

# The Effect of Primary School Size on Academic Achievement

Seth Gershenson  
Laura Langbein

*American University*

*Evidence on optimal school size is mixed. We estimate the effect of transitory changes in school size on the academic achievement of fourth- and fifth-grade students in North Carolina using student-level longitudinal administrative data. Estimates of value-added models that condition on school-specific linear time trends and a variety of teacher-by-school, student, and school-by-year fixed effects suggest that, on average, there is no causal relationship between school size and academic performance. However, two subgroups of interest are significantly harmed by school size: socioeconomically disadvantaged students and students with learning disabilities. The largest effects are observed among students with learning disabilities: A 10-student increase in grade size is found to decrease their math and reading achievement by about 0.015 test-score standard deviations.*

**Keywords:** *school size, school climate, special education, education production function*

SCHOOL quality, broadly defined, is an important predictor of educational attainment and labor-market success (e.g., Altonji & Mansfield, 2011). School size is one potential measure of school quality over which policymakers have some control. For example, the school consolidation movement in the United States in the middle of the 20th century was predicated on the notion that larger schools could offer more specialized instruction, increase administrative efficiency, and reduce per-student costs by exploiting economies of scale (e.g., Berry, 2006; Conant, 1959; Duncombe & Yinger, 2007). The movement successfully eliminated about 70% of schools and increased the average school enrollment from less than 100 to about 440 between 1930 and 1970 (Berry & West, 2010). More recently, primary school enrollments have been affected by desegregation policies such as Boston's Metco program (Angrist & Lang, 2004) and the increasing prevalence of school-closure and school-choice

programs (Brummet, 2014; Common Core of Data [CCD], 2011; Ravitch, 2011).<sup>1</sup> Today, the average U.S. primary school enrolls about 480 students and experiences annual enrollment fluctuations of about 30 students.<sup>2</sup>

However, the benefits of larger schools described in the preceding paragraph come at a cost, as larger schools have higher rates of student absences and social disorder that may hinder cognitive and social development (Gottfredson & DiPietro, 2011). Perhaps unsurprisingly, empirical evidence on the relationship between school size and academic performance is mixed (see Andrews, Duncombe, and Yinger, 2002; Cotton, 1996; and Leithwood and Jantzi, 2009, for thorough reviews of the literature). Kuziemko (2006) speculates that the "current confusion in the literature" is at least partly driven by the cross-sectional, correlational nature of many previous studies of school size. Similarly, Andrews et al. (2002) lament the literature's

general failure to utilize student-level data. Exceptions to these critiques include value-added style analyses that use instrumental-variables procedures to identify the impact of attending a small high school on educational achievement in Chicago (Barrow, Claessens, & Schanzenbach, 2013) and New York City (Schwartz, Stiefel, & Wiswall, 2013). Furthermore, Schwartz et al. (2013) find evidence of heterogeneous effects by schools' vintage, suggesting that the earlier literature's focus on parsimonious models that assumed homogeneous effects of school size may have further contributed to the mixed results.

A second shortcoming of the existing school-size literature is the underrepresentation of primary schools (Cotton, 1996; Leithwood & Jantzi, 2009; Ready & Lee, 2006), as the mechanisms through which school size influences academic performance likely differ between primary and secondary schools. This may be partially attributable to the recent "small schools" movement's focus on high schools. However, understanding the relationship between primary school size and achievement is no less important, as the recent wave of school closures, school choice, and desegregation programs in the United States have altered enrollment patterns, and children undergo substantial developmental changes during these ages. For example, problems of chronic absences and school disengagement begin to manifest as early as first grade (Alexander, Entwisle, & Kabbani, 2001; Schoeneberger, 2012). Existing studies of the relationship between primary school size and academic performance are typically correlational and were conducted at the school or district level. For example, Lamdin (1995) and Chen (2007) examine school-level data from Baltimore and New York City, respectively, and find contradictory results. In a study notable for applying an instrumental-variables procedure to school-level panel data, Kuziemko (2006) finds a sizable negative relationship between primary school size and academic performance in Indiana during the 1990s.

The present study contributes to the school-size literature by exploiting within-school variation in enrollments in North Carolina's public primary schools between 2003 and 2010 to identify the effect of such changes on the academic performance of individual students. Our preferred identification strategy is similar to that used by

Hanushek, Kain, and Rivkin (2009) to investigate the effect of school diversity on educational achievement in that we include school size as a contemporaneous input in value-added models (VAMs) of the education production function that condition on various combinations of teacher-by-school fixed effects (FE), school-by-year FE, student FE, school-specific linear time trends, and a rich set of time-varying school characteristics that jointly predict enrollment and academic performance. The various FE and time trends are important and novel contributions of the present study as they account for the nonrandom sorting of teachers and students across schools and the unobserved trends and year-specific shocks that jointly influence enrollments and student achievement.

The empirical analysis utilizes linked student-, teacher-, and school-level administrative records from North Carolina's longitudinal data system. North Carolina's Department of Public Instruction (DPI) began collecting these data in the mid-1990s, which are now made available to researchers through the North Carolina Education Research Data Center (NCERDC). The NCERDC was formed in 2000 in cooperation with DPI, along with financial support from the Spencer Foundation, by researchers at Duke University and the University of North Carolina (Muschkin, Bonneau, & Dodge, 2011). Since then, the NCERDC has cleaned, coded, and standardized numerous longitudinal data files; created encrypted student and teacher identifiers that allow students, teachers, and schools to be linked while retaining individuals' anonymity; and created extensive codebooks for each data set along with instructions for linking students, teachers, and classrooms. The NCERDC is currently housed in the Social Science Research Institute at Duke University and continues to process each new wave of data collected by DPI.

The authors obtained the data used in the present study by submitting a formal research proposal, which included a description of the project, evidence of institutional review board (IRB) approval from the authors' home institution, and a data security plan, to the NCERDC.<sup>3</sup> Of course, we are not the first researchers to address policy-relevant research questions using these data. For example, NCERDC data have been used to study differences in the observed teacher characteristics in charter relative to traditional public schools

(Carruthers, 2012), the relationship between observable teacher characteristics and student test scores (Clotfelter, Ladd, & Vigdor, 2007), stability of value-added measures of teacher effectiveness across time (Goldhaber & Hansen, 2013) and subjects (Goldhaber, Cowan, & Walch, 2013), the persistence of estimated teacher effects over time (Jacob, Lefgren, & Sims, 2010), teacher peer effects (Jackson & Bruegmann, 2009), the effect of statewide early childhood programs on third-grade academic achievement (Ladd, Muschkin, & Dodge, 2014), the effect of NCLB's subgroup-specific accountability requirements on student achievement (Lauen & Gaddis, 2012), and the presence of nonrandom student-teacher assignments (Rothstein, 2010). This is not an exhaustive list; see the NCERDC website (provided in Note 3) for a full list of projects to which the NCERDC contributed data. The sheer volume of high-quality policy-relevant education research generated by the NCERDC-DPI partnership suggests that this is a successful model that other states and universities might emulate.

### **Theoretical Background and Literature Review**

Previous scholarship has considered the potential for school size to influence students' behavior and academic performance. Regarding the former, theoretical and empirical evidence suggests that larger schools experience higher rates of student indiscipline (e.g., Haller, 1992; Johnston, 2009; Leung & Ferris, 2008). Disorderly incidents may decrease academic achievement through some combination of diverting student attention, creating a fearful or disruptive environment, changing schools' social norms, and decreasing student attendance (Akerlof & Kranton, 2002; Gottfredson & DiPietro, 2011).

Through what mechanisms, then, might a change in school size affect achievement net of its effects on school disorder and average daily attendance? We interpret such effects as the result of changes in schools' climates (Akerlof & Kranton, 2002; Anderson, 1982; Welsh, Stokes, & Greene, 2000). Specifically, the National School Climate Center states that "School climate is based on patterns of students', parents' and school personnel's experience of school life and reflects norms, goals, values, interpersonal

relationships, teaching and learning practices, and organizational structures."<sup>4</sup> School size affects school climate by changing the school's stock of social capital (Coleman, 1988) and "sense of community" (Wynne & Ryan, 1997). For example, there is less frequent and less direct communication between teachers, administrators, and students in larger schools (Gottfredson & DiPietro, 2011). Akerlof and Kranton (2002) argue that students in small schools benefit by being better able to identify with the school and with each other. Boccardo, Schwartz, Stiefel, and Wiswall (2013) provide empirical support for these claims by showing that students have better interpersonal relationships in New York City's old small schools. Furthermore, the benefits of small schools may spill over into the community, as Dee, Ha, and Jacob (2006) provide suggestive evidence that small rural schools promote parental involvement in the form of volunteering at school and participation in Parent Teacher Associations.

Numerous studies, reviewed by Andrews et al. (2002), Cotton (1996), and Leithwood and Jantzi (2009), have investigated the relationship between school size and academic achievement. However, much of the existing literature provides mixed evidence and cannot be given a causal interpretation (Kuziemko, 2006). The recent studies that do provide arguably causal estimates (e.g., Barrow et al., 2013; Bloom, Levy Thompson, & Unterman, 2010; Schwartz et al., 2013) are not particularly relevant to the present study, both because they focus on high schools and because they examine the effect of variation in school size generated by the "small high schools" movement. In addition to the likelihood that the mechanisms through which secondary school size affects academic achievement are different from those in primary schools, these studies exploit a fundamentally different type of variation in school size from that investigated in the present article. For example, Bloom et al. (2010) and Bloom and Unterman (2014) investigate the impact of attending a "Small School of Choice" (SSC) on high school graduation and college readiness. New York City built 123 SSCs in the early to middle 2000s that were designed to educate traditionally disadvantaged students. The SSCs were oversubscribed so admissions were determined by random lottery, which

allowed MDRC researchers to tease out the causal effect of attending an SSC. However, although the SSCs did have positive impacts on students' academic progress and achievement, it is impossible to disentangle the effect of school size from the other innovations simultaneously implemented in SSCs (e.g., curricular changes; Bloom & Unterman, 2014).

Only two studies have used student-level data to attempt to identify the causal effect of transitory variation in school size on academic achievement in the context of primary schools. We review each in turn both to contextualize the results and identify the novel contributions of the present study. First, Lee and Loeb (2000) estimated the effect of school size using student-level achievement data in a value-added context. The authors estimated hierarchical linear models of the 1997 math achievement of sixth- and eighth-grade students in Chicago Public Schools that controlled for lagged math achievement. The authors categorized schools as small (<400 students), medium, or large (>750 students) and found that math-achievement gains were significantly greater in small schools than in medium- or large-sized schools, but the authors found no significant difference between medium and large schools. However, the lack of panel data prohibited the authors from using school FE to control for unobserved school-level heterogeneity that may be correlated with school size. Second, McMillen (2004) conducted a similar analysis of third- and fifth-grade end-of-grade test scores for North Carolina's 1997 third-grade cohort. Because 2 years transpired between the current and lagged test scores, the author measured school size as the 2-year (fourth and fifth grade) average school size experienced by each student. Again, the lack of repeated observations of schools over time prevented the use of school FE. The author found no direct effect of school size on academic achievement.

The present study expands on these early works in three important ways. First, by using school-level panel data we are able to control for both the nonrandom sorting of teachers and students across schools and unobserved school heterogeneity by conditioning on various combinations of teacher-by-school, student, and school-by-year FE as well as linear school time trends. Second, rich administrative data from a

state as diverse as North Carolina enables the present study to provide results that are more generalizable to the U.S. student population than the results of a district-level analysis, test for differential effects of school size by schools' and students' observable characteristics, and control for observed school-level attributes that jointly predict achievement gains and school size. Finally, by updating these early analyses using data from the modern era of school choice, accountability, charter schools, and school closings, we are able to estimate the effect of school size in the context of current education policy and test for differential effects of school size in charter schools.

## **Data**

We use student-level administrative records on the population of third- through fifth-grade students in North Carolina's public schools between 2003 and 2010 to estimate VAMs of the education production function. The data contain end-of-grade math and reading scores, student demographics, classroom identifiers, and a set of potentially time-varying school-level characteristics.<sup>5</sup> Although the generalizability of North Carolina's educational context is potentially limited, two redeeming features make the data well suited for the present analysis. First, repeated observations of both schools and students over time enable FE estimators that exploit within-school and within-student changes in school size over time. Second, detailed annual school-level administrative data on total enrollment; average daily attendance; suspensions, expulsions, and crimes per 1,000 students; charter status; grade span; geographic locale; and the demographic composition of the student body facilitate tests for heterogeneous effects of school size and provide time-varying information on the attributes that likely covary with school size and predict student achievement. Furthermore, North Carolina is home to a geographically, socioeconomically, and demographically diverse student population: The analytic sample contains 691,450 unique students who attended 1,417 unique schools in North Carolina between 2004 and 2010.<sup>6</sup> The analytic sample is restricted to students who attended the same school in third through fifth grades to avoid conflating the effect

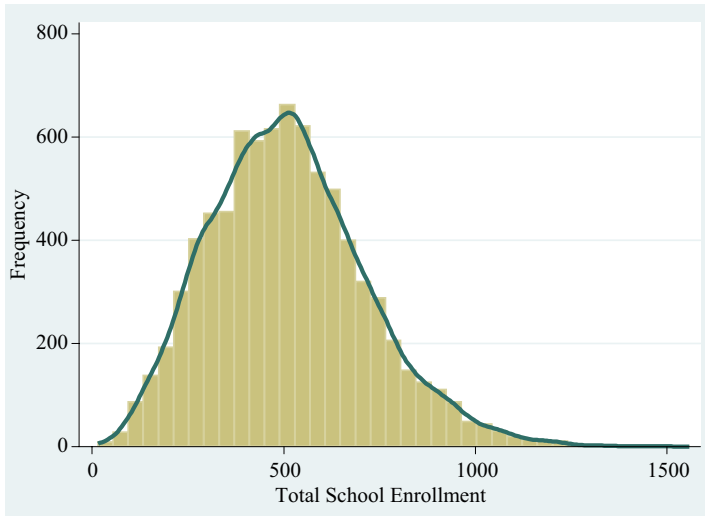


FIGURE 1. *The distribution of school size.*

Note. A kernel density plot is overlaid on the histogram. School-years are the unit of observation ( $N = 8,084$ ).

of changing schools with that of changes in enrollment; however, including the approximately 50,000 students who changed schools during this time does not change the qualitative results. Importantly, because school size is taken from administrative enrollment data, these sample restrictions do not create systematic measurement error in the school-size variables.

The independent variable of interest is school size, which can be operationalized as either total student enrollment or total grade-level enrollment (Ready & Lee, 2006). We consider both measures, as Manski (1993) and Epple and Romano (2011) stress the importance of appropriately identifying peer “reference groups.” Enrollment counts in the administrative data reflect schools’ average daily membership during the course of each academic school-year. However, there are several theoretical and methodological reasons to prefer grade-level enrollment as a measure of school size. First, students and teachers within each grade level primarily interact with one another (Ready & Lee, 2006). Second, schools’ enrollments may vary across grades due to demographic trends in the district. Third, schools’ total enrollments depend on grade spans (Lee & Loeb, 2000). Finally, the school-by-year FE identification strategy of Hanushek et al. (2009), which we describe in the following section, can only be implemented using grade-level enrollments.

When possible, each model is estimated twice: once using total school enrollment as the measure of school size and once using total enrollment in the student’s current grade as the measure of school size.

Figure 1 depicts histogram and kernel density estimates of the school-size distribution in which size is measured as total school enrollment and school-years are the unit of observation. The school-size distribution is approximately symmetric about the mean of 512 students but has a long right tail that contains schools with 1,000 or more students. Figure 2 does the same for the distribution of total grade enrollments, and again the distribution is approximately symmetric with a long right tail. Summary statistics of the school-size distribution are reported in Table 1.

North Carolina’s primary schools moderately increased in size between 2004 and 2008, before shrinking in 2009 and 2010. Table 1 also reports school-size summary statistics conditional on school type, as previous research suggests potential heterogeneities in the effect of school size by vintage (Schwartz et al., 2013) and geographic locale (Crispin, 2012). Charter schools are substantially smaller than traditional public schools and the 87 new schools introduced between 2004 and 2008 tend to be slightly larger than existing schools. Urban and suburban schools are similar in size and both tend to be larger than rural

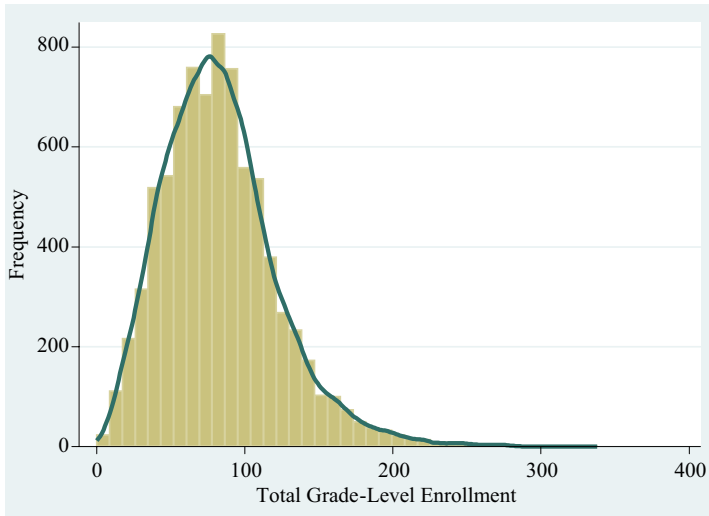


FIGURE 2. *The distribution of grade size.*

Note. A kernel density plot is overlaid on the histogram. School-years are the unit of observation ( $N = 8,084$ ).

TABLE 1  
*Primary School Size in North Carolina, 2004 Through 2010*

	School enrollment		Grade enrollment		N
	M	SD	M	SD	
All schools	512	205	83	39	8,084
2004	495	204	82	40	1,272
2005	510	202	85	39	1,221
2006	514	208	83	38	1,217
2007	523	213	82	40	1,244
2008	525	208	83	40	866
2009	504	202	82	38	941
2010	513	199	85	38	1,323
Charter schools	316	206	37	24	138
Traditional public schools	515	204	84	39	7,946
New schools	522	167	85	39	87
Not new schools	512	206	83	39	7,997
Urban schools	538	208	86	38	2,753
Rural schools	485	205	77	38	3,822
Suburban schools	530	193	93	42	1,509
Within-school annual change in size					
Absolute value	33	46	11	10	6,074
Percent change	7.2		15.8		
Decreases	-35	55			2,738
Decreases			-12	11	2,931
Increases	33	38			3,255
Increases			12	10	2,931

Note. School-years are the unit of analysis. School size is measured by total student enrollment. The analytic sample contains 1,417 unique primary public schools. *SD* = standard deviation.



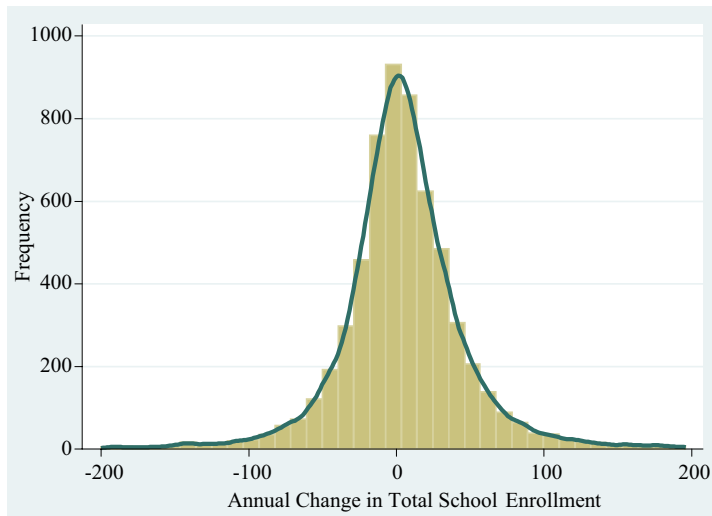


FIGURE 3. *Within-school annual changes in total school enrollments.*

*Note.* A kernel density plot is overlaid on the histogram. School-years are the unit of observation, and 1 year of data are lost in creating the annual changes ( $N = 6,074$ ).

schools. Appendix Table A1 similarly describes the population of U.S. primary schools for the same time period using data from the National Center for Educational Statistics' CCD.<sup>7</sup> North Carolina's primary schools are about 25 students larger than the national average, although average within-school annual changes in enrollment are similar to corresponding U.S. averages.

Because the identification strategy described in the following section exploits within-school variation in enrollments over time, it is also useful to describe the annual enrollment changes experienced by North Carolina's primary schools between 2004 and 2010. The lower panel of Table 1 shows that the average annual change in school enrollment was 33 students and that average increases and decreases were of roughly the same magnitude. Similarly, the average annual change in grade-specific enrollment was 11 students, and again, increases and decreases were equally likely and of the same size, on average. In percentage terms, the average within-school change in total enrollment was about 7%, and the average within-school change in grade-level enrollment was about 16%. Figures 3 and 4 plot the distributions of annual within-school changes in total and grade-level enrollments, respectively. Both are approximately normally distributed and centered on small changes in enrollment. Although changes are generally small, there are

occasional changes in school enrollments of 100 to 200 students and changes in grade-level enrollments as large as 40 students. The following section investigates the possible sources of such changes and how schools accommodate changes in enrollment.

Online Appendix Table B1 (see the online appendix at <http://epa.sagepub.com/supplemental>) reports summary statistics for all variables included in the empirical models at both the student-year and school-year levels.<sup>8</sup> The analytic sample is approximately evenly split between the fourth and fifth grades, poor and nonpoor, male and female, and White and non-White students. The majority of North Carolina's primary schools are K–5, and only 2% are classified as charter schools. Disorderly incidents are rare in the estimation sample, which is unsurprising given the ages of primary school students; still, they do occur.

### How and Why Does School Size Change?

This section presents some descriptive regressions that investigate why school enrollments change over time and how schools accommodate such changes. Changes in a school's enrollment must be associated with one or more of the following: a change in grade span, a change in the number of classrooms, or a change in average

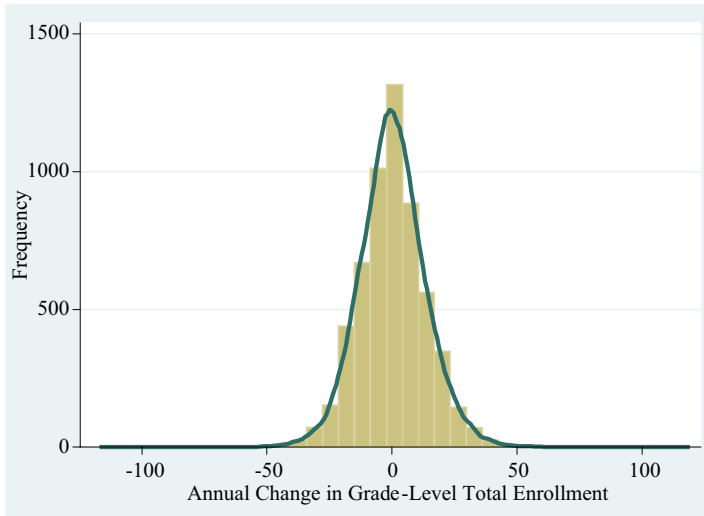


FIGURE 4. *Within-school annual changes in total grade-level enrollments.*

*Note.* A kernel density plot is overlaid on the histogram. School-years are the unit of observation, and 1 year of data are lost in creating the annual changes ( $N = 6,074$ ).

class size. About 10% of the 1,417 schools in the analytic sample experienced a change in grade span between 2004 and 2010. Column 1 of Table 2 reports within-school (school FE) estimates of the effect of grade span on total enrollment.<sup>9</sup> The point estimate of 37.7 indicates that, on average, a one-grade increase in grade span from one academic year to the next increases a school's total enrollment by about 38 students. However, the regression's  $R^2$  indicates that grade-span changes account for less than 5% of the total within-school variation in enrollments observed between 2004 and 2010. Still, given that 10% of schools did experience a change in grade span during the time period analyzed in the present study, as a sensitivity analysis, we will estimate the baseline model on the restricted sample of schools that did not change grade spans between 2004 and 2010.

Columns 2 through 10 of Table 2 further investigate three dimensions of the relationship between school size and school structure.<sup>10</sup> The first panel reports estimated effects of total school enrollment on grade-specific enrollments. Again, these estimates were generated by bivariate school-FE regressions. The coefficients on total enrollment suggest that at least between kindergarten and sixth grade, enrollment changes are evenly distributed across grades. Specifically, the point estimates suggest that on average, a one-student increase in school enrollment is

associated with a 0.15 student increase in grade size. The seventh- and eighth-grade point estimates are slightly smaller, although not significantly so. They are also less precisely estimated, which is likely due to the significantly smaller number of K–8 schools in the analytic sample. Overall, the results reported in the first panel of Table 2 suggest that variation in school size is driven by students changing schools and idiosyncratic variation across districts in cohort size as opposed to systematic increases in the size of incoming public school cohorts. The sources of variation in school enrollments are further investigated in Table 3.

Having shown that the within-school variation in school size is approximately evenly distributed across grade levels, the second and third panels of Table 2 examine the relationship between changes in grade size and the average size and number of that grade's classrooms, respectively. The effect of grade-specific enrollment on average class size is smaller in the lower grades, suggesting that schools are more likely to accommodate increased enrollments in lower grades, particularly kindergarten through third grade, by adding additional classrooms. Still, even in the higher elementary grades, the effects of enrollment increases on average class sizes are modest: For example, a 10-student (1  $SD$ ) increase in fifth-grade enrollment is associated



TABLE 2  
Structural Relationships Between School Size, Grade Span, Grade Size, and Average Class Size

	Total enrollment	Kindergarten	First grade	Second grade	Third grade	Fourth grade	Fifth grade	Sixth grade	Seventh grade	Eighth grade
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grade span	37.7 (4.8)***									
Total enrollment by grade										
Total school enrollment		0.153 (0.005)***	0.157 (0.004)***	0.153 (0.003)***	0.149 (0.003)***	0.151 (0.003)***	0.147 (0.003)***	0.145 (0.021)***	0.137 (0.014)***	0.124 (0.012)***
Adj. $R^2$	0.03	0.41	0.48	0.48	0.45	0.46	0.44	0.26	0.34	0.28
Average class size by grade										
Total grade enrollment		0.029 (0.002)***	0.028 (0.002)***	0.029 (0.002)***	0.029 (0.002)***	0.042 (0.003)***	0.043 (0.003)***	0.059 (0.020)***	0.105 (0.031)***	0.090 (0.024)***
Adj. $R^2$		0.03	0.02	0.03	0.03	0.02	0.03	0.04	0.05	0.06
Classrooms per grade										
Total grade enrollment		0.044 (0.001)***	0.046 (0.001)***	0.045 (0.001)***	0.044 (0.001)***	0.038 (0.001)***	0.037 (0.001)***	0.038 (0.003)***	0.035 (0.003)***	0.036 (0.003)***
Adj. $R^2$		0.63	0.56	0.54	0.53	0.40	0.38	0.45	0.35	0.42
$n$	9,262	7,591	8,865	8,909	9,092	9,170	8,870	1,325	864	840

Note. Each cell contains the coefficient from a bivariate regression that controls for school fixed effects. The dependent variables in columns 2 to 10 are grade-specific enrollments, average class sizes, and classrooms per grade. Parentheses contain standard errors that are robust to clustering at the school level. Enrollments and average class sizes are measured on a one-to-one student level. Grade span is a count of the number of grade levels contained in the school. Controlling for year fixed effects or restricting the sample to schools that did not experience a change in grade span during the sample's time frame yields qualitatively similar results. Sample sizes change across specifications because schools contain different grade levels.

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

TABLE 3

*Relationships Between Area Characteristics and School Enrollments*

Area	ZIP code		LEA		County	
	(1)	(2)	(3)	(4)	(5)	(6)
Closure(s)	10.60 (4.67)**	10.70 (4.63)**	-0.30 (3.22)	-1.99 (3.12)	2.29 (2.34)	3.21 (2.28)
Opening(s)	-15.90 (1.99)***	-43.59 (4.69)***	-11.36 (1.48)***	-14.43 (2.16)***	-9.03 (1.40)***	-8.36 (1.91)***
Area enrollment	0.04 (0.01)***	0.05 (0.01)***	0.002 (0.001)***	0.003 (0.001)***	0.002 (0.001)**	0.002 (0.001)**
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes	No	Yes
Adj. $R^2$	0.04	0.06	0.01	0.02	0.01	0.02
Area-by-year means and [SD]						
Closure(s)	0.01		0.02		0.06	
Opening(s)	0.18		0.22		0.27	
Area enrollment	1,161		3,922		6,708	
[SD]	[1,042]		[7,946]		[10,039]	
Within [SD]	[164]		[635]		[864]	
$n$	4,043		1,197		700	

Note. Regression analytic sample contains 9,262 school-years and 1,417 unique schools. The dependent variable in all models is school total enrollment. Parentheses contain standard errors that are robust to clustering at the school level. Lost and Gained are binary indicators that equal 1 if the school's area lost and gained schools since the previous academic year, respectively, and 0 otherwise. Thus, the omitted category is "no change" in the number of area schools. Total area enrollments are computed as the sum of all total enrollments reported by area schools. Brackets contain standard deviations of continuous variables [SD]. Within refers to within-area SD. FE = fixed effects; LEA = Local Educational Agency (i.e., school district).

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

with an average increase in fifth-grade class size of about 0.4 students. The "classroom per grade" results in the last panel of Table 2 are consistent with the average class size results, again suggesting that enrollment increases in the lower primary grades are more likely to result in new classrooms than are enrollment increases in the higher primary grades, although the effects are fairly similar across grade levels. Generally, a 25-student (2.5 SD) increase in any grade's enrollment is associated with about one new classroom in that grade.

Together, the descriptive results presented in Table 2 provide two general insights into how within-school variation in total enrollment relates to schools' average grade and class sizes. First, increases in total school enrollment are approximately evenly distributed across grade levels. Second, increases in grade size are more likely to be accommodated by increasing the number of classrooms than by increasing average class size,

particularly in the lower primary grades (i.e., K-5). These findings provide context for interpreting the main analyses of the relationship between school size and student achievement reported below. The first finding, that enrollment changes are approximately evenly distributed across grade levels, seems to rule out one potential source of within-school variation in total enrollments: systematic trends in the size of incoming kindergarten cohorts during the sample time period.

By definition, changes over time in a school's total enrollment are driven by students entering and exiting the school. One source of variation in total enrollment, alluded to above, is that at the start of each new school-year the cohort in the school's highest grade level that graduated the previous spring is "replaced" by the new cohort entering the school's lowest grade level. The other source of within-school variation in total enrollments is student mobility. Generally,

student mobility includes students who enter or leave the state of North Carolina, enter or leave particular school districts, and make intradistrict school changes. The latter may include moves induced by school closures and school openings.

Table 3 investigates how school closures, school openings, and total enrollments in schools' catchment areas affect school enrollments. Three geographic catchment area definitions are considered. From smallest to largest, they are the school's 5-digit ZIP code, Local Education Agency (LEA; i.e., school district), and county. Each model conditions on school FE, thus exploiting within-school variation in catchment area characteristics and school size, and is estimated both with and without academic year FE. Importantly, the academic year FE control for statewide trends in public primary school enrollments.

The effect of school closure(s) in a given school's catchment area on the school's total enrollment is positive and statistically significant only when the catchment area is defined as the school's 5-digit ZIP code; school closures in a school's district or county have no significant effect on remaining schools' enrollments. The point estimate of approximately 11 suggests that schools within ZIP codes that experienced at least one school closure had an average enrollment increase of 11 students (about 25% of the within-school enrollment increase *SD* shown in Table 1). Similarly, school openings in a school's catchment area have negative, statistically significant effects on the enrollments of remaining schools, regardless of how the catchment area is defined. The effects of school openings are generally larger in magnitude than the corresponding increases in school size associated with school closures, perhaps because openings are significantly more frequent than closures during the sample's time period. Such effects are largest when schools' catchment areas are defined as schools' ZIP codes, which is intuitive given that these schools provide the most direct competition to existing schools.

The total number of students enrolled in public primary schools in the catchment area is also significantly related to individual schools' enrollments, although the effects are modestly sized. For example, the within-ZIP-code standard deviation (*SD*) in total primary enrollment is 164, so

a 1-*SD* increase in ZIP code enrollment is associated with increases of about eight students—or 25% of the average within-school enrollment increase in schools' total enrollments. Once again, the effects are strongest when the catchment area is defined by ZIP code.

Together, the results presented in Table 3 suggest that school openings, school closures, and geographical variation in the growth of primary-school-age populations play modest roles in the within-school variation in total school enrollments observed in the analytic sample. Moreover, statewide trends in public primary school enrollments do not appear to be driving changes in school-level variation in total enrollments. The remaining within-school variation in enrollments seems to be driven by student mobility and idiosyncratic variation in the sizes of schools' entering and graduating cohorts.

## Method

The identification strategy includes school size as a contemporaneous input in VAMs of the education production function. We estimate student-level models so that we can test for heterogeneity in the effect of school size by observable student characteristics and control for teacher FE. We take a basic lag-score VAM specification as a point of departure, as Guarino, Reckase, and Wooldridge (2015) find this approach the most robust to a variety of potential nonrandom student-teacher assignment scenarios, and similar specifications are commonly used in analyses of the effects of educational inputs.<sup>11</sup> Formally, we model student *i*'s end-of-grade test score (*y*) in year *t* as

$$y_{ijst} = \alpha y_{i,t-1} + \beta X_{it} + f(\text{size}_{st}) + \delta Z_{st} + \lambda c_{jst} + \pi \bar{p}_{-i,jst} + \omega_{js} + \tau_t + u_{ijst}, \quad (1)$$

where *j* indexes teachers and *s* indexes schools; *X* is a vector of student characteristics including race, gender, grade level, English proficiency, the presence of a documented learning disability, and poverty status; *f*(size) is a general function of school size; *Z* is a vector of potentially time-varying school characteristics including the demographic and socioeconomic composition of the student body, the size and credentials of the instructional staff, geographic locale (i.e., rural, suburban, or urban), grade span, charter status,

average daily attendance, and the numbers of crimes, suspensions, and expulsions per 1,000 students;  $c$  is student  $i$ 's class size;  $\bar{p}$  is the average of student  $i$ 's year- $t$  classmates' lagged test scores, which is a commonly used control for classroom peer effects;  $\omega$  is a teacher-by-school FE;  $\tau$  is a year FE; and  $u$  is an idiosyncratic error term that contains student  $i$ 's potentially time-varying unobserved ability and test-score measurement error. Standard errors are made robust to clustering at the school level, which makes statistical inference robust to serial correlation within schools *and* students because the estimation sample is restricted to students who did not change schools between third and fifth grades (Angrist & Pischke, 2009).

Five aspects of Equation 1 warrant further discussion. First, we consider several specifications of  $f(\text{size})$ : linear, quadratic, and logarithmic functions of school size as well as sets of categorical indicators for quartiles of the school-size distribution. Similarly, we interact school size with a variety of student- and school-level observed characteristics to test for heterogeneity in the relationship between school size and student achievement.

Second, note that the specifications of  $f(\text{size})$  discussed above restrict the effects of increases and decreases in school size to be symmetric. To test whether the effects of increases and decreases in school size are symmetric, we also consider a linear switching specification of  $f(\text{size})$  in which  $d$  equals 1 if size increased between years  $t - 1$  and  $t$ , and 0 otherwise:

$$f(\text{size}_{st}) = \gamma_1 d_{st} \text{size}_{st} + \gamma_2 (1 - d_{st}) \text{size}_{st}. \quad (2)$$

It is then straightforward to test the hypothesis that  $\gamma_1$  equals  $\gamma_2$ .

Third, two potential concerns are that schools close endogenously and new-school sites are endogenously determined. To verify that the results are not driven by school openings or closings, we estimate the baseline model Equation 1 on the balanced panel of schools that were consistently open between 2004 and 2010 (Wooldridge, 2010).

Fourth, the teacher-by-school FE are crucial to the identification strategy as they control for the quality of individual students' teachers, the sorting of teachers across schools, and unobserved

time-invariant school effects during teachers' school-specific spells. Because the teacher-by-school FE control for unobserved school effects that remain constant during each teacher's spell in a given school, time-invariant teacher and school FE are redundant in Equation 1. However, the time-varying school-level characteristics in  $Z$  perform the equally important role of controlling for school characteristics that potentially vary with school size and predict gains in student performance. Similarly, there may be unobserved school-specific trends that jointly predict enrollments and achievement. Thus, we augment Equation 1 in two ways to ensure that the estimated effect of school size is not biased by the presence of time-varying unobserved school effects: We add either school-specific linear time trends (Wooldridge, 2010) or school-by-year FE (Hanushek et al., 2009) to Equation 1. The latter specification can only identify the effect of school size when size is measured as total grade enrollment, as total school enrollment is perfectly collinear with the school-by-year FE. The school-by-year FE estimates are identified by same-year, same-school differences between fourth- and fifth-grade enrollments.

Finally, a potential concern with the lagged test-score specification of Equation 1 is that unobserved student ability not captured by the lagged test score is left in the error term, which may be correlated with the model's covariates. Guarino et al. (2015) find that ordinary least squares (OLS) on Equation 1 is the most robust estimator in the presence of nonrandom student-teacher assignments, which is likely the case in North Carolina (Rothstein, 2010). However, when estimating the effect of a school-level characteristic rather than teacher effects, student sorting into schools is a potentially greater source of bias than student sorting into classrooms, and the former may be driven by relatively time-invariant student characteristics (Brummet, 2014). Accordingly, we examine the robustness of Equation 1 to conditioning on time-invariant unobserved student heterogeneity (i.e., student FE) in a gain-score VAM (Guarino et al., 2015). Specifically, we modify Equation 1 by decomposing the error term, restricting  $\alpha$  to equal 1, and subtracting the lagged test score from both sides of the equation, which yields:

$$\Delta y_{ijst} = \beta X_{it} + f(\text{size}_{st}) + \delta Z_{st} + \lambda c_{jst} + \pi \bar{p}_{-i,jst} + \omega_{js} + \theta_i + \tau_t + e_{ijst}. \quad (3)$$

TABLE 4

*Value-Added Estimates of School Size's Effect on Achievement*

Model	Lag score	Asymmetric lag score model	Lag score, balanced panel	Lag score, constant grade spans	Lag score + school time trends	Lag score + school-by-year FE	Gain score	Gain score + student FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Math achievement</b>								
School size	-0.010 (0.004)**	-0.012 (0.004)***	-0.010 (0.005)*	-0.009 (0.004)**	-0.007 (0.005)		-0.010 (0.004)**	-0.011 (0.018)
$(1-d)\gamma_2$		-0.013 (0.005)***						
Adjusted $R^2$	0.67	0.67	0.67	0.67	0.67		0.002	-0.23
Grade size	-0.002 (0.001)**	-0.002 (0.001)*	-0.002 (0.001)*	-0.003 (0.001)**	-0.001 (0.001)	0.0004 (0.002)	-0.001 (0.001)	-0.007 (0.008)
$(1-d)\gamma_2$		-0.002 (0.001)						
Adjusted $R^2$	0.67	0.67	0.67	0.67	0.67	0.74	0.002	-0.23
<b>Reading achievement</b>								
School size	-0.006 (0.003)*	-0.008 (0.003)**	-0.007 (0.004)*	-0.005 (0.003)*	-0.002 (0.004)		-0.006 (0.004)*	0.007 (0.017)
$(1-d)\gamma_2$		-0.009 (0.003)**						
Adjusted $R^2$	0.61	0.61	0.61	0.61	0.61		0.003	-0.41
Grade size	-0.001 (0.001)	-0.001 (0.001)	-0.0004 (0.001)	-0.001 (0.001)	0.0001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.002 (0.008)
$(1-d)\gamma_2$		-0.002 (0.001)						
Adjusted $R^2$	0.61	0.61	0.61	0.61	0.61	0.68	0.003	-0.41
$N$	1,034,490	1,034,490	584,370	946,356	1,034,490	1,034,490	1,034,490	1,034,490
$N$ (schools)	1,417	1,417	635	1,273	1,417	1,417	1,417	1,417

*Note.* Each column reports estimates from four distinct regressions: school size and grade size models for both math and reading achievement. School size is measured in hundreds of students and grade size is measured in tens of students. Parentheses contain standard errors that are clustered at the school level. All specifications include the full set of student- and school-level controls described in the text and summarized in Online Appendix Table B1 (see the online appendix at <http://epa.sagepub.com/supplemental>), in addition to teacher-by-school FE. All estimated coefficients for the math and reading specifications in column 1 are reported in Online Appendix Table B2. For each row in column 2, the first estimate is the estimated size coefficient for school-years in which size increased over the previous year, and the second estimate ( $\gamma_2$ ) is the size coefficient for school-years in which size either decreased or did not change since the previous year. None of the asymmetric-pair estimates in column 2 are statistically significantly different from each other at the 95% confidence level. FE = fixed effects.

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

The presence of student FE (or the school-by-year FE mentioned above) in addition to the teacher-by-school FE creates computational difficulties, as traditional FE estimators are infeasible due to the combination of unbalanced panels and the high dimensionality of the problem (Abowd, Kramarz, & Margolis, 1999). Thus, we estimate Equation 3 using the two-way-FE estimator proposed by Mittag (2012).

## Results

### *Main Results*

Each row of Table 4 corresponds to a different regression and reports the estimated coefficient on school size, measured as either total school enrollment or total grade-level enrollment, for a variety of specifications and samples that assume a linear relationship between school

size and student achievement. The top panel of Table 4 reports results for math achievement and the bottom panel reports results for reading achievement. Column 1 reports baseline estimates of the lag-score model given in Equation 1.<sup>12</sup> Column 2 estimates the asymmetric switching specification defined by Equation 2. Column 3 estimates the baseline specification of Equation 1 using a balanced panel of schools that were consistently in operation between 2004 and 2010. Column 4 estimates the baseline model on the restricted sample of schools that did not experience a change in grade span between 2004 and 2010.

The estimated coefficients in the first four columns of Table 4 are quite similar, which suggests two things. First, the effects of increases and decreases in school size are approximately symmetric, as none of the pairs of coefficients in column 2 are significantly different from one another other at the 95% confidence level. Second, the results are not driven by endogenous school closings, the performance of newly opened schools, or schools that changed grade span. Both the school- and grade-level measures are statistically significant in the math-achievement VAMs and tend to be about twice as large as the corresponding estimates in the reading achievement VAMs. Only the school-enrollment estimates are even marginally statistically significant in the reading achievement VAMs. The finding that school size has less, if any, effect on reading than on math achievement is consistent with studies of the relationship between high school size and student achievement (e.g., Crispin, 2012) and with the general finding that educational interventions have relatively stronger impacts on math achievement (e.g., Hanushek & Rivkin, 2010). This may be due to the structure of reading tests (Measures of Effective Teaching, 2010) or the fact that schools do in fact have larger impacts on the development of math skills as children learn and develop reading skills at home (Currie & Thomas, 2001).

The estimated coefficient on total school enrollment in the top panel of column 1 suggests that a 100-student increase in school size lowers math scores by 0.01 of a test-score standard deviation (*SD*). Although this effect appears small,

recall that the *SD* of school size is about 200 students. Thus, a 1-*SD* increase in school size is associated with a 0.02-*SD* decrease in math achievement. This effect is arguably practically significant as it is approximately 20% of the effect of a 1-*SD* increase in teacher effectiveness (e.g., Hanushek & Rivkin, 2010). However, as discussed in the methodology section, the estimates in column 1 might be biased by time-varying unobserved school characteristics that jointly predict enrollment and achievement. To control for this possibility, columns 5 and 6 of Table 4 augment Equation 1 to condition on school-specific linear time trends and school-by-year FE, respectively.

Controlling for linear time trends in column 5 reduces the estimated effect of school size on math achievement by 30% to 50%, and even more so for reading. None of the estimates remain even marginally statistically significant. The school-by-year FE estimates in column 6 actually turn positive but are quite small and indistinguishable from zero. These results suggest that the negative coefficients on school size observed in column 1 were at least partially driven by time-varying unobserved school characteristics.

Column 7 of Table 4 reports the estimated coefficients on school size in gain-score specifications that are otherwise identical to Equation 1, to provide context for comparisons between estimates of Equations 1 and 3. The point estimates in column 7 are generally similar to those in column 1, indicating that the results are robust to moving from the lag-score to gain-score specification. Column 8 of Table 4 reports two-way FE estimates of Equation 3, which conditions on student FE. The math results remain negative and similar in magnitude to those of the other specifications considered in Table 4, but are less precisely estimated. The imprecision of the two-way FE estimates is unsurprising, as only 2 years of data are available per student, the student FE are jointly insignificant, and the adjusted  $R^2$  actually turns negative. Once again, none of the two-way FE reading estimates are statistically significant.

Taken as a whole, the results reported in Table 4 provide no evidence of a causal relationship between transitory changes in school



TABLE 5

*Tests for Nonlinear Effects of Total Grade Enrollment on Achievement*

Specification	Math achievement			Reading achievement		
	Log	Quadratic	Quartiles	Log	Quadratic	Quartiles
	(1)	(2)	(3)	(4)	(5)	(6)
ln(size)	0.0069 (0.0149)			0.0031 (0.0123)		
Size		0.0024 (0.0041)			-0.0021 (0.0032)	
Size <sup>2</sup>		-0.0001 (0.0002)			0.0001 (0.0001)	
1st quartile			0.0003 (0.0071)			-0.002 (0.0060)
4th quartile			0.0098 (0.0071)			0.0045 (0.0065)
Size APE		0.0007 (0.0017)			0.0003 (0.0010)	
<i>F</i> test <i>p</i> value		0.84			0.43	
Adjusted <i>R</i> <sup>2</sup>	0.74	0.74	0.74	0.68	0.68	0.68

Note.  $N = 1,034,490$ . Grade size is measured in tens of students. Parentheses contain standard errors that are robust to clustering at the school level. All specifications include the full set of student- and school-level controls summarized in Online Appendix Table B1 (see the online appendix at <http://epa.sagepub.com/supplemental>), as well as the teacher-by-school and school-by-year FE included in column 4 of Table 3. In columns 3 and 6, the first and third quartiles of grade size are 45 and 72 students. In columns 2 and 5, *p* values are reported for *F* tests of the joint significance of the grade size variables. APE = average partial effect; FE = fixed effects.

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

size and academic achievement, regardless of whether school size is measured at the school or grade level. Specifically, the modestly sized, marginally significant negative effects of school size reported in column 1 vanish when either linear time trends, school-by-year FE, or student FE are controlled for. However, the nonfinding in Table 4 could be due to the linear specification of school size. For example, the true relationship between school size and student achievement might be nonlinear or might vary by student or school characteristics. We subsequently investigate these hypotheses using the school-by-year FE specification, which is arguably the most robust to the presence of unobserved school-year specific shocks (Hanushek et al., 2009), and hence our preferred specification. Similar analyses using the linear time trend specification yield qualitatively similar results.

### *Nonlinear Effects of School Size*

Table 5 tests for nonlinearities in the relationship between school size and student achievement using three nonlinear specifications of  $f(\text{size})$  in the preferred school-by-year FE specification: the natural log, a quadratic function, and categorical indicators for the first and fourth quartiles of the size distribution (omitting the second and third quartiles as the reference group). Columns 1 to 3 of Table 5 estimate these three specifications for math achievement, and columns 4 to 6 do so for reading achievement. None of the estimated coefficients on the nonlinear size terms are individually significantly different from zero. For the quadratic specifications reported in columns 2 and 5 of Table 5, the average partial effects are statistically insignificant and similar in magnitude to those reported in column 6 of Table 4. Moreover, the quadratic terms in Table 5 are jointly



TABLE 6

*Tests for Heterogeneous Effects of Total Grade Enrollment on Achievement*

	Math achievement	Reading achievement
	(1)	(2)
Size (S)	-0.0007 (0.0034)	0.0018 (0.0024)
S × Limited English proficient	-0.0008 (0.0009)	-0.0005 (0.0008)
S × Learning disability	-0.0155 (0.0006)***	-0.0144 (0.0005)***
S × Poverty	-0.0005 (0.0004)	-0.0011 (0.0004)***
S × Male	0.0005 (0.0002)**	0.0003 (0.0002)
S × Charter	-0.0668 (0.0954)	-0.0081 (0.0554)
S × New school	-0.0127 (0.0135)	0.0191 (0.0206)
S × K-8 school	0.0021 (0.0108)	0.0035 (0.0099)
S × Urban school	0.0053 (0.0045)	0.0011 (0.0034)
S × Rural school	-0.0004 (0.0042)	-0.0014 (0.0033)
<i>F</i> test <i>p</i> value		
Student	0.98	0.15
School	0.55	0.98
All	0.88	0.60

Note.  $N = 1,034,490$ . Grade size is measured in tens of students. Parentheses contain standard errors that are robust to clustering at the school level. All specifications include the full set of student- and school-level controls summarized in Online Appendix Table B1 (see the online appendix at <http://epa.sagepub.com/supplemental>), as well as teacher-by-school FE and either linear school time trends or school-by-year FE. The  $p$  values are reported for  $F$  tests of the joint significance of the student-level interactions, school-level interactions, and all interactions. FE = fixed effects.

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

insignificant as well. Taken as a whole, the results reported in Table 5 provide no evidence of a nonlinear relationship between school size and academic achievement and confirm that the nonfindings in Table 4 are not driven by an assumed linear relationship.

#### *Heterogeneous Effects of School Size*

Table 6 tests for heterogeneity in the relationship between school size and student achievement by both student and school observed characteristics. This is accomplished by interacting school size with four student and five school characteristics in the preferred school-by-year FE specification. The observed student characteristics are binary indicators for administratively classified Limited English Proficiency (LEP), subject-specific learning disabilities (i.e., math and reading), poverty status, and gender. The observed school characteristics are binary indicators for charter school, new school, kindergarten through eighth-grade school (as opposed to Kindergarten through

fifth grade), and geographic locale (urban and rural, with suburban serving as the omitted reference category). Because the interaction terms are not jointly significant, either as a group or separately by the student- and school-level interactions, there is no evidence that the results of this analysis are distorted by collinearity among the interaction terms. For this reason, as well as for presentational ease, we include the nine interaction terms simultaneously in one model; however, it is worth noting that adding each interaction term to the baseline specification individually yields qualitatively similar estimates—and patterns of statistical significance—of the interaction effects.

There is reason to believe that students may differentially respond to school size, as Krueger (1999) finds that the effect of class size varies by students' socioeconomic status (SES), and Baydar and Brooks-Gunn (1991) find differential effects of child care on cognitive development by SES and child's gender. The marginal benefit of individual attention provided by smaller schools may be greater for disadvantaged students who

receive less attention at home (Gershenson, 2013; Guryan, Hurst, & Kearney, 2008). Similarly, the effect of school size would differ by students' LEP and learning disability status if exceptional students either encounter difficulties navigating the social and administrative environments of large schools or benefit from larger schools' economies of scale in the provision of services to exceptional students.

The most striking result in Table 6 is that students in larger schools who have administratively classified learning disabilities perform significantly worse on math and reading standardized tests. The interaction effects are similar in magnitude for both math and reading, indicating that a 10-student increase in total grade enrollment reduces the achievement of learning-disabled students by about 0.015 test-score *SD*. This is an arguably practically significant difference, as the sample *SD* of grade enrollments is about 40 students and the average annual change in grade enrollments was 11 students. The poverty interaction is also strongly statistically significant for reading achievement, perhaps because reading skills are primarily developed at home and high-SES households are able to provide more or higher quality resources to the development of children's reading skills (Currie & Thomas, 2001). However, the interaction term is an order of magnitude smaller than that of learning disabilities and is unlikely to be practically significant. Interestingly, none of the school-level interaction terms are statistically significant at traditional confidence levels, although the point estimates of the charter-school and new-school interactions are negative and large in size.

## Discussion

The present study uses rich student-level administrative records from North Carolina's public school system that include repeated observations of both students and schools over time to estimate the effect of school size on academic achievement in a value-added framework. Estimates of naive VAMs suggest that a 1-*SD* increase in school size reduces math achievement by about 0.02 test-score *SD* while the effect on reading achievement is even smaller and only

marginally statistically significant. However, when the likely endogeneity of school size is accounted for either by linear school time trends or school-by-year FEs, the estimated effect of school size on math achievement decreases in magnitude and loses its statistical significance. Therefore, the primary analysis provides no evidence of a causal relationship between school size and student achievement, regardless of whether school size is measured at the school or grade level, nor do we find any evidence of a nonlinear relationship between school size and student achievement, at least within the range of school sizes observed in North Carolina between 2004 and 2010.

However, there do appear to be important heterogeneities in the relationship between school size and student achievement. Specifically, the math and reading achievement of students with learning disabilities, and the reading achievement of socioeconomically disadvantaged students, are disproportionately harmed by increases in school size. Interestingly, the effect of school size does not vary by geographic locale or in charter schools.

Students with learning disabilities may be particularly sensitive to increases in school size either because larger schools are less able to match exceptional students to the relevant support programs or because exceptional students are more sensitive to the weaker social bonds likely inherent in larger schools. Specifically, a 10-student increase in grade size is found to reduce the math and reading achievement of exceptional students by about 0.015 test-score *SD*, which is an arguably practically significant reduction in learning. This modest effect is most easily interpreted relative to the effects of other potential educational interventions that are considered to be practically significant. For example, it is approximately 15% of the effect of a 1-*SD* increase in teacher effectiveness (e.g., Hanushek & Rivkin, 2010). Arguably, these are nontrivial impacts on student achievement as schools in the analytic sample frequently experience annual changes in grade-level enrollments as large as 20, 30, and even 40 students.

The general finding of no overall, causal effect of school size on standardized test scores is also policy relevant and merits discussion, as educators,

reformers, and policymakers have debated the costs and benefits of school size for more than 100 years. This nonfinding is consistent with studies of the small high school movements in Chicago and New York by Barrow et al. (2013) and Schwartz et al. (2013), respectively, whose instrumental-variables estimates find no significant effect of school size on high school students' standardized test scores. These results suggest that the successes of New York City's SSCs (e.g., Bloom et al., 2010; Bloom & Unterman, 2014) were at least partly due to the curricular and organizational changes enacted at the same time as the enrollment reductions. It is also worth noting that Barrow et al. (2013) and Bloom and Unterman (2014) find significant effects of school size on graduation rates, suggesting that improved school climates in smaller schools strengthen noncognitive skills such as motivation and grit that are valued in the workplace and associated with long-run socioeconomic outcomes. It would thus be useful for future research to investigate the effects of school size on character skills in the primary school context.

Finally, the finding in the present study that school size influences the academic achievement of exceptional and disadvantaged students over and above its influence on class size, student indiscipline, and attendance highlights the potentially nuanced importance of school climate in the educational process, which encompasses a number of behaviors, attributes, and cultural and social norms. This raises deeper questions of how and why school climate is a function of school size and why certain subsets of the student population are particularly influenced by school climate. It would be useful for future research to attempt to get inside the "black box" of school size's role in the educational process and further unpack the effects of school size and school climate. For example, do instructional practices or rates of teacher attrition systematically vary by school size? Similarly, additional surveys and qualitative analyses of student, parent, teacher, and administrator perceptions of school climate similar to those analyzed by Lee and Loeb (2000) and Boccardo et al. (2013) may prove useful as policy implications and recommendations for best practice ultimately depend on the channels through which school size and school climate influence students' development and long-run socioeconomic outcomes.

## Appendix

TABLE A1  
*Primary School Total Enrollments in the United States, 2004–2010*

Schools	<i>M</i>	<i>SD</i>	<i>N</i>
All	484	276	446,968
2004	487	289	63,729
2005	486	285	63,445
2006	484	280	62,945
2007	480	274	64,243
2008	481	271	64,466
2009	484	267	64,311
2010	486	264	63,829
Charter	305	250	15,200
TPS	490	275	431,768
Urban	536	271	124,440
Rural	360	262	132,728
Suburban	536	261	189,800
K–5	476	217	208,006
K–8	490	318	238,962
Within-school annual change in size			
Absolute value	33	65	379,238
Decreases	–31	49	191,562
Increases	37	79	180,114

*Note.* School-years are the unit of analysis. School size is measured by total school enrollment. The final sample includes all U.S. public schools in which the highest grade was fourth, fifth, sixth, seventh, or eighth grade. A total of 72,771 unique primary schools operated in the United States between 2004 and 2010. TPS = traditional public school.

## Acknowledgments

We thank the North Carolina Education Research Data Center (NCERDC) for providing the data. We also thank Laura Crispin, Peter Hinrichs, Ben Ost, David Pitts, and Jeffrey Weinstein for helpful comments. Michael Hayes provided excellent research assistance. Any remaining errors are our own.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Gershenson gratefully acknowledges financial support from an American University Faculty Research Support Grant.

## Notes

1. There was a well-organized and well-funded “small high schools” movement in several urban school districts, most notably in New York City and Chicago, during the 1990s and 2000s (Schwartz et al., 2013). The movement was funded by the Gates, Carnegie, and Annenberg Foundations and the U.S. Department of Education (Barrow et al., 2013). However, because the movement focused entirely on reducing the size of high schools, it is only tangentially related to the present study. See U.S. Department of Education (2006), Ravitch (2011), and Shear and Smerdon (2003) for further discussion of the modern “small high schools” initiative.

2. Source: Authors’ calculations using the Common Core of Data. See Appendix Table A1.

3. Formal instructions for requesting the data are publicly available on NCERDC’s website. See <https://ssri.duke.edu/data-it-services/north-carolina-education-research-data-center-ncercd>.

4. See <http://www.schoolclimate.org/> for additional information.

5. The end-of-grade tests are state mandated, criterion referenced, and vertically aligned. We standardize test scores by grade, year, and subject to have mean 0 and  $SD$  1 (Ballou, 2009).

6. Third graders and academic-year 2003 observations do not appear in the estimation sample, as these data are used to create lag scores and gain scores for use in the value-added models.

7. The CCD is publicly available here: <http://nces.ed.gov/ccd/>.

8. Test-score means and  $SD$  are not precisely 0 and 1, respectively, because they were standardized using all available test scores.

9. This estimate was generated by regressing total enrollment on grade span and a full set of school indicators. Adding year indicators to the regression does not significantly change the point estimate on grade span.

10. The estimates reported in columns 2 to 10 of Table 2 are robust to conditioning on year fixed effects and to dropping schools that changed grade spans from the analytic sample.

11. Rothstein (2010) finds evidence of nonrandom student–teacher assignments in North Carolina.

12. Estimated coefficients for the full set of covariates included in the lag-score specification of column 1 are generally of the expected sign and are reported in Online Appendix Table B2 (see the online appendix at <http://epa.sagepub.com/supplemental>).

## References

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67, 251–333.

- Akerlof, G. A., & Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40, 1167–1201.
- Alexander, K. L., Entwisle, D. R., & Kabbani, N. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record*, 103, 760–822.
- Altonji, J. G., & Mansfield, R. (2011). The role of family, school, and community characteristics in inequality in education and labor-market outcomes. In G. Duncan & R. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children’s life chances* (pp. 207–228). New York, NY: Russell Sage Foundation.
- Anderson, C. S. (1982). The search for school climate: A review of the literature. *Review of Educational Research*, 52, 368–420.
- Andrews, M., Duncombe, W., & Yinger, J. (2002). Revisiting economics of size in American education: Are we any closer to a consensus? *Economics of Education Review*, 21, 245–262.
- Angrist, J. D., & Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston’s Metco Program. *The American Economic Review*, 94, 1613–1634.
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricists’ companion*. Princeton, NJ: Princeton University Press.
- Ballou, D. (2009). Test scaling and value-added measurement. *Education Finance and Policy*, 4, 351–383.
- Barrow, L., Claessens, A., & Schanzenbach, D. W. (2013, March). *The impact of Chicago’s small high school initiative* (Working Paper No. 18889). Cambridge, MA: National Bureau of Economic Research.
- Baydar, N., & Brooks-Gunn, J. (1991). Effects of maternal employment and child-care arrangements on preschoolers’ cognitive and behavioral outcomes: Evidence from the children of the National Longitudinal Survey of Youth. *Developmental Psychology*, 27, 932–945.
- Berry, C. R. (2006). School consolidation and inequality. *Brookings Papers on Education Policy*, 9, 49–75.
- Berry, C. R., & West, M. R. (2010). Growing pains: The school consolidation movement and student outcomes. *Journal of Law, Economics, & Organization*, 26, 1–29.
- Bloom, H. S., Levy Thompson, S., & Unterman, R. (2010). *Transforming the high school experience*. New York, NY: MDRC.
- Bloom, H. S., & Unterman, R. (2014). Can small high schools of choice improve educational prospects for disadvantaged students? *Journal of Policy Analysis and Management*, 33, 290–319.

- Boccardo, J., Schwartz, A. E., Stiefel, L., & Wiswall, M. (2013, November 8–10). *Beyond academics: Do small schools have better learning environments?* Presentation at the APPAM 2012 Fall Conference, Baltimore, MD.
- Brummet, Q. (2014). The effect of school closings on student achievement. *Journal of Public Economics*, 119, 108–124.
- Carruthers, C. K. (2012). New schools, new students, new teachers: Evaluating the effectiveness of charter schools. *Economics of Education Review*, 31, 280–292.
- Chen, G. (2007). School disorder and student achievement: A study of New York City elementary schools. *Journal of School Violence*, 6, 27–43.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26, 673–682.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120.
- Common Core of Data. (2011). *Numbers and types of public elementary and secondary schools from the common core of data: School year 2009–2010*. National Center for Education Statistics. Retrieved from [http://nces.ed.gov/pubs2011/pessschools09/tables/table\\_01.asp](http://nces.ed.gov/pubs2011/pessschools09/tables/table_01.asp)
- Conant, J. B. (1959). *The American high school today: A first report to interested citizens*. New York, NY: McGraw Hill Book.
- Cotton, K. (1996). *School size, school climate and student performance*. Portland, OR: Northwest Regional Educational Laboratory.
- Crispin, L. (2012). *School size and student achievement: Does one size fit all?* Unpublished Manuscript, Ohio State University, Columbus.
- Currie, J., & Thomas, D. (2001). Early test scores, socioeconomic status, school quality, and future outcomes. *Research in Labor Economics*, 20, 103–132.
- Dee, T. S., Ha, W., & Jacob, B. A. (2006). The effects of school size on parental involvement and social capital: Evidence from the ELS: 2002. *Brookings Papers on Education Policy*, 9, 77–97.
- Duncombe, W., & Yinger, J. (2007). Does school district consolidation cut costs? *Economics of Education Review*, 22, 193–201.
- Epple, D., & Romano, R. (2011). Peer effects in education: A survey of the theory and evidence. In J. Benhabib, A. Bisin, & M. Jackson (Eds.), *Handbook of social economics* (Vol. 1B, pp. 1053–1163). Amsterdam: North-Holland, Elsevier.
- Gershenson, S. (2013). Do summer time-use gaps vary by socioeconomic status? *American Educational Research Journal*, 50, 1219–1248.
- Goldhaber, D., Cowan, J., & Walch, J. (2013). Is a good elementary teacher always good? Assessing teacher performance estimates across subjects. *Economics of Education Review*, 36, 216–228.
- Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, 80, 589–612.
- Gottfredson, D. C., & DiPietro, S. M. (2011). School size, social capacity, and student victimization. *Sociology of Education*, 84, 69–89.
- Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, 10, 117–156.
- Guryan, J., Hurst, E., & Kearney, M. (2008). Parental education and parental time with children. *The Journal of Economic Perspectives*, 22(3), 23–46.
- Haller, E. J. (1992). High school size and student indiscipline: Another aspect of the school consolidation issue? *Educational Evaluation and Policy Analysis*, 14, 145–156.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2009). New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27, 349–383.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100, 267–271.
- Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4), 85–108.
- Jacob, B. A., Lefgren, L., & Sims, D. P. (2010). The persistence of teacher-induced learning. *Journal of Human Resources*, 45, 915–943.
- Johnston, S. L. (2009). Improving the school environment to reduce school violence: A review of the literature. *Journal of School Health*, 79, 451–465.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114, 497–532.
- Kuziemko, I. (2006). Using shocks to school enrollment to estimate the effect of school size on student achievement. *Economics of Education Review*, 25, 63–75.
- Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina. *Journal of Policy Analysis and Management*, 33, 162–187.
- Lamdin, D. J. (1995). Testing for the effect of school size on student achievement within a school district. *Education Economics*, 3, 33–42.



- Lauen, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis, 34*, 185–208.
- Lee, V. E., & Loeb, S. (2000). School size in Chicago elementary schools: Effects on teachers' attitudes and students' achievement. *American Educational Research Journal, 37*, 3–31.
- Leithwood, K., & Jantzi, D. (2009). A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research, 79*, 464–490.
- Leung, A., & Ferris, J. S. (2008). School size and youth violence. *Journal of Economic Behavior and Organization, 65*, 318–333.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*, 531–542.
- McMillen, B. J. (2004). School size, achievement, and achievement gaps. *Education Policy Analysis Archives, 12*(58), 1–24.
- Measures of Effective Teaching. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching project*. Seattle, WA: Bill & Melinda Gates Foundation.
- Mittag, N. (2012). *New methods to estimate models with large sets of fixed effects with an application to matched employer-employee data from Germany* (FDZ-Methodenreport). Nuremberg, Germany: Institute for Employment Research.
- Muschkin, C., Bonneau, K., & Dodge, K. (2011). *North Carolina education research data center: Final report to the Spencer Foundation* (Grant# 200300138). Retrieved from [http://childandfamilypolicy.duke.edu/pdfs/projects/NCERDC\\_SpencerFoundationReport.pdf](http://childandfamilypolicy.duke.edu/pdfs/projects/NCERDC_SpencerFoundationReport.pdf)
- Ravitch, D. (2011). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.
- Ready, D. D., & Lee, V. E. (2006). Optimal context size in elementary schools: Disentangling the effects of class size and school size. *Brookings Papers on Education Policy, 9*, 99–135.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics, 125*, 175–214.
- Schoeneberger, J. A. (2012). Longitudinal attendance patterns: Developing high school dropouts. *The Clearing House, 85*, 7–14.
- Schwartz, A. E., Stiefel, L., & Wiswall, M. (2013). Do small schools improve performance in large, urban districts? Causal evidence from New York City. *Journal of Urban Economics, 77*, 27–40.
- Shear, L., & Smerdon, B. (2003). *Mapping the terrain: Year 1 of the evaluation of the Bill & Melinda Gates Foundation's National School District and Network Grants Program*. Retrieved from [http://smallhs.sri.com/documents/Mapping\\_the\\_Terrain.pdf](http://smallhs.sri.com/documents/Mapping_the_Terrain.pdf)
- U.S. Department of Education. (2006). *Smaller Learning Communities program*. Washington, DC: Author.
- Welsh, W., Stokes, R., & Greene, J. R. (2000). A macro-level model of school disorder. *Journal of Research in Crime & Delinquency, 37*, 243–283.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: MIT Press.
- Wynne, E., & Ryan, K. (1997). *Reclaiming our schools: Teaching character, academics, and discipline*. Upper Saddle River, NJ: Merrill.

### Authors

SETH GERSHENSON is an assistant professor of public policy at American University. His research focuses on the causes, consequences, and potential remedies of socioeconomic and demographic achievement gaps in K–12 education.

LAURA LANGBEIN is a professor of public policy at American University. Her publications include studies of bureaucratic discretion, pay-for-performance, intrinsic motivation, and corruption, with applications in environment and education policy, and a textbook on program evaluation.

Manuscript received September 30, 2013

Revision received September 5, 2014

Accepted January 23, 2015