} + \beta_4 N_{it} + \beta_5 C_{jgkt} + \beta_6 T_{jgkt} \\ & + \beta_7 A_{ijgk(t-1)} + \delta_{kg} + \varepsilon_{igjkt}, \end{aligned} \tag{5}$$

where δ_{kg} represents school-by-grade fixed effects. In more detail, δ_{kg} is a set of binary variables for each school-grade combination, identifying if a student was in a given grade in a particular school, such as grade 3 in school 228 (for each school-grade combination, 1 indicates yes and 0 indicates no). This set of school-grade indicator variables leaves out one school-grade category as the reference group.

The rationale behind this particular model, then, is that school-by-grade fixed effects account for systematic grade-by-grade differences in a particular school that may affect the number of grade-retained peers as well as achievement, such as grade-specific academic policies or teacher assignment policies, as described previously. With school-by-grade fixed effects, any pattern that is unique to a particular institution for a given grade will be held constant in the model specification, and estimates are identified solely on within-school, across-classroom variance for each grade.

Finally, even with the use of a proxy for individual fixed effects (a true measure of individual fixed effects is employed as a test of robustness) as well as school and school-by-grade fixed effects, the models thus far have been constructed under the assumption that unobserved factors are time invariant. However, there is the possibility that time-varying unobserved factors may be influencing schools and grades and hence the number of retained students as well as the achievement of nonretained classmates. For example, if there were a decline in school quality over time, a school might be attracting decreasingly effective teachers (i.e., with a poorer set of classroom management skills and instructional capability) year after year, which would potentially bias the number of retained students in the room as well as the achievement of nonretained students in the same class. To account for such differences over time, a final modification includes school-by-grade-by-year fixed effects:

$$\begin{aligned} A_{ijgkt} = {} & \beta_0 + \beta_1 R_{ijgkt} + \beta_2 I_{it} + \beta_3 F_{it} + \beta_4 N_{it} + \beta_5 C_{jgkt} + \beta_6 T_{jgkt} \\ & + \beta_7 A_{ijgk(t-1)} + \delta_{kgt} + \varepsilon_{igjkt}, \end{aligned} \tag{6}$$

where the term δ_{kgt} represents school-by-grade-by-year fixed effects. In the application of this model to this study's data, δ_{kgt} is a set of binary variables for each school-grade-year combination that indicates if a student was in a given grade in a particular school in a given academic year, such as being in grade 3 in school 228 in the academic year 2000–2001 (for each school-grade-year combination, 1 indicates yes and 0 indicates no). As consistent with the other fixed-effects models in this study, this set of school-grade-year indicator variables leaves out one school-grade-year category as the reference group. Applying this final equation makes it highly unlikely that variations in the unobserved time-varying within-school, within-grade environment would bias the estimated spillover effects of the number of grade-retained peers.

Results

Baseline and Fixed-Effects Results for Reading

Table 3 provides unstandardized coefficient estimates and Huber/White/sandwich robust standard errors, adjusted for classroom clustering for specifications examining the spillover effect of grade-retained peers on the standardized reading achievement outcomes of their nonretained classmates. Recall that the sample includes only those students who have been continuously promoted throughout their schooling, so that it is possible to examine the effect of having retained classmates.

Column 1 presents the results from a baseline model of analysis from equation (3), in which only observable student, residential neighborhood, classroom, and teacher covariates are included. The models become more robust in columns 2–4, in which school, school-by-grade, and school-by-grade-by-year fixed effects are incorporated, respectively.

This process, by which more complex models build on previous models, demonstrates several noteworthy points initially. First, the table suggests that a model containing only observable characteristics explains almost half of the variance of current reading achievement—approximately 50%. This is logical, as the lagged achievement measure included in the model is assumed to soak up all historical information about the student not pertaining to the current school year (Hanushek et al. 2002). Second, the inclusion of school, then school-by-grade, and then school-by-grade-by-year fixed effects improves the explained portion of the variance of reading achievement. The greatest improvement is derived in moving from column 1 to column 2 and subsequently moving from column 2 to column 3: nonetheless, each moderation to the model has improved its ability to explain the variance in achievement. Third, the measures of the classroom counts of retained students are consistently statistically significant ($p < .001$), regardless of which model is examined in the table. Hence, the inclusion of more complex fixed-effects models does not veer from supporting the first hypothesis in this study—that having retained peers decreases the standardized testing performance for nonretained classmates. Moreover, each model consistently suggests that there is a negative spillover effect of being in classrooms with a greater number of previously retained classmates.

Assessing in more detail the coefficients pertaining to the effects of having retained peers (which are located in the top row of table 3) indicates the following results. First, the baseline model suggests that having a larger number of grade-retained classmates is negatively related to reading standardized testing performance, even after controlling for a wide range of observable classroom characteristics, including average peer ability and class size. The measure of effect sizes in this evaluation, as adopted by many education empiricists in nonexperimental studies, is the standardized beta coefficient (e.g., Caldas 1993; Gottfried 2011; Hoxby 2000; McEwan 2003). The relationship between the number of grade-retained peers and reading achievement corresponds to an effect size of approximately -0.09σ. This effect size is consistent with prior quasi-experimental research on classroom peer effects, in which effect sizes of less than 10% of a standard deviation are standard (Ammermueller and Pischke 2006; Hoxby 2000).

Thus, an increase by one standard deviation in the number of classmates who have been grade retained is associated with a statistically significant negative decline in reading performance for nonretained students—one that goes above and beyond student, residential neighborhood, teacher, and classroom attributes and one that goes above and beyond robust models accounting for the multilevel structure of the data (i.e., students within classrooms). It is possible to put this effect into perspective by examining the effect sizes of other coefficients in column 1. Within the framework of this empirical model, the negative effect size of the number of grade-retained classmates (-0.09σ) is slightly larger than the effect of being a special education student (-0.08σ). An alternative perspective is that having two retained students as peers may lower other students' achievement by 9% of a standard deviation. Thus, putting a nonretained student into a room with a one standard deviation higher than average number of retained peers is equivalent to the effect size of being a special education student. In fact, the effect size of the number of grade-retained classmates is more than double the effect size of being a high-poverty student (i.e., free lunch recipient, -0.04σ) and is approximately two-thirds the size of the black-white test score gap (-0.15σ).

The results from employing school, school-by-grade, and school-by-grade-by-year models are presented in columns 2–4, respectively. The coefficient on the number of retained peers increases between the baseline model and the subsequent model employing solely school fixed effects (i.e., between models 1 and 2). Without proceeding any further, a model accounting for only the unobserved school environment (without taking into account additional unobserved levels of data) would have suggested a prior underestimation of the effect of grade-retained peers in the baseline model, as evidenced by the change in the size of the coefficients between models 1 and 2. However, the consistency in the estimates in the subsequent analyses in models 3 and 4 suggests that in addition to relying on school fixed effects, it is necessary to include measures that account for unobserved grade and year factors.

Turning to these two final models in the table, there is an overall reduction in the point estimate when grade fixed effects are incorporated in both column 3 (school-by-grade model) and column 4 (school-by-grade-by-year model).

TABLE 3

Estimates of Spillover Effects of Retained Classmates and Other Inputs on SAT9 Reading for Nonretained Students

	No FE (1)	School FE (2)	School-by-Grade FE (3)	School-by-Grade-by-Year FE (4)
Key variable:				
Number of retained peers	−.55***	−.82***	−.21***	−.23***
	(.07)	(.07)	(.06)	(.06)
Effect size[a]	$-.09\sigma$	$-.13\sigma$	$-.03\sigma$	$-.04\sigma$
Model controls:				
Student inputs: demographics:				
Male	−1.21***	−1.24***	−.96***	−.98***
	(.17)	(.15)	(.14)	(.14)
Black	−4.55***	−3.56***	−2.23***	−2.24***
	(.37)	(.36)	(.33)	(.32)
Latino	−3.29***	−2.78***	−1.35***	−1.43***
	(.45)	(.42)	(.37)	(.39)
Asian	−.06	.52	.81*	.75^{+}
	(.52)	(.42)	(.40)	(.41)
Other	−3.59^{+}	−3.18	−2.25	−2.98
	(2.07)	(2.28)	(2.21)	(2.28)
Student inputs: academic indicators:				
1-year lagged SAT9 reading score	.60***	.60***	.72***	.72***
	(.01)	(.01)	(.01)	(.01)
Special education	−6.57***	−7.38***	−2.34***	−2.40***
	(.70)	(.71)	(.63)	(.66)
English language learner	−3.16***	−2.94***	−.79	−.83
	(.58)	(.63)	(.57)	(.57)
Free lunch	−1.34***	−.76***	−.63***	−.63***
	(.17)	(.15)	(.14)	(.14)
Behavior problem	−3.13***	−2.90***	−1.41***	−1.44***
	(.36)	(.35)	(.31)	(.31)
Days absent	−.08***	−.08***	−.06***	−.06***
	(.01)	(.01)	(.01)	(.01)
Student inputs: census block characteristics:				
Census block white	.84	.45	.70^{+}	.71^{+}
	(.52)	(.44)	(.39)	(.38)
Census block at/below poverty	−1.50	−.50	−.10	−.11
	(1.97)	(1.40)	(1.30)	(1.33)
Census block vacant	−2.27	.96	1.00	1.43
	(1.54)	(1.21)	(1.16)	(1.15)
Median block income	.00**	.00	.00	.00
	(.00)	(.00)	(.00)	(.00)

TABLE 3 (*Continued*)

	No FE (1)	School FE (2)	School-by-Grade FE (3)	School-by-Grade-by-Year FE (4)
Classroom inputs: teacher descriptors:				
Male	−5.45***	−6.15***	−1.77	.10
	(1.14)	(1.28)	(1.23)	(1.07)
Black	−7.73***	−8.39***	-1.23^{+}	−2.24**
	(.74)	(.76)	(.70)	(.78)
Latino	−11.75***	−11.03***	-4.29^{+}	−3.81
	(2.01)	(2.22)	(2.40)	(3.45)
Asian	−11.77**	−14.38***	−2.18	−8.84**
	(3.82)	(2.14)	(3.09)	(.95)
Master's degree	−3.85***	−4.94***	.08	−.14
	(.87)	(1.09)	(.73)	(.72)
Classroom inputs: nonretention variables:				
Class size	−.12*	−.13	−.21**	−.30***
	(.07)	(.08)	(.08)	(.09)
Mean reading score	.11***	.18***	.11***	.09***
	(.01)	(.02)	(.02)	(.03)
N	17,289	17,289	17,289	17,289
R^2	.49	.53	.61	.62

NOTE.—Robust Huber-White standard errors adjusted for clustering within classrooms are in parentheses. All regressions include a constant. FE = fixed effects.

[a]Effect sizes in this study are estimated as the standardized beta coefficient.

$^{+}p < .10$.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Thus, these two most rigorous of empirical models suggest that there are unobservable grade and time effects that were previously influencing the estimates of the number of retained classmates—hence exemplifying the importance of relying jointly on school, grade, and year fixed effects. Not including them led to an upward bias in the estimates in column 1, and the biased upward estimates in column 2 can be explained by not having accounted for the particular unobserved school-grade environment or school-grade-year environment.

In more detail, there appear to be the most significant changes in the transition from column 2 to column 3, in which grade fixed effects are incorporated into the analysis. First, the explained portion of the variance has the greatest change between these two columns, indicated by the change in values of R^2. Also, the largest change in the magnitude of the coefficient estimates of grade-retained peers occurs between columns 2 and 3: the inclusion of year fixed effects in column 4 does not drastically change the estimate of grade-retained classmates from that in column 3. Further, while incorporating year fixed effects in column 4 does improve the explained portion of the variance (and is in fact the highest across all four columns), it has not changed dramatically compared to column 3. Hence, there is a consistency in these increasingly complex models in the latter two models of the table.

Most important, in both columns 3 and 4, the coefficient estimates remain statistically significant ($p < .001$) and similar in magnitude, thereby suggesting that a consistent deleterious effect continues to persist from having classmates who have previously been grade retained even after controlling for unobservable characteristics and for unobserved school, grade, and year factors. The effect sizes in columns 3 and 4, while smaller than in columns 1 and 2, continue to hover in the range of effect sizes as defined by prior literature on classroom peer effects (Ammermueller and Pischke 2006; Hoxby 2000). The negative effect sizes thus continue to portray that a one standard deviation increase in the number of grade-retained classmates has a pervasive negative effect on standardized reading achievement, even under more robust models that, in addition to controlling for a range of observable variables, also account for multiple unobserved levels of nested data.

Briefly examining the control variables in table 3 provides the following results. Black and Latino students tend to have lower reading and math test scores compared to white students (Caldas and Bankston 1997; Ogbu 1989; Summers and Wolfe 1977). Similarly, the coefficients of being a special education student, being a free lunch recipient, or having behavioral issues are negative and consistently statistically significant in all models in the table. Additionally, a larger number of days absent is negatively related to reading achievement, consistent with prior literature (Dryfoos 1990; Finn 1993; Gottfried 2009; Lehr et al. 2003; Stouthamer-Loeber and Loeber 1988).

The table does not indicate significant coefficients relating to student-level neighborhood attributes. The results of student-level neighborhood characteristics may suggest that it is the students' characteristics and school environment that affect achievement levels in the education production function model. As for teacher characteristics, although the teacher effects are strong in models 1 and 2, they have an overall lack of significance with the inclusion of grade fixed effects in column 3. In conjunction with the improved explained portion of the variance in moving from column 2 to column 3, accounting for the unobserved school and grade (and school-by-grade) environment proves to provide a more robust description of these teacher effects. Indeed, the overall lack of significance in the reading achievement models 3 and 4 is consistent with many education production studies, including Hanushek (1986).

As for nonretention classroom characteristics, the results here indicate that class size, as measured by the number of students in a single classroom, may not be a consistently significant factor in student achievement (Henderson et al. 1978). Finally, being in a classroom with a higher average of peer reading achievement is in general a significant, positive predictor of testing performance.

Baseline and Fixed-Effects Results for Math

Because the decline in math performance continues to be a significant issue in US educational policy, this study has differentiated between reading and math achievement models so that it is possible to distinguish between the effects of grade-retained classmates in the two separate subject areas. Doing so is particularly relevant within this study's sample of urban students, as previous research has suggested that the achievement gap in math achievement begins to widen for minority and high-poverty students in urban districts as early as in elementary school (Balfanz and Byrnes 2006).

Table 4 provides a similarly structured analysis for how having classmates who have been grade retained relates to nonretained students' SAT9 math achievement. The overall interpretation of the estimated coefficients on the number of grade-retained peers is consistent with that of reading. Moreover, the sizes of the coefficients on the first two models are smaller than those for reading; however, once school-by-grade and school-by-grade-by-year fixed effects are incorporated in models 3 and 4, the coefficients are larger for math achievement outcomes than for reading achievement. Thus, if it can be assumed that the model incorporating school-by-grade-by-year fixed effects portrays the most robust empirical model, then according to model 4, the negative result from having classmates who have been previously grade retained is larger in magnitude for math in terms of unstandardized coefficients.

Focusing on the final, full model in table 4, compared to reading, the larger unstandardized coefficient and quantitatively similar effect size in math are in line with previous literature on classroom peer effects, as was also the case for reading achievement. Regardless of the academic subject area, however, the interpretation is consistent and clear: a one standard deviation increase in the number of peers who have been retained in a given school year is related to a statistically significant standard deviation decrease in SAT9 math performance.

In addition, the interpretation of student, neighborhood, teacher, and classroom control variables has remained generally similar as well to those in the

TABLE 4

Estimates of Spillover Effects of Retained Classmates and Other Inputs on SAT9 Math for Non-retained Students

	No FE (1)	School FE (2)	School-by-Grade FE (3)	School-by-Grade-by-Year FE (4)
Key variable:				
Number of retained peers	−.30***	−.43***	−.26***	−.30***
	(.07)	(.08)	(.08)	(.08)
Effect size[a]	−.04σ	−.05σ	−.03σ	−.04σ
Model controls:				
Student inputs: demographics:				
Male	−.05	−.06	−.06	−.04
	(.16)	(.14)	(.13)	(.13)
Black	−4.70***	−4.06***	−3.42***	−3.56***
	(.43)	(.34)	(.32)	(.31)
Latino	−1.79***	−2.34***	−1.68***	−1.75***
	(.49)	(.44)	(.39)	(.34)
Asian	1.80***	1.88***	2.06***	1.93***
	(.57)	(.43)	(.42)	(.42)
Other	−.50	−1.67	.59	−1.45
	(2.43)	(2.26)	(2.13)	(2.24)
Student inputs: academic indicators:				
1-year lagged SAT9 math score	.68***	.67***	.70***	.70***
	(.01)	(.01)	(.01)	(.01)
Special education	−2.33**	−2.97***	−1.69*	−1.67*
	(.73)	(.74)	(.72)	(.70)
English language learner	−2.07***	−1.49*	−1.03[+]	−1.16[+]
	(.68)	(.68)	(.61)	(.64)
Free lunch	−1.33***	−1.19***	−1.39	−1.35***
	(.24)	(.21)	(.21)	(.21)
Behavior problem	−2.50***	−2.40***	−2.17***	−2.25***
	(.32)	(.33)	(.34)	(.32)
Days absent	−.11***	−.11***	−.10***	−.10***
	(.01)	(.01)	(.01)	(.01)
Student inputs: census block characteristics:				
Census block white	1.25**	.06	.15	.16
	(.51)	(.53)	(.51)	(.48)
Census block at/below poverty	−1.91	−.54	−.17	.04
	(2.01)	(1.50)	(1.48)	(1.41)
Census block vacant	−1.67	−.49	−.37	−.56
	(1.54)	(1.12)	(1.15)	(1.09)
Median block income	.00	.00	.00	.00
	(.00)	(.00)	(.00)	(.00)

TABLE 4 (*Continued*)

	No FE (1)	School FE (2)	School-by-Grade FE (3)	School-by-Grade-by-Year FE (4)
Classroom inputs: teacher descriptors:				
Male	.27	.11	.42	−.42
	(.89)	(.84)	(.83)	(.73)
Black	−2.37***	−2.28***	−.45	−2.24**
	(.61)	(.65)	(.67)	(.72)
Latino	−3.98+	−2.96	.38	.77
	(2.39)	(2.28)	(2.58)	(2.43)
Asian	7.02*	6.72+	4.52	3.30*
	(2.95)	(4.07)	(3.69)	(1.41)
Master's degree	.50	.28	1.45**	−.18
	(.62)	(.59)	(.51)	(.74)
Classroom inputs: nonretention variables:				
Class size	−.01	−.03	.03	.36
	(.06)	(.07)	(.08)	(.96)
Mean math score	.08***	.05**	.09***	.09***
	(.02)	(.02)	(.02)	(.02)
N	23,436	23,436	23,436	23,436
R^2	.54	.57	.62	.64

NOTE.—Robust Huber-White standard errors adjusted for clustering within classrooms are in parentheses. All regressions include a constant. FE = fixed effects.
[a]Effect sizes in this study are estimated as the standardized beta coefficient.
+$p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

reading models; this yields evidence of the consistency of the empirical model chosen for this study.

Results by Nonretained Student Characteristic

Tables 5 and 6 disaggregate the analyses in tables 3 and 4, respectively, to determine if different demographic groups are differentially affected by being in the same classrooms as grade-retained peers. Each cell represents the coefficient on "the number of grade-retained classmates" from a fully interacted model based on the demographic characteristic indicated in the leftmost column.

TABLE 5

Estimates of Spillover Effects on SAT9 Reading Achievement for Nonretained Students, Delineated by Demographic Characteristic

	No FE (1)	School FE (2)	School-by-Grade FE (3)	School-by-Grade-by-Year FE (4)
Estimates from table 3	−.55***	−.82***	−.21***	−.23***
	(.07)	(.07)	(.06)	(.06)
Male	−.51***	−.78***	−.18***	−.21**
	(.08)	(.09)	(.07)	(.08)
Female	−.59***	−.86***	−.24***	−.27***
	(.07)	(.08)	(.06)	(.07)
White	−.59***	−.67***	−.19***	−.16
	(.13)	(.14)	(.12)	(.13)
Black	−.60***	−.87***	−.20**	−.21***
	(.09)	(.09)	(.06)	(.07)
Latino	−.68***	−.66***	−.19	−.24
	(.15)	(.17)	(.14)	(.17)
Asian	−.76***	−.81***	−.34^{+}	−.38^{+}
	(.18)	(.21)	(.21)	(.22)
Free lunch recipient	−.52***	−.78***	−.22**	−.24***
	(.08)	(.09)	(.07)	(.07)
Not a recipient of free lunch	−.58***	−.86***	−.21***	.22***
	(.08)	(.08)	(.07)	(.08)

NOTE.—Robust Huber-White standard errors adjusted for clustering within classrooms are in parentheses. Each cell represents a separate regression, which includes the same control variables as in tables 3 and 4. FE = fixed effects.
$^{+}p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

For reading achievement, the effects of having grade-retained classmates differ by gender and race—hence providing support for the second hypothesis in this study. Nonretained girls are slightly more affected by having retained classmates than are nonretained boys. Nonretained minority students are slightly more affected than are nonretained white students by having a greater number of retained classmates. Free lunch recipients and students who do not receive free lunch are generally affected in a similar way by having greater numbers of classmates who have been grade retained.

Table 6 suggests that the results for the demographic characteristics of the nonretained students in the room differ on the basis of the testing subject;

TABLE 6

Estimates of Spillover Effects on SAT9 Math Achievement for Nonretained Students, Delineated by Demographic Characteristic

	No FE (1)	School FE (2)	School-by-Grade FE (3)	School-by-Grade-by-Year FE (4)
Estimates from table 4	−.30***	−.43***	−.26***	−.30***
	(.07)	(.08)	(.08)	(.08)
Male	−.24***	−.39***	−.22***	−.25**
	(.09)	(.09)	(.08)	(.10)
Female	−.36***	−.49***	−.32***	−.36***
	−.08	−.09	−.09	−.09
White	−.41***	−.46***	$-.23^{+}$	−.34**
	(.14)	(.13)	(.12)	(.13)
Black	−.19*	−.39***	−.23*	−.27*
	(.09)	(.10)	(.10)	(.11)
Latino	−.51**	$-.33^{+}$	−.22	−.17
	(.18)	(.18)	(.21)	(.23)
Asian	−1.10***	−1.14***	−.65***	−.71**
	(.22)	(.25)	(.24)	(.27)
Free lunch recipient	−.23**	−.38***	−.23**	−.24**
	(.09)	(.09)	(.09)	(.09)
Not a recipient of free lunch	−.37***	−.49***	−.30***	−.36***
	(.08)	(.09)	(.08)	(.09)

NOTE.—Robust Huber-White standard errors adjusted for clustering within classrooms are in parentheses. Each cell represents a separate regression, which includes the same control variables as in tables 3 and 4. FE = fixed effects.

$^{+}p < .10$.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

hence, the distinction between reading and math performance is a crucial one. For math achievement, there is generally a slightly larger difference across all models between the coefficient estimates for nonretained girls and nonretained boys than there was for reading achievement. Nonretained black students are slightly less affected by the number of retained classmates than are nonretained white students in math. Out of all racial and ethnic groups, however, Asian students are the most negatively affected by having a larger number of retained classmates in math achievement. Finally, the differential between free lunch recipients and students who do not receive free lunch is much larger in terms of math achievement than it was for reading achievement.

Test of Robustness

In addition to biases on the estimates that may arise from unobserved school, grade, and time-varying factors, there may also be biases that result from within-school sorting (i.e., various student groupings across classrooms in a given grade). On the one hand, principals might be more likely to assign more capable nonretained students—that is, students who are least likely to experience negative effects of having a classmate who has been retained—into classrooms that have greater numbers of students who have been retained. In this case, a negative effect of having a grade-retained classmate would be biased downward. On the other hand, principals might have policies that place less capable nonretained students with less effective teachers and also with a greater number of grade-retained classmates as a way to sort students by academic ability. In this case, the match of nonretained students (but who have a high probability of academic failure) with less effective teachers and with grade-retained students would upwardly bias the negative spillover effects of grade-retained peers. Given the potential biases in the data, a student fixed-effects model is employed:

$$A_{ijgkt} = \beta_0 + \beta_1 R_{ijgkt} + \beta_2 I_{it} + \beta_3 F_{it} + \beta_4 N_{it} + \beta_5 C_{jgkt} + \beta_6 T_{jgkt} + \beta_7 A_{ijgk(t-1)} + \delta_i + \varepsilon_{igjkt}, \tag{7}$$

where δ_i represent student fixed effects. Analogous to previous models in this study, student fixed effects are binary indicators for each individual student (based on student identification code) in which one student category is omitted as the reference.

In utilizing this student fixed-effects model, the importance of relying on longitudinal data becomes evident. Effectively, this equation is conducting a within-student analysis, where it is possible to measure how changes to going in and out of rooms with various student groupings (including varying numbers of retained peers) are related to changes in achievement.[10] Student fixed effects essentially control for all unobserved student confounders that remain constant over time, and solely time-varying characteristics remain in the equation. Accounting for within-student variation, this eliminates any bias caused by correlation between the regressors and the unobserved influences, such as sorting.

Table 7 presents the results for both reading and math. For the sake of comparison, the first section of the table presents the estimates from the school-by-grade-by-year fixed-effects models from table 3 for reading and table 4 for math. The second section of table 7 provides the estimates of the spillover effects of the number of grade-retained peers based on a model of student

TABLE 7

Test of Robustness

	Reading	Math
Column 4 estimates of the number of grade-retained peers from tables 3 and 4, respectively	−.23***	−.30***
	(.06)	(.08)
Effect size[a]	$-.04\sigma$	$-.04\sigma$
Estimate of number of retained peers using student fixed effects	−.18**	−.31**
	(.07)	(.12)
Effect size[a]	$-.03\sigma$	$-.04\sigma$
Time-varying covariates included in student fixed-effects model?	Y	Y
N	17,289	23,436
R^2	.79	.78

NOTE.—Robust Huber-White standard errors adjusted for clustering within classrooms are in parentheses.

[a]Effect sizes in this study are estimated as the standardized beta coefficient.

$^{+}p < .10$.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

fixed effects. The table suggests that results for these student fixed-effects models are comparable to prior fixed-effects models in the analysis, hence providing additional support for the third hypothesis in this study. Also consistent with prior results, the unstandardized effect for math achievement is larger than that for reading achievement; this indicates that moving into classrooms with a larger number of grade-retained classmates poses a greater educational risk in terms of achievement scores for nonretained students for math in particular.

Overall, the results throughout table 7 suggest an interpretation similar to the one before: that a significant relationship exists between grade-retained classmates and achievement, one that suggests causality as supported by the fixed-effects strategies (Schneider et al. 2007). Though the coefficients are slightly different for the testing outcomes of this student fixed-effects methodology in table 7, the effect sizes are equivalent to those in tables 3 and 4. Hence, the analyses continue to uphold that having a greater number of grade-retained classmates in elementary school has negative implications that are generalizable to multiple performance metrics.

Discussion

This study has yielded novel insight into the achievement ramifications of having grade-retained classmates. Although much of the literature has focused on the effects of retention as it pertains to the outcomes of those who have been retained, this research has brought forth the issue that the effects of this practice extend beyond those grade-retained students: there is also a peer effect, that is, an adverse relationship between the number of retained classmates and the academic outcomes of nonretained students in the same classroom. In researching this area, this study has parsed out a new distinction between classroom composition and subsequent outcomes, as no prior research has examined peer effects specifically as they play out for nonretained students in urban schools. Doing so for elementary school classrooms has not only its methodological advantages (i.e., children are contained in a single classroom throughout the day and year, thereby allowing for a clear identification of the peer group) but also policy implications regarding how to identify the predictors of early schooling risk.

Given the longitudinal sample of elementary school students in the Philadelphia School District, this study implemented three methodological approaches. The first began with the assessment of a baseline specification of student achievement for nonretained students, in which SAT9 reading and math were modeled on observable individual, residential neighborhood, teacher, and classroom characteristics, including peer ability. The coefficients on the number of retained classmates indicated negative, significant relationships between having a higher number of grade-retained classmates and student-level achievement in both reading and math.

As a second approach to evaluating the spillover effects of grade-retained classmates, the study incorporated multilevel fixed effects, including school, school-by-grade, and school-by-grade-by-year models. The intention was to account for unobservable factors that may be influencing both the number of grade-retained classmates and student achievement outcomes. For example, it was hypothesized that schools with consistently more effective teachers might have a smaller need to retain students, and yet the students on average perform better on exams because of these effective teachers. Hence, this would bias the estimates of having grade-retained classmates. The coefficients on the number of grade-retained peers, though increased with the inclusion of only school fixed effects, were reduced in magnitude with the introduction of the school-by-grade and school-by-grade-by-year fixed-effects models. This highlights not only the importance of accounting for the unobserved school environment that may be influencing both the number of grade-retained classmates and achievement but also the significance of accounting for grade- and time-specific variation.

In a final approach, a student fixed-effects strategy was employed in which the estimates of having grade-retained classmates were identified from within-student variation in the number of grade-retained peers a student experiences in different classrooms over time. As explained, doing so allowed this evaluation to account for unobserved school sorting. What remained in the analyses were time-varying counts of grade-retained classmates (in addition to all time-varying control variables) in an attempt to isolate the spillover effect of those retained peers. Using these fixed effects, the analyses have pointed toward evidence of causality (Schneider et al. 2007): controlling for student and neighborhood characteristics and school, grade, and year fixed effects, the point estimates consistently indicated that students with higher numbers of grade-retained classmates have worse educational outcomes.

What truly stands out from this analysis, then, is that consistently, across all three methods employed in this study, the results indicate a pervasive, negative spillover effect on standardized achievement from having a greater number of grade-retained classmates. This effect remained negative and significant, even after controlling for student, residential neighborhood, teacher, and classroom characteristics (including average peer ability and class size) and after modeling for school, year, grade, and student fixed effects.

While providing evidence that the findings are generalizable across multiple measures of achievement, the results also suggest that there are differential outcomes based on reading versus math achievement for the nonretained students in the sample. In general, the results indicate that the negative spillover effects of having a higher number of retained classmates are larger in math in terms of NCE test scores. This result is germane within this study's urban school sample because minority and high-poverty students continue to disproportionately fall behind in math (Balfanz and Byrnes 2006). Moreover, these results are also differentiated by demographic characteristics. Girls tend to be more highly affected by having a greater number of retained peers, particularly in math achievement. Additionally, there is evidence of a differential effect based on minority and poverty status across testing subjects.

Given these results, this study demonstrates that strictly focusing the evaluation of retention on the outcomes of the retained students proves to provide an incomplete understanding of how retention policies permeate student outcomes. On the basis of the new analyses in this study, a more refined picture has been drawn: that there are significant relationships between classroom composition and academic achievement pertaining specifically to how retained students affect their peers.

The practice of retention, then, not only has negative effects on the individual who is retained but also has the potential to diminish academic outcomes for other students, perhaps via a strain in classroom resources. If the classroom instructional environment can be characterized as a public good as Lazear (2001) has suggested, then any deviance from regular instruction caused by one student can be interpreted as a negative externality onto his or her peers. Lower-performing students require more attention and help from their teachers (Everston 1982), while higher-performing classmates are less demanding of teacher time (Brown and Saks 1987; Everston and Emmer 1981). Hence, lower-performing classmates, like retained students, can negatively affect their classmates' achievement because they require additional instructional attention from the teacher. Thus, teachers must divert their attention away from instructing the entire class and instead toward academic remediation and potential behavior discipline. This potential misallocation of a teacher's resources (i.e., a strain on the teacher's instructional time and attention) materializes for other students in the classroom as a decline in academic performance, which has been quantified in this article. As such, peer effects from students who have been retained may have arisen because the classroom environment has diminished in quality.

Hence, there are several data-driven policy implications that arise from this study. First, previous research suggests that grade-retained students experience higher instances of negative academic and social consequences from this school practice (e.g., Jimerson 2001a, 2001b). This study contributes new insight by indicating how retained students also have negative spillover effects on their classmates. Thus, by understanding the effects of retention on retained students (i.e., the contribution of previous studies) but also by quantifying the classroom-level peer effect (i.e., the contribution of this study), researchers and practitioners can more efficiently assess how retention policies increase academic risk for individual students and peers. Practically, schools could more efficiently identify and target resources to those classrooms with particularly high numbers of retained students; doing so might mitigate patterns of educational decline for all students affected by this practice, thereby improving the use of classroom and school resources and subsequent test performance.

Second, it has been shown to be particularly useful to identify the characteristics of the nonretained students in the same classroom. This research has found a differential effect of having grade-retained classmates based on individual characteristics such as race and gender. Hence, researchers and practitioners must consider these differences in understanding and remedying the extent to which retained classmates may affect the outcomes of other students. This will enable a more precise assessment of who is placed at greatest risk of educational failure from having grade-retained classmates.

Third, the fact that the data and analyses in this study were multilevel and longitudinal illuminates how schools can more effectively utilize multiple sources of information over time. For example, this study demonstrates the importance of evaluating both student- and classroom-level data. By assessing student-level data in conjunction with classroom measures, schools can guide their policies to most efficiently address their institutional-specific issues by examining the precise effects of having individual students in specific classrooms. This study also demonstrates the importance of relying on longitudinal data, as it has been possible in this research to document analytically how being grade retained in one year can have negative ramifications on the outcomes of other students in subsequent schooling years.

Fourth, the focus on an urban school district has allowed this evaluation to document the effects of retention in high-needs schools—a particularly understudied aspect in both literatures on retention and the peer effects of retention. The analysis in the present research has demonstrated not only that the spillover effects of retained students are significant but that they remain consistently harmful across multiple measures of achievement and differentially for students from different demographic backgrounds, including gender, race, and poverty status. Thus, this supports the premise that a distinctive, and potentially causal, relationship exists between classroom peers and achievement for urban youths.

Conclusion

This research has brought to the foreground a new dimension of the analysis relating to grade retention, as well as peer effects. In essence, by demonstrating a pervasive negative effect of retained classmates, as derived from multiple quasi-experimental methods on a panel data set of urban schoolchildren, this study corroborates the predominant conclusions of prior research on retention. Namely, this practice of grade retention is associated with negative outcomes (Allen et al. 2009; Holmes 1989; McCoy and Reynolds 1999; Meisels and Liaw 1993; Reynolds 1992; Roderick and Nagaoka 2005). The unique contribution of this study is that the negative outcomes of this schooling practice are not restricted to the retained students themselves; rather, this study has found that retention can affect the academic outcomes for other members of the classroom. Moreover, this study has also facilitated an opportunity for urban educational experiences to be further delineated and for policy implications of this practice to be more thoroughly discussed.

Additional insight into these issues could be based on three research extensions. First, though the data used in this study were multilevel and highly comprehensive of a student's demographic background and classroom environment, further detail on reasons why students were retained would yield additional benefits. For instance, in this study, the data did not designate whether a student was retained for academic, behavioral, or other reasons, such as relative age for the grade. However, a data set that included specific information regarding why students were retained would enable future research to differentiate the effects of retained classmates into more distinctive classroom effects. Perhaps having classmates who were retained for academic purposes would yield distinctive outcomes compared to those classmates who were retained for developmental reasons. Additionally, having detailed information on reasons for retention would allow for studies of subgroups of retained peers by type on other subgroups of retained peers by type, for example, the effects of students who were continuously promoted but "should" have been retained on the outcomes for students who were retained but "should" have been continuously promoted.

Second, this study evaluated standardized reading and math achievement outcomes. Future research may entail evaluating the noncognitive outcomes of nonretained students who have greater numbers of retained peers in their classrooms. Noncognitive measures continue to gain momentum in the educational research literature (e.g., Duckworth and Seligman 2005, 2006; Heckman and Rubinstein 2001). A longitudinal data set that included both cognitive and noncognitive measures would be able to approach these research questions. The policy implications from this continued line of work would allow for practitioners to use data to identify which outcomes have the strongest significant relationships associated with being in classrooms with retained peers. In this way, schools could more efficiently evaluate how peer effects place students at risk, both academically and beyond, and develop policies and programs to target specific outcomes most influenced by spillover effects of retention.

Finally, there are many advantages to evaluating students within a single, large urban school district, particularly because these youths are at extremely high risk for educational failure. That being said, differential results and interpretations may arise from the evaluation of other school systems. A research extension, thus, is that the methods employed in this study could be implemented on multilevel, longitudinal data from other districts to assess the generalizability of the new findings emanating from this study: that in addition to previous research upholding that grade retention negatively affects the outcomes of retained students, there are also spillover effects that have negative consequences for the performance of nonretained students in the same classroom.

Notes

1. Data used in the reading and math regressions present different sample sizes depending on the standardized test subject area. However, those subsamples employed for each subject test are not statistically different. Additionally, although student ob-

servations were dropped from the analysis because of missing test score data, the correlation between missing data and test scores is extremely small.

2. To address missing data, two subsidiary analyses were conducted. First, missing data were replaced with mean sample values without missing data for the variable, and additionally, a dummy variable for each variable of missing data was included in the analyses (e.g., Fletcher 2010). Second, multiple imputation was conducted with 10 sets of imputations. Model results were aggregated across the multiply imputed data sets using standard procedures. In both approaches, the results from analyses with imputed missing observations were compared to original estimates. The results were not significantly altered by the inclusion of missing data. Hence, the original specifications do not appear to be unduly biased by missing observations.

3. The measure of free lunch is a strict measure of high poverty, as it includes only those students who are on fully funded programs. Reduced lunch is not included in the data, which would increase the percentage of low-income families to much higher numbers though would dilute the current measure from high poverty to poverty.

4. The name of each student's teacher in each year appears as part of the student's record; that information is extracted from the student's report card along with the classroom number.

5. Less than 1% of the retained sample had been retained more than once. This percentage is consistent with the result in Roderick and Nagaoka (2005). Additionally, approximately 16% of the sample was grade retained, which is consistent with prior research on grade retention in urban schools (e.g., McCoy and Reynolds 1999).

6. A subsequent analysis tested if there was a differential effect based on which grade-retained students had repeated. However, the results did not prove to be significant.

7. An empirical model including squared terms of the number of retained peers was tested in order to search for nonlinearities in the data. The main results remained significant though the squared terms were not. Moreover, the total variance explained did not increase. Thus, these terms were not included in the final model presented in this study.

8. To address the allocation of students to classrooms in more detail, two ancillary models were tested in which the number of grade-retained peers was regressed on all other covariates in the main specification. In all models, there is a lack of statistical significance on class size and mean classroom ability, thereby suggesting that a nonsystematic relationship exists between the number of grade-retained peers and observable classroom characteristics.

9. It is also possible to account for the nested structure of the data using hierarchical linear modeling techniques. However, clustering by classroom and the combination of school, grade, and year fixed effects were purposefully selected in this study rather than hierarchical linear modeling. Clustering occurs after estimation and does not have the complexity of estimating variance components that exists with hierarchical linear modeling. Clustering with fixed effects also relies on fewer preestimation assumptions about the error term than does hierarchical linear modeling, thereby reducing the potential errors in estimation of coefficients throughout the analysis (Primo et al. 2007). Additionally, hierarchical linear modeling appears to be particularly useful when trying to determine where variance lies across multiple levels, which is outside the scope of this study.

10. Although it is possible to conduct this analysis with a binary indicator of being

in a room with retained students versus not, doing so would diminish the richness in variation that exists by having the peer effect of retained students in classroom counts.

References

Alexander, Karl L., Doris R. Entwisle, and Carrie S. Horsey. 1997. "From First Grade Forward: Early Foundations of High School Dropout." *Sociology of Education* 70 (2): 87–107.

Allen, Chiharu S., Qi Chen, Victor L. Willson, and Jan N. Hughes. 2009. "Quality of Research Design Moderates Effects of Grade Retention on Achievement: A Meta-Analytic, Multi-level Analysis." *Educational Evaluation and Policy Analysis* 31 (4): 480–99.

Ammermueller, Andreas, and Jörn-Steffen Pischke. 2006. "Peer Effects in European Primary Schools: Evidence from PIRLS." Working Paper no. 12180, National Bureau of Economic Research, Cambridge, MA.

Anderson, Gabrielle E., Shane R. Jimerson, and Angela D. Whipple. 2005. "Students' Ratings of Stressful Experiences at Home and School: Loss of a Parent and Grade Retention as Superlative Stressors." *Journal of Applied School Psychology* 21 (1): 1–20.

Balfanz, Robert, and Vaughan Byrnes. 2006. "Closing the Mathematics Achievement Gap in High-Poverty Middle Schools: Enablers and Constraints." *Journal of Education for Students Placed at Risk* 11 (2): 143–59.

Brown, Byron W., and Daniel H. Saks. 1987. "The Microeconomics of the Allocation of Teachers' Time and Student Learning." *Economics of Education Review* 6 (4): 319–32.

Caldas, Stephen J. 1993. "Reexamination of Input and Process Factor Effects in Public School Achievement." *Journal of Educational Research* 86 (4): 206–14.

Caldas, Stephen J., and Carl Bankston III. 1997. "Effect of School Population Socioeconomic Status on Individual Academic Achievement." *Journal of Educational Research* 90 (5): 269–77.

Cho, Rosa M. 2012. "Are There Peer Effects Associated with Having English Language Learner (ELL) Classmates? Evidence from the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K)." *Economics of Education Review* 31 (5): 629–43.

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: US Government Printing Office.

Corman, Hope. 2003. "The Effects of State Policies, Individual Characteristics, Family Characteristics, and Neighborhood Characteristics on Grade Repetition in the United States." *Economics of Education Review* 22 (4): 409–20.

Daniel, Larry G., and Debra A. King. 1997. "Impact of Inclusion on Academic Achievement, Student Behavior and Self-Esteem, and Parental Attitudes." *Journal of Education Research* 91 (2): 67–80.

Downing, June, Joanne Eichinger, and Lilly J. Williams. 1997. "Inclusive Education for Students with Severe Disabilities: Comparative Views of Principals and Educators at Different Levels of Implementation." *Remedial and Special Education* 18 (3): 133–42.

Dreyfoos, Joy G. 1990. *Adolescents at Risk: Prevalence and Prevention.* New York: Oxford University Press.

Duckworth, Angela L., and Martin E. P. Seligman. 2005. "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents." *Psychological Science* 16 (12): 939–44.

Duckworth, Angela L., and Martin E. P. Seligman. 2006. "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." *Journal of Educational Psychology* 98 (1): 198–208.

Ekstrom, Ruth B., Margaret E. Goertz, Judith M. Pollak, and Donald A. Rock. 1986. "Who Drops Out of High School and Why? Findings from a National Study." *Teachers College Record* 87 (3): 356–73.

Ergenbright, Kate. 2010. "Special Education Students Disciplined Twice as Often." *Texas Tribune*, August 10.

Everston, Carolyn M. 1982. "Differences in Instructional Activities in Higher- and Lower-Achieving Junior High English and Math Classes." *Elementary School Journal* 82 (4): 329–50.

Everston, Carolyn M., and Edmund T. Emmer. 1981. "Effective Management at the Beginning of the School Year in Junior High Classes." *Journal of Educational Psychology* 74 (4): 485–98.

Figlio, David N. 2007. "Boys Named Sue: Disruptive Children and Their Peers." *Education Finance and Policy* 2 (4): 376–94.

Finn, Jeremy D. 1989. "Withdrawing from School." *Review of Educational Research* 59 (2): 117–42.

Finn, Jeremy D. 1993. *School Engagement and Students at Risk*. Washington, DC: National Center for Education Statistics.

Fletcher, Jason M. 2009. "The Effects of Inclusion on Classmates of Students with Special Needs: The Case of Serious Emotional Problems." *Education Finance and Policy* 4 (3): 278–99.

Fletcher, Jason M. 2010. "Spillover Effects of Inclusion of Classmates with Emotional Problems on Test Scores in Early Elementary School." *Journal of Policy Analysis and Management* 29 (1): 69–83.

Friesen, Jane, Ross Hickey, and Brian Krauth. 2010. "Disabled Peers and Academic Achievement." *Education Finance and Policy* 5 (3): 317–48.

Gottfried, Michael A. 2009. "Excused versus Unexcused: How Student Absences in Elementary School Affect Academic Achievement." *Educational Evaluation and Policy Analysis* 31 (4): 392–415.

Gottfried, Michael A. 2011. "Absent Peers in Elementary Years: The Negative Classroom Effects of Unexcused Absences on Standardized Testing Outcomes." *Teachers College Record* 113:1597–1632.

Greene, Jay P., and Marcus A. Winters. 2007. "Revisiting Grade Retention: An Evaluation of Florida's Test-Based Promotion Policy." *Education Finance and Policy* 2 (4): 319–40.

Greene, Ross W., Sara K. Beszterczey, Tai Katenstein, Kenneth Park, and Jennifer Goring. 2002. "Are Students with ADD More Stressful to Teach?" *Journal of Emotional and Behavioral Disorders* 10 (2): 79–89.

Hanushek, Eric A. 1979. "Conceptual and Empirical Issues in the Estimation of Education Production Functions." *Journal of Human Resources* 14 (3): 351–88.

Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24 (3): 1141–78.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18 (5): 527–44.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2002. "Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities?" *Review of Economics and Statistics* 84 (4): 584–99.

Hayworth, Penny. 2009. "The Quest for a Mainstream EAL Pedagogy." *Teachers College Record* 111 (9): 2179–2208.
Heckman, James J., and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review* 91 (2): 145–49.
Henderson, Vernon, Peter Mieszkowski, and Yvon Sauvageau. 1978. "Peer Group Effects and Educational Production Functions." *Journal of Public Economics* 10 (1): 97–106.
Holmes, C. Thomas. 1989. "Grade-Level Retention Effects: A Meta-Analysis of Research Studies." In *Flunking Grades: Research and Policies on Retention*, ed. Lorrie A. Shepard and Mary Lee Smith. London: Falmer.
Holmes, C. Thomas, and Kenneth M. Matthews. 1984. "The Effects of Nonpromotion on Elementary and Junior High School Pupils: A Meta-Analysis." *Review of Educational Research* 54 (2): 225–36.
Hong, Guanglei, and Stephen W. Raudenbush. 2005. "Effects of Kindergarten Retention Policy on Children's Cognitive Growth in Reading and Mathematics." *Educational Evaluation and Policy Analysis* 27 (3): 205–24.
Hong, Guanglei, and Bing Yu. 2008. "Effects of Kindergarten Retention on Children's Social-Emotional Development: An Application of Propensity Score Method to Multivariate, Multilevel Data." *Developmental Psychology* 44 (2): 407–21.
House, Ernest R. 1999. "Race and Policy." *Education Policy Analysis Archives* 7 (16).
Hoxby, Caroline. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115 (4): 1239–85.
Huber, Peter J. 1967. "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, *Theory of Statistics*, ed. Lucien Lecam and Jerzy Neyman. Berkeley: University of California Press.
Hughes, Jan N., Qi Chen, Felix Thoemmes, and Oi-man Kwok. 2010. "An Investigation of the Relationship between Retention in First Grade and Performance on High Stakes Tests in Third Grade." *Educational Evaluation and Policy Analysis* 32 (2): 166–82.
Jackson, Gregg B. 1975. "The Research Evidence on the Effects of Grade Retention." *Review of Educational Research* 45 (4): 225–36.
Jimerson, Shane R. 1999. "On the Failure of Failure: Examining the Association between Early Grade Retention and Education and Employment Outcomes during Late Adolescence." *Journal of School Psychology* 37 (3): 243–72.
Jimerson, Shane R. 2001a. "Meta-Analysis of Grade Retention Research: Implications for Practice in the 21st Century." *School Psychology Review* 30 (3): 420–37.
Jimerson, Shane R. 2001b. "A Synthesis of Grade Retention Research: Looking Backward and Moving Forward." *California School Psychologist* 6:46–59.
Jimerson, Shane R., Gabrielle E. Anderson, and Angela D. Whipple. 2002. "Winning the Battle and Losing the War: Examining the Relation between Grade Retention and Dropping Out of High School." *Psychology in the Schools* 39 (4): 441–57.
Jimerson, Shane R., and Amber M. Kaufman. 2003. "Reading, Writing, and Retention: A Primer on Grade Retention Research." *Reading Teacher* 56 (7): 622–35.
Johnson, Genevieve M. 2005. "Student Alienation, Academic Achievement, and WebCT Use." *Educational Technology and Society* 8 (2): 179–89.
Karabenick, Stuart A., and Phyllis A. C. Noda. 2004. "Professional Development Implications of Teachers' Beliefs and Attitudes toward English Language Learners." *Bilingual Research Journal* 28 (1): 55–75.

Karweit, Nancy L. 1999. "Grade Retention: Prevalence, Timing, and Effects." Report 70, Center for Research on the Education of Students Placed at Risk, Center for Social Organization of Schools, Johns Hopkins University.

Lavy, Victor M., Daniele Paserman, and Analia Schlosser. 2008. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom." Working Paper no. 14415, National Bureau of Economic Research, Cambridge, MA.

Lazear, Edward P. 2001. "Educational Production." *Quarterly Journal of Economics* 116 (3): 777–803.

Lefgren, Lars. 2004. "Educational Peer Effects and the Chicago Public Schools." *Journal of Urban Economics* 56 (2): 169–91.

Lehr, Camilla A., Anastasia Hansen, Mary F. Sinclair, and Sandra L. Christenson. 2003. "Moving beyond Dropout towards School Completion: An Integrative Review of Data-Based Interventions." *School Psychology Review* 32 (3): 342–64.

Link, Charles R., and James G. Mulligan. 1991. "Classmates' Effects on Black Student Achievement in Public School Classrooms." *Economics of Education Review* 10 (4): 297–310.

Lorence, Jon, and Anthony G. Dworkin. 2006. "Elementary Grade Retention in Texas and Reading Achievement among Racial Groups: 1994–2002." *Review of Policy Research* 23 (5): 999–1033.

Mantzicopoulos, Panayota, and Delmont Morrison. 1992. "Kindergarten Retention: Academic and Behavioral Outcomes through the End of Second Grade." *American Educational Research Journal* 29 (1): 182–98.

McCoy, Ann R., and Arthur J. Reynolds. 1999. "Grade Retention and School Performance: An Extended Investigation." *Journal of School Psychology* 37 (3): 273–98.

McEwan, Patrick J. 2003. "Peer Effects on Student Achievement: Evidence from Chile." *Economics of Education Review* 22 (2): 131–41.

Meisels, Samuel J., and Fong-Ruey Liaw. 1993. "Failure in Grade: Do Retained Students Catch Up?" *Journal of Educational Research* 87 (2): 69–77.

Morgan-D'Atrio, Cindy, John Northup, and Lynn LaFleur. 1996. "Toward Prescriptive Alternatives to Suspensions: A Preliminary Evaluation." *Behavioral Disorders* 21 (2): 190–200.

Morrison, Frederick J., Elizabeth M. Griffith, and Denise M. Alberts. 1997. "Nature-Nurture in the Classroom: Entrance Age, School Readiness, and Learning in Children." *Developmental Psychology* 33 (2): 254–62.

Natriello, Gary, Edward L. McDill, and Aaron M. Pallas. 1990. *Schooling Disadvantaged Children: Racing against Catastrophe*. New York: Teachers College Press.

Neidell, Matthew, and Jane Waldfogel. 2008. "Cognitive and Noncognitive Peer Effects in Early Education." *Review of Economics and Statistics* 92 (3): 562–76.

Newmann, Fred. 1981. "Reducing Student Alienation in High Schools: Implications of Theory." *Harvard Educational Review* 51 (4): 546–64.

Ogbu, John U. 1989. "The Individual in Collective Adaptation: A Framework for Focusing on Academic Underperformance and Dropping Out among Involuntary Minorities." In *Dropouts from School: Issues, Dilemmas, and Solutions*, ed. Lois Weis, Eleanor Farrar, and Hugh G. Petrie. Albany: State University of New York Press.

Pagani, Linda, Richard E. Tremblay, Frank Vitaro, Bernard Boulerice, and Pierre McDuff. 2001. "Effect of Grade Retention on Academic Performance and Behavioral Development." *Development and Psychopathology* 13 (2): 297–315.

Primo, David M., Matthew L. Jacobsmeier, and Jeffrey Milyo. 2007. "Estimating the Impact of State Policies and Institutions with Mixed-Level Data." *State Politics and Policy Quarterly* 7 (4): 446–59.

Reynolds, Arthur J. 1992. "Grade Retention and School Adjustment: An Explanatory Analysis." *Educational Evaluation and Policy Analysis* 14 (2): 101–21.

Roderick, Melissa, and Jenny Nagaoka. 2005. "Retention under Chicago's High-Stakes Testing Program: Helpful, Harmful, or Harmless?" *Educational Evaluation and Policy Analysis* 27 (4): 309–40.

Schmidt, Mary A. 2000. "Teachers' Attitudes toward ESL Students and Programs." In *Inclusive Education: A Casebook and Readings for Prospective and Practicing Teachers*, ed. Suzanne Wade. Mahwah, NJ: Erlbaum.

Schneider, Barbara, Martin Carnoy, Jeremy Kilpatrick, William H. Schmidt, and Richard J. Shavelson. 2007. *Estimating Causal Effects Using Experimental and Observational Designs*. Washington, DC: American Education Research Association.

Shepard, Lorrie A. 1989. "A Review of Research on Kindergarten Retention." In *Flunking Grades: Research and Policies on Retention*, ed. Lorrie A. Shepard and Mary Lee Smith. London: Falmer.

Shepard, Lorrie A., and Mary Lee Smith. 1990. "Synthesis of Research on Grade Retention." *Educational Leadership* 47 (8): 84–88.

Stouthamer-Loeber, Magda, and Rolf Loeber. 1988. "The Use of Prediction Data in Understanding Delinquency." *Behavioral Sciences and the Law* 6 (3): 333–54.

Summers, Anita, and Barbara Wolfe. 1977. "Do Schools Make a Difference?" *American Economic Review* 67 (4): 639–52.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–38.

Willms, J. Douglas. 1986. "Social Class Segregation and Its Relationship to Pupils' Examination Results in Scotland." *American Sociological Review* 51 (April): 224–41.

Willson, Victor L., and Jan N. Hughes. 2006. "Retention of Hispanic/Latino Students in First Grade: Child, Parent, Teacher, School, and Peer Predictors." *Journal of School Psychology* 44 (1): 31–49.

Zabel, Jeffrey E. 2008. "The Impact of Peer Effects on Student Outcomes in New York City Public Schools." *Education Finance and Policy* 3 (2): 197–249.

Zimmer, Ron W., and Eugenia F. Toma. 2000. "Peer Effects in Private and Public Schools across Countries." *Journal of Policy Analysis and Management* 19 (1): 75–92.