Optimal Context Size in Elementary Schools: Disentangling the Effects of Class Size and School Size

---

DOUGLAS D. READY AND VALERIE E. LEE

Young children's learning—and how their learning is distributed by social background—may be influenced by the structural and organizational properties of their school. This study focuses on one important structural dimension of these educational contexts: *size*. Over the past several decades, various elements of the size of educational contexts have become a major focus of researchers, politicians, and corporate leaders. Billions of public and private dollars have been invested in reforms to reduce the size and scope of both classrooms and schools. Unlike many educational reform initiatives, these downsizing efforts have found support from virtually every quarter. A united front of stakeholders has coalesced behind the notion that "smaller is better." Although size-reduction policies are well intentioned, their effectiveness is unclear, and some efforts have produced unintended and even undesirable consequences. Moreover, their cost-effectiveness has seldom been considered.

Based on results from the famous Tennessee class-size experiment, California invested billions of dollars encouraging its schools to limit classes in the early grades to no more than twenty students. Quite recently, the push to reduce the size of high schools has been accompanied by enormous financial support from foundations and the federal government in an effort to encourage schools-within-schools, small learning communities, and small stand-alone schools. Curiously, these important policy initiatives—reduced class size and reduced school size—have not been simultaneously considered within elementary school contexts. However, the effects of class size may be a function of school size, the effects of school size may be a function of class size, or both. The lack of research that simultaneously considers these potentially related elements of size is somewhat surprising.

Despite the groundswell of public support for smaller educational settings, the empirical base regarding the confounding effects of various components of elementary school size remains quite sparse, particularly if only methodologically sound studies are considered. Moreover, crafting size-reduction polices that faithfully reproduce the findings of experimental and quasi-experimental studies is a challenging task. In short, efforts to reduce various elements of size in elementary schools may be an instance where policy is far in front of research.

Background

Determining how the size of educational contexts may influence student outcomes can be conceptualized and measured at multiple levels. Decisions regarding the appropriate unit of analysis are important, as each level may uniquely influence student learning. It seems logical to assume that the social and structural consequences of size would be strongest where they most directly affect the daily activities of teaching and learning. For example, at the elementary school level, a focus on class size seems most reasonable. Unlike high school students, elementary school students spend more time in a single classroom. However, non- and quasi-experimental studies of elementary school class size rarely account for school size—clearly a problem, as class size may be a function of school size. Moreover, as most high schools contain the same grades (nine through twelve), examining the effects of school size on student outcomes in those contexts seems quite appropriate. Conversely, the grade spans that elementary schools include vary widely, with K–3, K–6, and K–8 schools all relatively common. If elementary schools contain fewer grades (K–3, for example), each grade is likely to include more students and classes. Thus grade size may be an additional element of context size in elementary schools.

Unfortunately, the research on these separate (yet related) elements of elementary school size is generally quite weak. Apart from the recent class-size experiments in Tennessee and Wisconsin, research in this area generally employs small and nonrepresentative samples, relies on cross-sectional data, and suffers from numerous other methodological limitations. Moreover, the theoretical justifications behind these studies often rest on literature reviews. One strange result is a circular chain, wherein literature reviews often cite other literature reviews rather than solid empirical studies. In one sense this is understandable, given the scarcity of high-quality research on the topic. Our review focuses on class size in elementary schools, as the research on other components of elementary school context is limited in both quality and quantity.

Research on Class Size

In 1985 Tennessee initiated a longitudinal class-size reduction experiment that would serve as the foundation for similar efforts across the country.[1] The experiment, titled Project STAR (Student/Teacher Achievement Ratio), randomly assigned several thousand kindergartners to one of three within-school experimental conditions: a small class enrolling between thirteen and seventeen children, a large class enrolling between twenty-two and twenty-six children with a single teacher, or a large class with a teacher and an aide. At the end of kindergarten, the achievement of children in small classes was almost one month ahead of the achievement of children in the other two classroom conditions; by the end of first grade, the same children were almost two months ahead (ES = 0.2–0.25 standard deviation).

Although Project STAR is generally considered the premier educational study with a randomized design in contemporary educational research, the study has garnered some criticism.[2] Because participation in STAR required at least three classrooms at each grade level—a small class, a large class, and a large class with an aide—only larger schools participated in the study. Moreover, student attrition from the treatment group was substantial and potentially nonrandom: only 48 percent of the original treatment group participated through third grade, and children who left the sample may have been lower achieving.[3] Teachers with smaller classes were also aware that they were part of the intervention group. Not only did many teachers enter the study already convinced that smaller classes were superior, but the state was simultaneously considering universal class-size reductions.[4] In this sense, such teachers may have induced experimenter expectancies.[5] However, class-size effects may be *under*estimated in the Tennessee experiment, as "large" classrooms enrolled only twenty-six students. The nationally representative ECLS-K data we employ in this study indicate that a substantial proportion of U.S. elementary schools offer classes enrolling more than twenty-six students.

In 1996 Wisconsin launched a similar (although more modest) class-size reduction experiment titled SAGE (Student Achievement Guarantee in Education). Unlike STAR, the SAGE design was randomized between, not within, schools. Kindergarten through third-grade classrooms in SAGE schools enrolled only fifteen students, compared to classrooms of twenty-one to twenty-five in the control schools.[6] Wisconsin's program differed from Tennessee's in another way: it targeted low-income schools—both SAGE and control schools enrolled substantial numbers of children living in poverty. Despite these differences in design and study participants, findings from the SAGE program are comparable to those from the Tennessee study: children in SAGE schools experienced higher achievement gains than their control school counterparts (ES = 0.2 standard deviation).[7]

In 1996 California used the STAR findings to justify a program that offered districts $650 for every child enrolled in a classroom with twenty or fewer students. In general, evaluations of California's efforts have been formative rather than summative. Unlike the evaluations of the class-size initiatives in Tennessee and Wisconsin, the California design was not experimental. All districts were permitted to receive funds and reduce class sizes simultaneously, rendering meaningful evaluation virtually impossible, as comparison groups were not available. For example, on average, by the end of third grade, children in the "treatment group" were enrolled in smaller classes for only one year more than those in the "control group." Moreover, selection bias was quite apparent, in that low-income schools were the last to implement smaller classes, despite the financial incentives for doing so. Even if these critical design flaws are ignored, estimating the relationship between student learning and class size would not be possible—the data permit only cross-sectional comparisons, as students' cognitive skills were not assessed in the early grades. Although evaluators report class-size "effects," we agree with their judgment that findings regarding student achievement are "inconclusive."[8]

As policy interest in the size of educational contexts increases, it is important to evaluate size-reduction efforts on two additional criteria: cost and unintended consequences. First, as class-size reduction programs are quite expensive, school districts and taxpayers are (rightly) interested in whether such costly investments are educationally sound.[9] California currently spends more than $1.6 billion a year on its efforts to reduce class enrollments below twenty—a number still larger than the ideal class size identified in the Tennessee and Wisconsin experiments. Although policies that seek to reduce class size are very popular among teachers and parents, the educational return on such a substantial investment remains unclear.

Second, several unintended and undesirable consequences accompanied California's class-size-reduction policy. By definition, large-scale, class-size-reduction programs require many more teachers, and California did not have a surplus of qualified teachers. As such, many districts hired teachers lacking full credentials to staff new classrooms, a practice that runs counter to the "highly qualified teacher" provisions within the federal No Child Left Behind legislation. Prior to class-size reduction, only 1.8 percent of California's K–3 public school teachers were uncertified; by the second year of the program, 12.5 percent lacked full credentials. Moreover, schools serving socioeconomically disadvantaged students were disproportionately forced to hire uncertified and inexperienced teachers.[10]

Another unintended, but serious, consequence in California flowed from the need to create 18,000 additional classrooms virtually overnight. Already crowded low-income districts often had inadequate facilities to accommodate new classrooms.[11] Many schools and districts not only adopted year-round calendars, but also transformed teacher lounges, gymnasiums, auditoriums, libraries, labs, special education facilities, and even storage rooms into classrooms. Again, these issues are not particular to California. Almost 60 percent of large school districts nationally that received federal class-size-reduction funds reported difficulty locating adequate classrooms for their new teachers.[12]

Research on School Size

Extant and relevant empirical studies of school size are typically characterized by a host of problems—defined in terms of level, outcomes, design, and quality. Regarding the first problem—level—almost all school-size studies have focused on high schools. It is unclear whether research findings regarding high school size are generalizable to elementary schools. The second problem—outcomes—refers to the fact that most research on size is not longitudinal, relying on simple correlations between school size and student achievement status, rather than achievement growth over time. Achievement status is quite different from learning.

A third general problem concerns study design. Not only are the majority of school-size studies cross-sectional rather than longitudinal, but most assume that the relationship between school size and student outcomes is linear. In our own research, we document a distinctly nonlinear relationship between high school size and student learning.[13] Another common, but fundamental, design flaw is that almost no research on school size recognizes that questions regarding school size and student outcomes are multilevel. Thus the large majority of school-size research examines the relationship with aggregate data (that is, size effects on school-average achievement). This approach ignores the fact that size may differentially influence learning, based on students' social background. Moreover, size effects may interact with such basic school characteristics as racial or social class composition.[14]

Using multilevel methods and a longitudinal design, Lee and Smith conclude that achievement gains are largest in medium-size high schools (600–900 students), although schools with somewhat smaller enrollments are more equitable, in terms of the relationship between social background and achievement gains.[15] Although the same size range is ideal in schools differentiated by their concentration of minorities and socioeconomic status (SES), size has stronger effects on student learning in schools educating less-advantaged populations. In another study, we explore how the size of Chicago's K–8 elementary schools influences achievement gains for seventh and eighth graders, both directly and through teachers' attitudes. That study finds favorable effects for smaller elementary schools (below 400 students) but no differences between medium and large schools (more than 750 students).[16] Moreover, teachers' willingness to take responsibility for their students' learning is greater in smaller schools. A more thorough and complete review of the literature on school size is available in the paper by Darling-Hammond, Ross, and Milliken in this volume.

Research on Grade Span

Another area that has received little empirical scrutiny is grade span, a concept that describes how many and which grades are included within a single school. There are both structural and philosophical reasons arguing for narrow versus broad grade spans. Much of the literature on grade span focuses on middle and junior high schools, neglecting elementary school configurations, where the construct is equally valid. The decision about what grades to include in which schools is generally guided by matters of practical necessity rather than educational value. The size of existing buildings, enrollments, and fiscal resources determines grade spans more often than thoughtful attention to children's social and academic needs.[17]

Our interest in grade span is twofold. First, grade configurations influence the social and academic characteristics of schools. Socially, broader grade spans within schools create opportunities for older children to act as role models for younger peers.[18] These opportunities may occur both informally (for example, at recess and in the cafeteria) and formally (for example, through "reading buddy" and other activities). Academically, school principals craft goals for their school based partly on the grades the school serves. Principals in K–8 and K–12 schools are more likely to stress higher-order thinking over basic skills than principals in schools enrolling lower grades.[19] Moreover, broader grade spans facilitate teacher communication across grades, matching pedagogical strategies and expectations to children's developmental stages.[20]

Second, the grade configuration of elementary schools influences the number of children within each grade. Schools serving fewer grades typically have more students and classes per grade (for example, K–3); schools serving many grades typically have fewer students per grade (for example, K–8 and K–12). The same mechanisms may link school size and grade size to student outcomes. For example, schools that enroll more students per grade are more likely to sort students into tailored academic programs or even academically homogeneous classrooms, thus increasing the odds that children's learning will be stratified academically and socially. Indeed, some research uses grade size as a proxy for school size.[21] However, we choose to maintain an important distinction between these two elements of elementary school size.

Summary of Research on Size

In general, extant research favors smaller educational contexts, defined both in terms of school size and class size. However, the strands of research examining class size and school size are curiously independent and seldom combined into a single study. Despite extensive literature on school size in high schools and class size in elementary schools, these bodies of research do not inform one another. Research on *school* size focuses almost exclusively on secondary schools, whereas research on *class* size focuses entirely on elementary schools (and really at the lowest grades). It seems reasonable to assume that these size elements are related in U.S. elementary schools, despite the paucity of research exploring the connection.

Although policymakers have recently embraced a strong advocacy of small high schools (not necessarily with empirical support), research on high school size may not be applicable to elementary schools. And despite the well-designed Tennessee experiment, many nagging issues challenge the documented relationship between several elements of organizational size and student learning in elementary school. Although teachers at all levels favor smaller classes, basic issues of educational cost and efficiency cannot be ignored. In the policy arena, the size dimensions we consider in this study are amenable to direct policy manipulation. People who work in schools should recognize how the various elements of size work together—rather than thoughtlessly embrace the mantra that smaller is better. School practitioners, policymakers, and taxpayers may rightly ask, Better for whom? How small is "small"? Possibly better for some but harmful for others? In this study we bring together what are currently quite disparate strands of research to address some of these questions.

Research Questions

Our exploration of these issues differs from extant studies linking size to student outcomes in four important respects. First, we focus on elementary school size. Second, we conceptualize the size of educational contexts quite broadly, focusing on the relative impacts of *class size* and *school size,* while simultaneously accounting for *grade span*. Third, we explore the effects of these structural characteristics of elementary schools on both learning and the equitable distribution of that learning by children's social background, particularly race or ethnicity and socioeconomic status. Fourth, our research design provides considerable methodological leverage with which to disentangle the confounding effects on student learning of student background and the size of elementary school contexts.

The paper is organized around three research questions:

—*Effects on learning trajectories.* How can we characterize the relationship between elementary class size, school size, and student learning in reading and mathematics over the kindergarten and first-grade years? Of particular interest is whether class size is related to student learning once we account for school size, and vice versa.

—*Size effects on the social distribution of learning.* Do the effects of school and class size differ for children of different social backgrounds? If so, are smaller classes and schools more important for more disadvantaged students?

—*Changes in size effects over time and subject.* To what extent do the effects of these various elements of size differ between kindergarten and first grade and between learning in literacy and in mathematics? In other words, are certain elements more important in kindergarten than in first grade or for the development of literacy rather than mathematics skills?

Method of Research

This study is located within a type of research called "school effects," which investigates how school characteristics influence student outcomes. Most school-effects research centers on high schools. However, this type of research in elementary schools flows from a seminal study by Barr and Dreeben.[22] A few recent studies also focus on elementary school effects.[23] The school-effects tradition capitalizes on a basic notion in education: nesting. That is, students are nested in classrooms, and classrooms are nested in schools. At each level of nesting, different policies and practices influence students' experiences. In this study, we conceptually and analytically nest students within schools. Although we could logically have the classroom as the unit of analysis, we do not pursue this approach for three reasons. First, two of the three size dimensions (school size and grade span) are school-level phenomena. Second, class size is typically a function of school enrollments and district policies; class sizes within schools vary little. Third, the structure of the data we use does not support the classroom as a separate unit of analysis.

Data

In this study, we employ data from the Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K). Sponsored by the National Center for Education Statistics (NCES), these data are ideal for studying how organizational size influences children's learning, particularly with the statistical methods discussed below. The ECLS-K collection of base-year (1998) data followed a stratified design structure. The primary sampling units were geographic areas consisting of counties or groups of counties from which about 1,000 public and private schools offering kindergarten programs were selected. A target sample of about twenty-four children was then drawn from each school. In this chapter, we employ the first four data waves of ECLS-K, which include information on the same children in the fall and spring of kindergarten (waves 1 and 2) and the fall and spring of first grade, with a random subsample in the fall (waves 3 and 4). Beyond testing children with one-on-one untimed achievement tests at each wave, data were also collected from parents through structured telephone interviews, from each child's teacher, and from schools.[24] These rich data allow researchers to capture a longitudinal picture of a recent, large, and nationally representative cohort of young children as they move through elementary school.

Growth Curves within an HLM Framework

We employ hierarchical linear modeling (HLM) within a three-level growth-curve framework.[25] Specifically, we nest learning trajectories within children, who are nested within schools. Our level-1 HLM models estimate children's individual learning trajectories. At level 2, we model these learning trajectories as functions of children's social and academic background. At level 3—the focus of this study—we estimate the effects of organizational size on children's learning.

AN ALTERNATE GROWTH-CURVE APPROACH. Quantitative researchers traditionally have used analysis of covariance (ANCOVA) or gain-score models to measure change over time within individuals. Over the past several decades, however, social scientists have concluded that estimating change based on only two data points is inherently inadequate.[26] Myriad statistical and substantive issues have driven this methodological shift, although one central concern is shared: traditional approaches assume that variance in the outcome remains steady over time. This assumption *itself* implies that growth trajectories among individuals are perfectly parallel, "an entirely unrealistic state of affairs [that] is obvious even at the most casual glance."[27]

As an alternative approach, educational researchers are increasingly using three or more data points to model growth rates and learning trajectories. Such analyses entail both within-individual and between-individual components.[28] The first analytic phase estimates the growth rates of individuals, while the second phase focuses on the detection and *explanation* of systematic variance in individual growth rates.[29] An endless array of potential explanatory covariates exists, including the characteristics of individual children, their classrooms and teachers, schools, peers, and neighborhoods.[30] Our examination of the relationship between components of elementary school size and cognitive growth falls within this relatively new analytic framework.

CONCEPTUALIZING TIME. The ECLS-K data present a unique challenge to researchers interested in modeling children's cognitive growth over time. Longitudinal studies of student learning generally consider the timing of events as constant across cases (that is, "third grade" represents an identical value or construct). However, the dates on which the ECLS-K cognitive assessments were administered vary considerably across children, both within and between schools. This is understandable given the enormity of the data collection involved with ECLS-K and the time required for each one-on-one assessment. In addition to variability in testing dates, the starting and ending dates of academic years vary across schools.

The result of this variability in school exposure at each assessment is that children's opportunities to learn differed both within and between schools. For example, the time children were in school between the fall and spring kindergarten assessments ranged from almost four to over eight months, averaging about six months (although the school year is nine months). For some children, the fall assessments took place months into the school year and the spring assessments occurred several months before the end of the school year. As such, the assessments do not represent comparable events in time across children. Further complicating the analyses, children were in school for approximately half of the "summer vacation" between the spring kindergarten and fall first-grade assessments. Considering the rapid learning rates among young children, researchers who employ the ECLS-K data must take these concerns into account.[31]

Despite these analytic challenges, the structure of the ECLS-K data provides a unique methodological opportunity. Our level-1 models include three time-varying covariates that indicate individual children's exposure to school at each assessment: months of exposure to kindergarten, months of exposure to summer between kindergarten and first grade, and months of exposure to first grade. For example, at the time of the first assessment the average child had been "exposed" to more than two months of kindergarten, but zero months of summer and zero months of first grade. With the second assessment, the average child had experienced more than eight months of kindergarten but had not been exposed to summer or first grade. At the third assessment, the average child had been exposed to 9.5 months of kindergarten (a full year), 2.7 months of summer (the traditional summer vacation), and more than a month of first grade. At the point of the fourth and final assessment, the average child had been exposed to 9.5 months of kindergarten, 2.7 months of summer, and more than eight months of first grade.

These three measures of school exposure—each linked to the four assessment dates—allow us to model four distinct parameters: *initial status*, or children's achievement as they began kindergarten (literally, predicted achievement with exposure to zero days of kindergarten, zero days of summer, and zero days of first grade). Rather than initial *status*, the three remaining parameters are linear *learning* rates or slopes over *the kindergarten year*, *the summer between kindergarten and first grade*, and *the first-grade year*. Again, the variability in testing dates permits this "slopes as outcomes" approach, where the slopes are adjusted for exposure to school. An additional benefit of this approach is that at each analytic level, all coefficients are in an easily interpretable metric: points of learning *per month* in kindergarten, summer, and first grade. We also present our class-size and school-size estimates in effect-size (standard deviation) units. In this chapter, we focus on the estimates obtained from the kindergarten and first-grade parameters, which address our questions regarding the influence of educational size on student learning.[32]

LEARNING PATTERNS IN ECLS-K. As our focus in this study is on learning rather than achievement, we explore learning patterns from children's entry into kindergarten to the end of first grade. Figure 1 displays young children's observed learning trajectories during this period. As these learning trajectories are constructed from test scores, in the metric of the IRT (item response theory)-equated number-right scores available in ECLS-K, the scales differ for reading and math. Panel A depicts learning in reading; panel B depicts learning in mathematics. On average, the first testing occurred in late October 1998, the second in early May 1999, the third in early October 1999, and the fourth in early May 2000. Thus the observed trajectories begin and end on those dates.

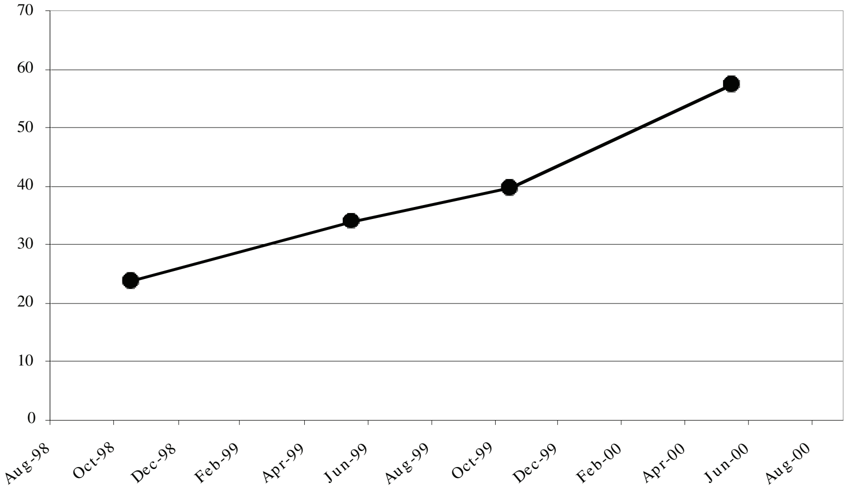
The panels in figure 1 suggest three similar trends for both subjects. First, we see a pattern of continual growth (or learning). Young children's achievement in these subjects increases steadily. Second, we see a slight decline in the learning slope over the summer between kindergarten and first grade, more so in reading than in math. Third, we see a slightly higher learning slope in first grade than in kindergarten. These slope variations are more noticeable in reading than in math.

Recall that our basic HLM growth-curve models estimate four learning parameters in each subject: initial status, monthly achievement gain over the kindergarten year, monthly gain over the summer, and monthly gain over the first-grade year. Figure 2 displays these patterns, in which achievement growth is adjusted for differences between when schools actually opened and closed and when the tests were administered to each child. The slopes of the lines in figure 2 may be expressed in a monthly learning metric.

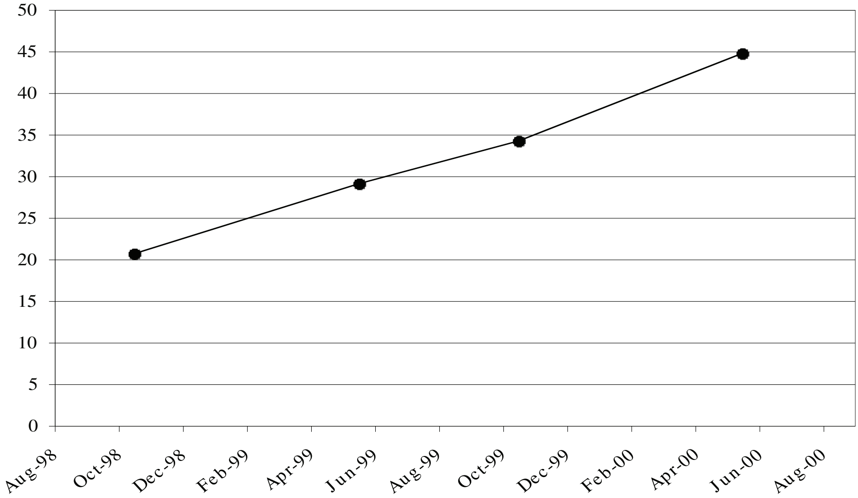
Estimates of the four parameters of interest—initial status and the three gain slopes—from figure 2 indicate that results in figure 1 are misleading in three ways. First, initial status (achievement at entry into kindergarten) is overestimated in figure 1, as the first testing was nearly two months (sometimes more) into the school year. Estimated initial status in September, from figure 2, is a few test points lower than the observed achievement several weeks later. Second, the time gap between the second and third assessments in figure 1 is too wide to isolate true summer learning (that is, only those months when school is not in session), as both assessments were made during the school year. In figure 2, learning trajectories extend across the entire kindergarten and first-grade school years, coinciding with the average time of school closing at the end of kindergarten and opening at the beginning of first grade. In-school time periods average 9.5 months; summer averages 2.7 months. Third—and most important—the slopes of the lines for the three growth parameters in figures 1 and 2 are different, especially during the summer. The estimated learning slopes are slightly steeper for the in-school periods in figure 2 than in figure 1. The most striking comparison shows that the estimated summer learning slope

Figure 1. ECLS-K Overall Learning Trajectories Based on Average Testing Times

Panel A. Reading achievement (HLM estimated scores)



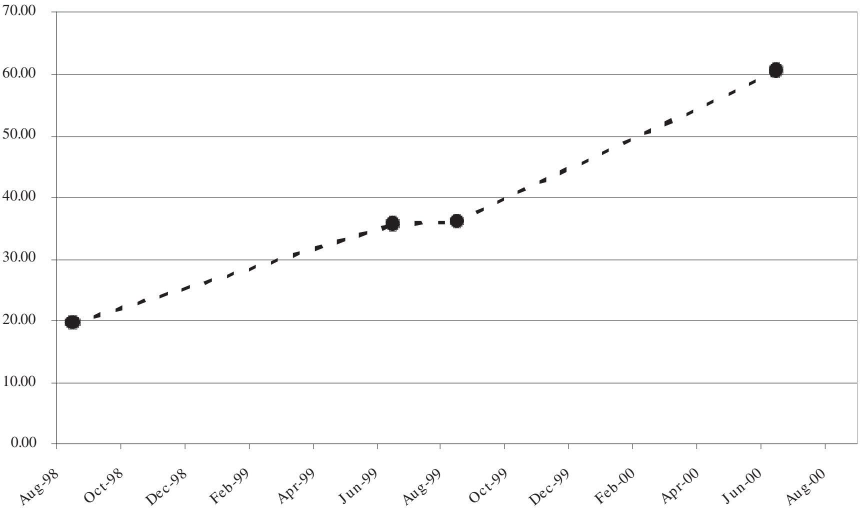
Panel B. Math achievement (HLM estimated scores)



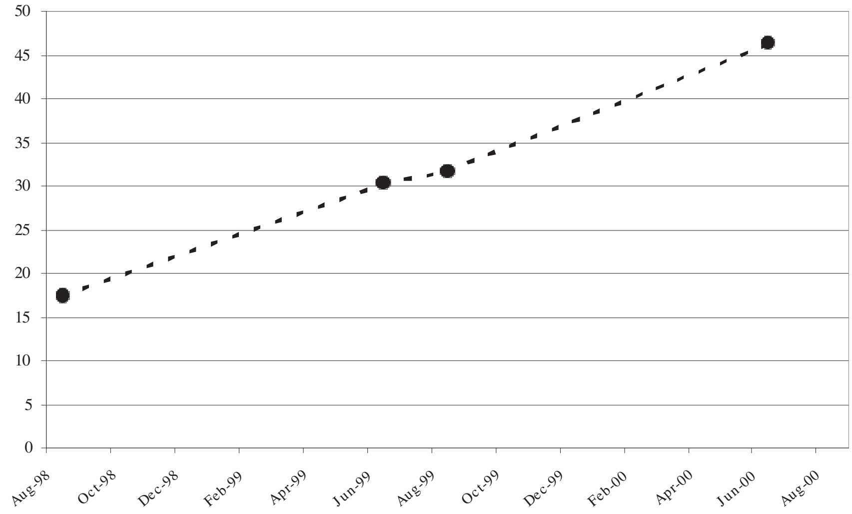
Source: Authors' calculations using ECLS-K data.
HLM = hierarchical modeling.

Figure 2. Estimated Overall Learning Trajectory, in School and out of School

Reading achievement (HLM estimated scores)



Math achievement (HLM estimated scores)



Source: Authors' calculations using ECLS-K data.
HLM = hierarchical modeling.

for reading in figure 2 is nearly flat when we model true out-of-school learning (rather than learning between the two assessment time points).

WEIGHTS. Because ECLS-K used a multistage stratified sampling design, the data include a series of design weights. As with other longitudinal NCES data sets, analyses using ECLS-K require the use of weights to compensate for unequal probabilities of selection within and between schools (for example, the intentional oversampling of Asian and Pacific Islander children) and nonresponse effects. Although our growth-curve models consider achievement at four waves of the ECLS-K data, the "1234" panel ECLS-K weights are only defined for children in the sample at time 3. Hence the use of those weights automatically restricts the sample to that small subgroup of children and schools with data at the beginning of first grade. Instead, these analyses employ the "124" panel weights, which retain the larger sample. Our descriptive and analytic analyses employ a child-level weight (C124CW0) to compensate for differential sampling both within and between schools. We use the ECLS-K school-level weight (S2SAQW0) with our school-level descriptive and multilevel analyses. Both weights are normalized to a mean of 1 to reflect the actual (smaller) sample sizes.

ANALYTIC SAMPLE. From the full ECLS-K sample, we constructed our analytic sample in two stages. First, we selected children who had a nonmissing weight, remained in the same school in kindergarten and first grade, advanced to the first grade following the 1998–99 kindergarten year, had complete data on gender, race or ethnicity, and socioeconomic status, and had test scores for at least three of the four literacy and mathematics assessments. We then selected schools that had a nonmissing weight, were not year-round schools, were public schools that offered kindergarten and first grade, and enrolled at least five ECLS-K children.[33] Our final analytic sample includes 25,545 literacy and 25,545 mathematics test scores nested within 7,740 children, who are nested within 527 public schools. An analysis of missing data revealed that our subsample is somewhat more socioeconomically advantaged than the full ECLS-K sample, with fewer language-minority children and fewer children from the lowest SES quintile. The loss of lower-SES and language-minority children mostly occurred when the sample was restricted by available test scores, as all testing was in English.

Measures

A central task of this study is to consider the most fruitful way to conceptualize various elements of public elementary school size. Although we had intended to include grade cohort size as a separate element, we found that kindergarten and first-grade cohort size was highly correlated with school size ($r = 0.75$). This is reasonable, given that elementary schools with larger enrollments generally enroll more students at each grade. However, this finding means that grade cohort and school size are not independent constructs. The focus of our study is on the effects of class size and school size on children's learning in the early grades.

CLASS SIZE. Because ECLS-K sampled only a modest number of children per school, most within-classroom sample sizes are quite small. This sampling design precluded our ability to conceptualize the classroom as a separate unit of analysis. Using an HLM fully unconditional model, we found that the vast majority (over 85 percent) of variance in class size is between (rather than within) schools. That is, classes at the same grade within the same school were very likely to enroll close to the same number of children. Thus we decided to consider class size as school-level aggregates (that is, separate averages of the kindergarten and first-grade class sizes in each school). Based on the Tennessee class-size parameters, we designated classes enrolling seventeen or fewer children as "small classes." For reasons discussed below, we designated classes enrolling twenty-five or more children as "large classes." In our multivariate HLM analyses, we compare schools with these large and small classes to those with medium-size classes (between seventeen and twenty-five students).

Neither the Tennessee nor the Wisconsin class-size experiments examined medium-size classrooms. Classrooms with enrollments between seventeen and twenty-two did not participate in Project STAR, and Wisconsin's SAGE program involved no classrooms enrolling between fifteen and twenty-one students. This is quite understandable, in that these evaluations sought to maximize their ability to identify class-size effects. However, the nationally representative ECLS-K data we employ in this study indicate that, in roughly half of all public schools, medium-size kindergarten and first-grade classrooms (enrollments between seventeen and twenty-five students) are the norm. Moreover, the Tennessee and Wisconsin experiments suffered from restricted class-size ranges: no classrooms in either experiment enrolled more than twenty-six students. The ECLS-K data again indicate that a considerable number of U.S. public school students in kindergarten and first grade are enrolled in classrooms larger than this. As such, the parameters of the "large" classrooms in this study are somewhat larger than those in either the Tennessee or the Wisconsin experiments.

SCHOOL SIZE. For three reasons, we chose not to employ a continuous measure of school size in our statistical models. First, the variable measuring elementary school size is positively skewed, with many more small than large schools. This non-normal distribution precludes its use as a continuous measure in our multivariate analyses, which assume normal distributions. Second, our previous research on school size suggests nonlinear relationships between school size and student learning.[34] Third, in addressing issues of interest to policymakers and school administrators, it is helpful to offer results that have clear substantive meaning. Although we could have transformed our school size measure using the natural logarithm, describing results in terms of "log size" can be a cumbersome venture. Thus we have constructed a series of dummy variables that identify small schools (fewer than or equal to 275 children), medium-small schools (276–400), medium-size schools (401–600), medium-large schools (601–800), and large schools (more than 800 students). In our multivariate analyses, medium-size schools are the uncoded comparison group.

DEPENDENT MEASURES: COGNITIVE ASSESSMENTS. Each ECLS-K cognitive assessment was administered individually, with an adult assessor spending fifty to seventy minutes with each child at each testing wave.[35] The literacy assessment at each wave was designed to measure both basic literacy skills (print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, word recognition) as well as more advanced reading comprehension skills (initial understanding, interpretation, personal reflection, and ability to demonstrate a critical stance). These advanced literacy skills, which were assessed through verbal dialogue between the child and the assessor, measured children's ability to identify the main points of a passage and connect text to their own personal experiences and assessed their critical thinking skills and ability to distinguish real versus imaginary content. Mathematics assessment items were designed to measure conceptual and procedural knowledge and problem solving, with items equally divided between number sense and measurement. The scores on both the reading and mathematics assessments at each wave were equated separately using item response theory, in order to make them appropriate measures of change over time. Our analyses use the IRT-scale scores.[36]

SOCIAL AND ACADEMIC BACKGROUND. Children's socioeconomic status is captured with a composite measure of parents' income, education, and occupational prestige (a z score; M = 0, SD = 1). Our analyses also employ a dummy-coded gender measure (girls = 1, boys = 0) and a measure indicating whether the child was a member of a traditionally underperforming racial or ethnic group (Hispanic, African American, Native American, and multiracial children = 1, white and Asian children = 0). The models further account for children's age (in months), whether the child lived in a single-parent home (yes

= 1, no = 0), and whether a language other than English was the primary home language (yes = 1, no = 0). Academic background is captured by whether the child was repeating kindergarten (yes = 1, no = 0) and attended full-day kindergarten (yes = 1, no = 0).

SCHOOL CHARACTERISTICS. The focus of this research is our level-3 (between-school) HLM models. In addition to the average class-size and school-size measures discussed above, to capture grade span we use a set of dummy-coded indicators to identify primary (K–3) schools, K–8 schools, and K–12 schools, which are each compared to elementary (K–6) schools in our multivariate HLMs. Our school-level models also incorporate composition controls for school-average SES (a z score) and high-minority enrollment (a dummy variable indicating non-white and non-Asian enrollments above 33 percent). Due to documented associations between urbanicity and school size, we include dummy-coded indicators of school location (large city, medium city, rural–small town, each compared to suburbs–urban fringe).

Results

We present both descriptive and multivariate results. Our descriptive results provide information about both children and schools, organized by the size of their classes and schools. We tested group mean differences for statistical significance with t tests (for continuous variables) and cross-tabulations (for categorical variables). We present our within-school and between-school multivariate and multilevel HLM results separately. Our within-school results describe the relationships between child-level characteristics and student learning. Our between-school models explore the effects of elementary school organizational size on student learning—the focus of this paper. All HLM results in tables are presented in the test score points per month (ppm) metric described earlier, although we also convert some to effect sizes (ES) and annual test-score point differences.

Descriptive Results

Table 1 presents information about schools and students organized by school size. A linear relationship is evident between school size and average kindergarten and first-grade class size, although the differences are small and mostly not statistically significant. We find stronger evidence of a (curvilinear) relationship between average-SES and school size. A 0.4 standard deviation

Table 1. Descriptive Statistics for Schools and Students, by School Size[a]

Indicator	*Small school*	*Medium-small school*	*Medium-size school*	*Medium-large school*	*Large school*
Schools (N = 527)					
Sample size	110	128	171	80	38
Average kindergarten class size	19.3	20.2	20.7	21.5	22.2
	(5.9)	(4.6)	(4.8)	(3.9)	(3.7)
Average first-grade class size	18.2***	19.3	20.4	21.4	21.2
	(4.8)	(3.0)	(3.2)	(3.1)	(4.2)
Average socioeconomic status[b]	–0.3*	–0.1	0.1	0.1	–0.2*
	(0.7)	(0.9)	(1.1)	(1.0)	(1.0)
Percent high-minority school[c]	18.2**	24.2	33.1	33.8	44.7
Percent primary school (K–3)	18.2***	6.3	4.6	5.0	2.6
Percent elementary school (K–6)	64.5***	82.7	83.9	82.5	86.8
Percent K–8 school	11.8	7.9	6.3	10.0	10.5
Percent K–12 school	5.5	3.1	5.2	2.5	0.0
Percent large city	9.1	7.0	10.9	21.3*	18.4
Percent medium-size city	20.9	22.6	24.6	17.5	13.2
Percent suburban or urban fringe	20.0***	32.8	38.9	48.8	50.0
Percent small town or rural	50.0***	32.6	25.3	12.5*	18.4
Students (N = 7,740)					
Sample size	1,004	1,510	2,733	1,615	878
Socioeconomic status[b]	–0.2***	–0.1***	0.0	0.1	–0.1*
	(0.9)	(1.0)	(1.0)	(1.0)	(0.9)
Age (months)	66.1	66.1	66.1	66.4	66.2
	(4.2)	(4.2)	(4.2)	(4.2)	(4.3)
Percent female	51.3	46.3	49.2	47.7	48.8
Percent full-day kindergarten	49.2	48.9	51.0	59.2***	71.9***
Percent minority (non-white or non-Asian)	24.0***	28.5	31.3	33.4	34.5
Percent English as a second language	3.6***	3.2***	6.7	7.6	8.0
Percent single-parent family	23.3	23.6	22.5	23.8	22.4
Percent repeating kindergarten	4.8*	2.7	3.1	4.9**	2.7

Source: Authors' calculations using ECLS-K data.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

a. Unweighted $N = 7,740$ children nested within 527 public schools. Small schools are up to 275 students; medium-small schools are 276–400 students; medium-size schools are 401–600 students; medium-large schools are 601–800 students; large schools are more than 800 students. All significance tests are compared to medium-size schools; standard deviations are in parentheses.

b. Measure is z scored (M = 0, SD = 1).

c. School enrollment is at least 33 percent non-white, non-Asian.

average SES gap separates small and medium-size schools ($p < 0.05$), and a 0.3 standard deviation gap separates large and medium-size schools. As subsequent results suggest, small schools tend to be rural (and lower SES) and large schools tend to be urban (and also lower SES). In short, public schools at both ends of the size continuum tend to serve socioeconomically disadvantaged students. Indeed, half of the small schools in our sample are located in small towns and rural areas, compared to only one-quarter of medium-size schools

($p < 0.001$). Further reflecting the small-town and rural character of these schools, less than one-fifth (18.2 percent) of small schools have high-minority enrollments, compared to almost one-third of medium-size schools ($p < 0.01$). In terms of grade span, compared to medium-size schools, small schools are less likely to be primary schools and more likely to be elementary schools ($p < 0.001$).

Mirroring these school-level descriptive statistics, the results in table 1 indicate that children in small, medium-small, and large schools tend to come from less-advantaged families than children in medium-size schools. Minority children are less likely to attend small compared to medium-size schools ($p < 0.001$). Children in medium-large and (especially) large schools are more likely to attend full-day kindergarten ($p < 0.001$). This may reflect the fact that many urban public schools (which tend to be larger) offer full-day kindergarten as a compensatory program. Children in small and medium-large schools are more likely to be kindergarten repeaters than those in medium-size schools. Children's age, gender, and single-parent status are not related to the size of the school they attend.

Table 2 presents information about schools and students organized by average class size in kindergarten and first grade. Although the sociodemographic relationships are similar to those found in table 1, there are some clear differences. Most notably, the relationship between school size and average class size becomes even more evident. Schools with small kindergarten and first-grade classes enroll roughly 130 fewer students than schools with medium-size kindergarten and first-grade classes ($p < 0.001$). Moreover, schools with large first-grade classes enroll almost 100 students more than those with medium-size first-grade classes ($p < 0.05$). Schools with large kindergarten and first-grade classrooms are also considerably more likely to enroll high proportions of minority students ($p < 0.05$) and to be located in large cities ($p < 0.01$). Conversely, schools with small classes are quite likely to be located in small towns and rural areas ($p < 0.01$) and in suburban and urban fringe communities ($p < 0.01$). Indeed, more than two-thirds of schools with small classes are located in these areas (40.2 and 29.0 percent of small kindergarten classes; 42.5 and 23.9 percent of small first-grade classes).

A curvilinear relationship between class size and socioeconomic status is evident. Children attending schools with small *and* large class sizes are less socially advantaged compared to those attending schools with medium-size classes ($p < 0.001$). Almost half of the children attending schools with large kindergarten (43.4 percent) and first-grade classrooms (45.9 percent) are members of racial or ethnic minority groups ($p < 0.001$). As with SES, schools with

Table 2. Descriptive Statistics for Schools and Students, by Class Size[a]

	Kindergarten			First grade		
Indicator	*Small classes*[a]	*Medium-size classes*	*Large classes*	*Small classes*	*Medium-size classes*	*Large classes*
Schools (N = 527)						
Sample size	107	341	80	113	374	40
Average enrollment	359.0***	484.2	473.2	344.9***	478.9	572.6*
	(156.7)	(212.8)	(265.9)	(188.4)	(206.2)	(272.9)
Average socioeconomic status[b]	–0.2*	0.1	–0.3**	–0.4***	0.1	-0.2*
	(0.8)	(1.0)	(0.9)	(0.8)	(1.0)	(1.0)
Percent high-minority school[c]	18.7*	29.1	41.3*	23.9	27.9	51.3*
Percent primary school (K–3)	4.7	9.0	7.5	6.2	9.0	2.6
Percent elementary school (K–6)	72.0**	83.4	70.0**	77.9	79.6	82.1
Percent K–8 school	15.0***	5.2	16.5***	9.7	8.0	15.4
Percent K–12 school	7.5*	2.3	6.3*	7.1*	3.4	0.0
Percent large city	4.9*	11.4	24.1**	5.3*	12.2	28.2**
Percent medium-size city	27.1	24.4	8.9**	28.3*	20.7	22.5
Percent suburban or urban fringe	29.0**	38.4	33.8	23.9**	39.4	35.0
Percent small town or rural	40.2**	25.9	32.9	42.5**	27.6	12.8*
Students (N = 7,740)						
Sample size	1,295	5,352	1,093	1,289	5,749	702
Socioeconomic status[b]	–0.1***	0.0	–0.2***	–0.2***	0.0	–0.1***
	(1.0)	(1.0)	(1.1)	(0.9)	(1.0)	(1.1)
Age (months)	66.3	66.2	66.1	66.6**	66.1	66.0
	(4.4)	(4.2)	(4.2)	(4.4)	(4.2)	(4.2)
Percent female	46.4	49.1	47.6	47.2	48.6	49.8
Percent full-day kindergarten	38.3***	56.8	62.9***	57.2	53.4	63.9***
Percent minority (non-white or non-Asian)	27.0	29.2	43.4***	30.6	29.2	45.9***
Percent English as a second language	3.2***	6.0	9.2***	4.0**	5.9	10.8***
Percent single-parent family	27.3***	21.2	28.8***	24.2*	22.4	27.7**
Percent repeating kindergarten	3.7	3.5	3.9	3.7	3.7	2.3

Source: Authors' calculations using ECLS-K data.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

a. Unweighted $N = 7{,}740$ children nested within 527 public schools. Small classes are up to seventeen students; medium-size classes are between seventeen and twenty-five students; large classes are more than twenty-five students. All significance tests are compared to schools with medium-size classes; standard deviations are in parentheses.

b. Measure is z scored (M = 0, SD = 1).

c. School enrollment is at least 33 percent non-white, non-Asian.

small *and* large classrooms also enroll greater proportions of children from single-parent homes and children for whom English is not the primary home language. Almost two-thirds of children in public schools with large kindergarten and first-grade classrooms receive full-day kindergarten, compared to slightly more than half of students in schools with medium-size classrooms ($p < 0.001$). Despite their relative socioeconomic disadvantage, children attending schools with small kindergarten classrooms are less likely to receive full-day kindergarten ($p < 0.001$); only 38.3 percent of children in schools whose kindergarten classes are small attend full-day kindergarten.

In sum, our descriptive results indicate a modest, but positive, relationship between public school size and class size. Thus the effects of each measure of context size should be estimated net of the other. Smaller schools (with smaller classes) are more likely to be located in rural areas, whereas larger classes (often in larger schools) are more often located in large cities. Although schools with high concentrations of minority students are more likely to be large (and to offer large classes), the relationship between SES and school size follows a different pattern. Both the largest and the smallest schools (with larger and smaller classes) enroll disproportionate numbers of socially disadvantaged children. It is clear from these descriptive differences that our estimates of class-size and school-size effects on young children's learning must include statistical controls for social background, school composition, school location, and grade span.

Within-School Results

Our within-school HLM models explore the associations between child-level characteristics and learning in kindergarten and first grade (see table 3). We speak of "learning" because the outcomes are gains over the kindergarten and first-grade years (figure 2). Although our between-school models represent the primary focus of this study, we briefly describe our child-level results here. Over the kindergarten year, girls gain more skills in literacy (0.13 test score points per month [ppm], $p < 0.001$) and mathematics (0.04 ppm, $p < 0.05$) than their male counterparts. Children attending full-day kindergarten learn considerably more than their peers attending half-day programs (0.26 and 0.15 ppm in literacy and mathematics, respectively; $p < 0.001$). These results are similar to findings from our previous research using ECLS-K.[37]

In previous work, we describe the considerable racial and socioeconomic disparities that characterize young children's achievement as they begin kindergarten.[38] The results in table 3 indicate that these social disparities actually

Table 3. Within-School Models of Kindergarten and First-Grade Literacy and Mathematics Learning[a]

Indicator	*Literacy learning*	*Mathematics learning*
Kindergarten		
Female	0.13***	0.04*
Full-day kindergarten	0.26***	0.15***
Age (months)	0.00	0.00
Socioeconomic status[b]	0.07**	0.04**
English as a second language	0.13*	0.09*
Single-parent family	–0.05	–0.02
Repeating kindergarten	–0.14	–0.16*
Minority (non-white or non-Asian)	–0.13**	–0.12***
Intercept	1.61***	1.29***
First grade		
Female	0.03	–0.04
Full-day kindergarten	–0.27***	–0.11**
Age (months)	–0.01	–0.01***
Socioeconomic status[b]	0.06*	–0.02
English as a second language	0.10	0.01
Single-parent family	–0.06	–0.01
Repeating kindergarten	–0.38	–0.14*
Minority (non-white or non-Asian)	0.01	0.03
Intercept	2.56***	1.56***

Source: Authors' calculations using ECLS-K data.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

a. $N = 7{,}740$ children nested within 527 public schools. All coefficients are in a points-per-month learning metric. All measures are grand-mean centered.

b. Measure is z scored (M = 0, SD = 1).

increase during kindergarten. Even after adjusting for many other child-level covariates, minority status is associated with reduced literacy and mathematics gain during kindergarten (–0.13 ppm in literacy, $p < 0.01$; -0.12 ppm in mathematics, $p < 0.001$). Conversely, higher-SES children tend to gain more skills: a 1 standard deviation increase in SES is associated with 0.07 ppm additional learning in literacy and 0.04 ppm additional learning in mathematics ($p < 0.01$).

Children for whom English is not the primary home language gain more literacy skills during kindergarten than their English-speaking counterparts (0.13 ppm in literacy, $p < 0.05$; 0.09 ppm in mathematics, $p < 0.05$). Unlike these potentially compensatory effects associated with language-minority status and full-day kindergarten, the results in table 3 challenge the efficacy of kindergarten retention practices. Kindergarten repeaters learn less than nonrepeaters in mathematics (–0.16 ppm, $p < 0.05$) and gain literacy skills at comparable rates to nonrepeaters during their second year of kindergarten.

We turn now to within-school results for the first grade. Our results in table 3 suggest that, over the first-grade year, children who did not attend full-day kindergarten "catch up" to their counterparts who did. Indeed, the learning advantages of full-day kindergarten are significantly reversed during first grade. Another departure from the kindergarten results is that minority and nonminority students learn at similar rates during first grade (that is, their learning rates are parallel), whereas they are disadvantaged in kindergarten learning in both subjects. However, higher-SES children continue to gain more skills in literacy during first grade ($p < 0.05$), but not in math. Accounting for the other covariates, first-grade literacy and mathematics learning are not related to gender, single-parent, or language status.

We also note here the different learning rates in each subject in kindergarten and first grade (comparing the intercepts in column 1 and column 2). Particularly in literacy—but also in mathematics—on average children gain considerably more skills in first grade than in kindergarten. The literacy intercepts indicate an adjusted average monthly gain of 1.61 ppm in kindergarten, but 2.56 ppm during first grade. Although the distinction is not as stark with mathematics, children gain 0.27 ppm more in first grade than in kindergarten (1.56 versus 1.29 ppm). These differential learning rates may reflect two phenomena: a generally stronger academic focus of most first-grade classrooms and the fact that virtually all first grades are full day. One implication for this study is obvious: less variability in kindergarten learning suggests less variability that may be explained as a function of elementary school organizational size.

EXPLORING EQUITY. In the HLM level-2 models presented in table 3, all child-level variables modeling learning are grand-mean centered, and their between-school variances are fixed to 0. Our original intention, as described in the second research question, was to explore whether class size or school size was associated with equity. The equity measures we considered are captured by the relationship between SES and learning in either subject at either grade—essentially four slopes-as-outcomes (two subjects, two grades). Although in some cases, these slopes vary significantly between schools, none of the four SES-learning slopes is related to either class size or school size. In other words, the class- and school-size effects we report here are similar across race and social class background. Thus we "fix" the four SES-learning slopes in our level-2 HLM models (similar to the slopes of the other child characteristics in these models). We also explore whether class-size and school-size effects—which we discuss below—are different for schools with different minority concentrations and social class compositions. We do find some interactions, but they are inconsistent. As such, we decided to focus on our main effects, which are themselves quite complicated.

BETWEEN-SCHOOL RESULTS. The major findings from our study of class size and school size are presented in table 4. The estimates obtained from our between-school level-3 HLM models are adjusted for both the child-level characteristics in table 3 as well as the school-level measures described earlier and displayed here. As the child-level estimates presented in table 3 change very little from the level-2 HLM models, we do not present them again here. Rather, we focus on our major findings regarding the relationship between organizational size and student learning. However, it is important to understand that the learning outcomes (the intercepts) of the level-3 HLM models shown in table 4 include the full set of controls shown in table 3.

Compared to children in schools with large kindergarten classes, children in schools with small kindergarten classes gain 0.10 ppm more in literacy and 0.08 ppm more in mathematics ($p < 0.10$). Expressed in terms of the standard deviation of the subject- and grade-specific learning slopes, these represent yearly (nine-month) effect sizes of 0.14 standard deviation in literacy and 0.15 standard deviation in mathematics. More noteworthy, children in schools with medium-size classrooms gain more in literacy (0.14 ppm, ES = 0.19 SD) and mathematics (0.08 ppm, ES = 0.15 SD) than those in large kindergarten classrooms ($p < 0.05$). Because we are interested in identifying "ideal" class sizes, we also estimated effects of small compared to medium-size classrooms in other HLM models not shown here. However, we found no differences in literacy or mathematics learning between schools offering small rather than medium-size kindergarten classrooms. The interpretation of these results suggests *detrimental* effects of large kindergarten class size rather than *beneficial* effects of small classes.

These findings support the conclusions from the Tennessee and Wisconsin class-size experiments. However, we extend their important findings. By including schools with medium-size classrooms in our models—which neither the Tennessee nor the Wisconsin experiments considered—our results suggest that schools may enjoy similar advantages by decreasing enrollment from large to mid-size classrooms. Moving to even smaller classes does not appear to provide additional academic benefits, even though such a change would surely require considerable additional costs.

Table 4 also documents significant class-size effects on literacy and mathematics learning during first grade, but with somewhat different patterns. Compared to those in large first-grade classes, children attending schools with small classes gain 0.19 ppm more in literacy (ES = 0.20 SD; $p < 0.05$) and 0.12

Table 4. Between-School Models of Kindergarten and First-Grade Literacy and Mathematics Learning[a]

Indicator	*Literacy learning*	*Mathematics learning*
Kindergarten		
Small classes[b]	0.10~	0.08~
Medium-size classes	0.14*	0.08*
Small school[c]	–0.04	–0.03
Medium-small school	0.02	0.02
Medium-large school	0.02	0.00
Large school	–0.03	–0.01
Primary school (K–3)[d]	–0.07	–0.06
K–8 school	–0.09	0.03
K–12 school	–0.08	–0.08
Large city[e]	0.05	0.04
Medium-size city	0.05	–0.02
Rural-small town	–0.11~	–0.05
Average socioeconomic status[f]	0.01	–0.04~
High-minority school[g]	–0.07	–0.12*
Random effect (intercept)	1.55***	1.28***
First grade		
Small classes[b]	0.19*	0.12**
Medium-size classes	–0.07	0.09*
Small school[c]	0.03	0.13*
Medium-small school	0.07	0.08
Medium-large school	0.04	0.02
Large school	–0.17*	–0.03
Primary school (K–3)[d]	–0.08	0.01
K–8 school	0.12	0.04
K–12 school	–0.57***	–0.03
Large city[e]	–0.01	0.08
Medium-size city	0.12~	0.18***
Rural or small town	–0.12	–0.05
Average socioeconomic status[f]	0.03	–0.05*
High-minority[g]	–0.18*	–0.13*
Random effect (intercept)	2.49***	1.43***

Source: Authors' calculations using ECLS-K data.

~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

a. $N = 7{,}740$ children nested within 527 public schools. All coefficients are in a child-level points-per-month of learning metric.

b. Small classes (up to seventeen students) and medium-size classes (between seventeen and twenty-five students) are compared to large classes (more than twenty-five students).

c. Small schools (275 students or fewer), medium-small (276–400 students), medium-large (601–800 students), and large schools (more than 800 students) are compared to medium-size schools (401–600).

d. Compared to elementary (K–6) schools.

e. Compared to suburban or urban fringe schools.

f . Measure is z scored (M = 0, standard deviation = 1).

g. School enrollment at least 33 percent non-white, non-Asian.

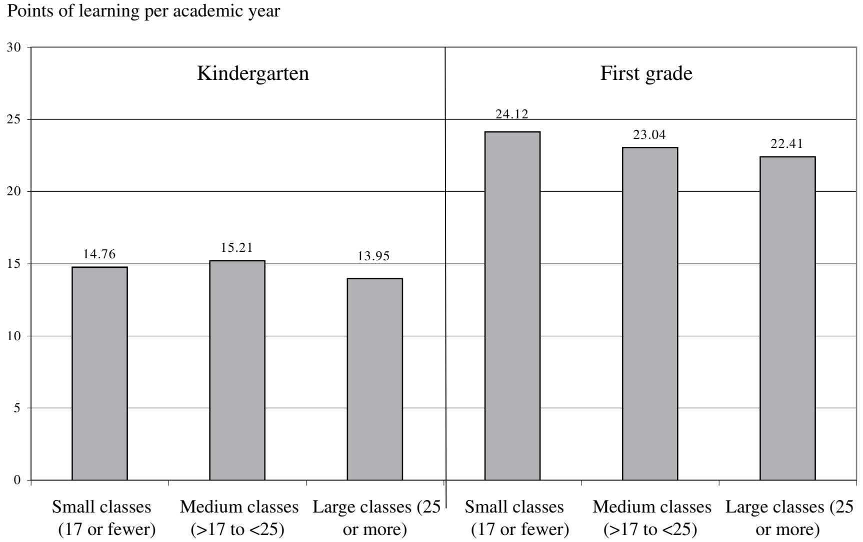
ppm more in mathematics (ES = 0.24 SD; $p < 0.01$). Children in schools with medium-size classrooms also learn more mathematics than their peers in schools offering large classrooms (0.09 ppm, ES = 0.18 SD; $p < 0.05$). As we do with kindergarten learning, we compare the learning rates associated with small and medium-size first-grade classrooms. Whereas we find no benefits of small compared to medium-size kindergarten classrooms, children in schools with small first-grade classrooms gain more literacy skills than those in schools with medium-size first-grade classrooms (0.12 ppm, ES = 0.13 SD; $p < 0.05$). In mathematics, however, we find no differences between schools offering small and medium-size classrooms.

Figures 3 and 4 offer simple illustrations of these class-size effects. Compared to both small and medium-size classrooms, the results shown in figure 3 indicate that children in schools with large kindergarten classes gain fewer literacy skills over the course of the kindergarten year. In first grade, however, the effects are somewhat different: schools with small classes have an advantage over schools with both medium and large classes. The results for mathematics learning presented in figure 4 are more dramatic. Children learn less in schools with large compared to both small and medium-size classrooms ($p < 0.05$), whereas the learning rates for children in schools with small and medium-size classes are similar. As explained earlier, schools with small classes have no advantage in terms of mathematics learning over those offering medium-size classes. However, clear negative effects are associated with schools offering large kindergarten and first-grade classes.

Once we take into account the types of students they enroll and other social and structural characteristics—notably average class size—we find less dramatic evidence of school-size effects on student learning in kindergarten, as shown in table 4. In first grade, however, size effects are more important. Students learn more mathematics per month in small compared to medium-size schools (0.13 ppm, ES = 0.24 SD; $p < 0.05$) and learn fewer literacy skills per month in large compared to medium-size schools (–0.17 ppm, ES = –0.18 SD; $p < 0.05$). The less dramatic findings regarding elementary school size (compared to class size) may be due to the self-contained nature of most kindergarten and first-grade classrooms, so that the classroom context may be more relevant to learning than the larger school context. As noted, unlike high school students, children's experiences in elementary school are generally influenced more by their classroom context, in which the vast majority of their experiences occur.

The findings about school-size effects on learning in literacy and mathematics in first grade may be clearer when displayed graphically. As with figures

Figure 3. Annual Literacy Learning, by Class Size and Grade

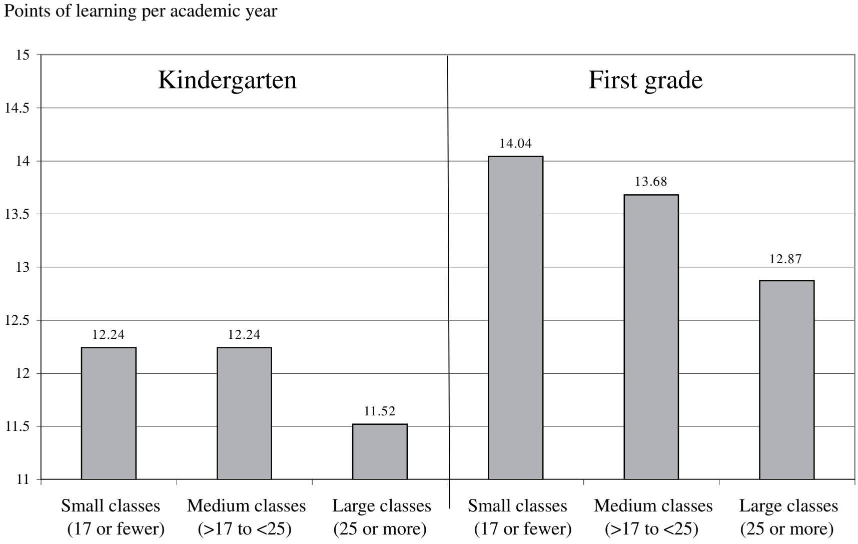


Source: Authors' calculations using ECLS-K data.

3 and 4, which document class-size effects over a full school year, figure 5 also presents annualized results.[39] Three findings are evident. First, learning in mathematics is advantaged in small schools (annual learning of 14.04 points a year versus 12.6 points a year in large schools). Second, learning in literacy is considerably disadvantaged in large schools (19.98 points a year versus 23.04 points a year in medium-small schools). Third, consistent with our earlier studies of school size in upper grades, school-size effects on learning in the lower-elementary grades are distinctly nonlinear. Given the substantial correlation between school size and average class size in this nationally representative sample of U.S. public elementary schools, as well as the association of school size with school grade span and location, we believe these findings are quite important.

Although they are not the focus of this chapter, table 4 indicates other school-level effects on student learning. In kindergarten, children in small-town and rural schools gain fewer literacy skills than their suburban counterparts (–0.11 ppm, $p < 0.10$). Another notable effect is the reduced first-grade literacy learning among children in K–12 schools compared to children in public elementary schools. Compared to those attending traditional K–6 schools, students in K–12 schools gain about 0.5 point less per month (0.57 ppm, $p < 0.001$). Moreover, children attending schools in medium-size cities

Figure 4. Annual Mathematics Learning, by Class Size and Grade



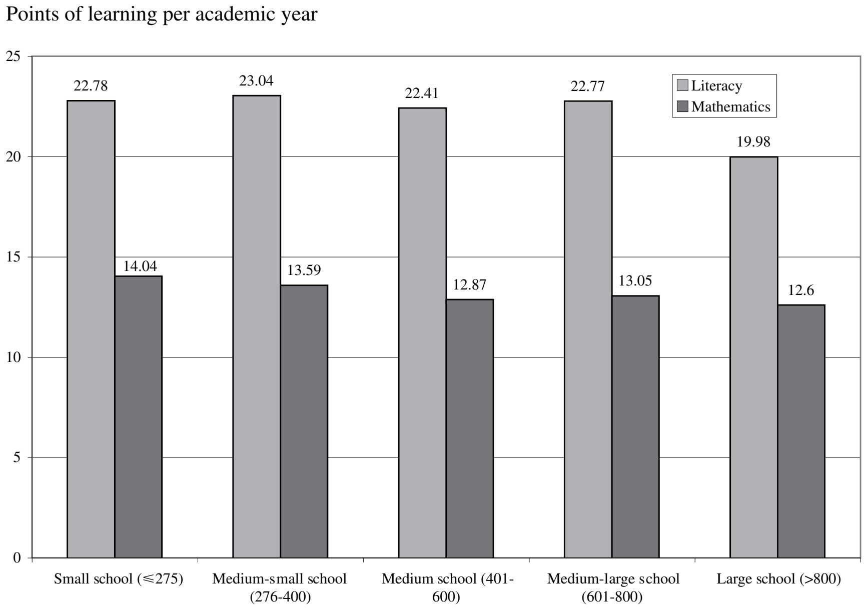
Source: Authors' calculations using ECLS-K data.

learn more in literacy and mathematics than their counterparts attending schools in suburban and urban fringes ($p < 0.001$). Even after accounting for the other child- and school-level covariates, children attending high-minority-enrollment schools gain fewer skills in kindergarten (–0.12 ppm in mathematics; $p < 0.05$) and in first grade (–0.18 ppm in literacy; –0.13 ppm in math; $p < 0.05$). In kindergarten and first grade, once we account for the other child- and school-level covariates, we find a small negative relationship between school-average SES and mathematics learning.

Revisiting Our Research Questions

Early in this chapter we pose three research questions around which we designed this study. Before expanding on the larger implications of the study, we summarize our results in terms of our guiding questions. Our first research question asks how class size and school size influence young children's learning trajectories in their first two years of formal schooling. Our results here are quite straightforward. Even in models that estimate class size and school size simultaneously for separate learning trajectories and also include substantial controls for other child and school characteristics, we find effects for both measures of context size. Kindergarteners attending schools with medium-size

Figure 5. Annual First-Grade Learning Rates, by School Size and Subject



Source: Authors' calculations using ECLS-K data.

classes learn more in both literacy and mathematics than their peers in schools with large classes. In first grade, children's learning rates are greater in schools with small compared to large classes. School-size effects, although more modest, are evident for children's learning in both subjects, but only in first grade. Mathematics learning is higher in small schools, and literacy learning is lower in the largest schools. Thus the findings for our first research question are consistent; in general, young children learn more in smaller contexts.

Our second research question focuses on the equitable distribution of learning, in particular by children's social class. Here, our findings are less satisfying but also quite consistent. We identify no relationship between either class size or school size on the equitable distribution of learning in either grade or subject.

Our third research question asks if the effects of the size of educational contexts are different for the same children's learning rates in kindergarten and first grade. In terms of findings about school size, the results do vary by grade. Although school size does not influence children's learning over the kindergarten year, we do find school-size effects on first graders' learning, generally favoring smaller schools. The class-size effects on learning also vary by grade. Kindergartners learn more in both subjects in schools with medium compared to larger classes, whereas first graders learn more in schools with small classes.

Discussion

Our results suggest robust class-size effects, net of school size, the types of students enrolled, and other school-level characteristics; the effects of both class size and school size are estimated in the same models. That is, the class-size effects we report here are independent of school size, and vice versa. To us, this says that these size effects are both real and important. In literacy and mathematics learning in both kindergarten and first grade, our study provides clear support for the findings from the Tennessee and Wisconsin class-size experiments: children learn more in small compared to large classes. However, our study adds several additional dimensions. First, we compare schools with small and large classes to those with medium-size classes—the type of classroom that elementary school students are most likely to experience. With kindergarten literacy and mathematics as well as first-grade mathematics, schools with small and medium classes do not differentially influence student learning (figures 3 and 4). Rather, schools with *large* classes are *detrimental* to student learning. Only in first-grade literacy learning do we find small class sizes to be more beneficial than medium-size classes.

Second, although our study is not an experimental one, our findings are net of a large set of statistical controls that are systematically linked with the size of school contexts: students' social background, whether they experienced full- or half-day kindergarten, school social composition, and several structural and organizational properties of schools, such as grade span and urbanicity. Third and most important is the structure of our analytic models, where class size and school-size effects are estimated simultaneously. The fourth advantage of this study over others that have considered this topic—perhaps the most important contribution—is the structure of our analysis of learning gains, where children's achievement is measured one-on-one in untimed tests of literacy and mathematics. Moreover, children's learning trajectories are estimated in a complex piecewise linear growth model that accounts for when children are in and out of school.

Are These Findings Large or Small?

As table 4 indicates, first graders in small classes learn almost 10 percent more per month in literacy than children in large classrooms (2.68 versus 2.49 ppm, ES = 0.20 SD). Translating this 0.19 monthly advantage into nine months of learning—the traditional school year—suggests that children in large compared to small classes finish first grade roughly three weeks behind (0.19 x 9 = 1.71, with an average monthly gain of 2.56), an approach we adopt in figures 3–5. Moreover, if children remained in the same elementary school for five or six years and if the class-size and school-size effects were constant over time, these differences would be very substantial: a roughly 10-point advantage for children in small over large classes by the end of sixth grade, or 4.5 months of additional learning. Our findings (particularly when presented in terms of effect sizes) are quite similar to those reported in the Tennessee and Wisconsin experiments, despite the fact that our student samples and methodological approaches differ considerably.

Our results also suggest that first-grade literacy gains are smaller in large elementary schools (those that enroll more than 800 children), and math gains are greater in small schools in the same grade. If these results were sustained over the elementary school years, they would be very large. Moreover, as indicated in table 1, large elementary schools are more likely to have large first-grade classes. This suggests that some children suffer the double disadvantage of attending large schools that offer large classes. Our estimates suggest that such children complete first grade almost 1.5 months behind children enrolled in small first-grade classrooms and schools with enrollments below 800. Again, if the double disadvantage were sustained, children's learning would be very adversely affected.

What causal mechanisms might explain the associations between class size and student learning? Teachers in smaller classes may know their students better and thus more easily tailor instruction to students' needs. Another explanation argues that, rather than instructional or pedagogical improvements, class-size effects may operate through improved classroom climate. Smaller classrooms may foster a more positive disciplinary environment, with fewer student disruptions. As a result, teachers in smaller classes may need to spend less time on classroom management, leaving more time for instruction. In this view, class size benefits may accrue from *student* rather than *teacher* changes in behavior. Future work on this topic might identify practices and processes that typify smaller classrooms. This is crucial, as class size per se may not be the issue, but rather the pedagogical approaches and classroom climates that accompany smaller classrooms.

How might large or small schools influence student learning in elementary schools? Beyond the mechanisms through which class size may influence learning, teachers in smaller schools—both across grades and within the same grade—may have more opportunity to collaborate, to discuss their instructional practices with their colleagues, and to share tips on how to best accommodate challenging children. The less complex organizational form of small schools facilitates collaboration and makes practice more transparent. Moreover, in larger elementary schools weak teachers and struggling students can "slip through the cracks." Similar to class size, we argue that school size does not have a direct effect on student learning, despite the fact that this is exactly the approach our study has, by necessity, taken.

Two Types of Small Schools

In this study, we find that organizational size—of both classes and schools—influences children's learning in literacy and mathematics in both kindergarten and first grade. However, once we account for the characteristics of students and their schools, class size plays a more consistent role than school size in young children's cognitive development. This finding raises questions that are rarely discussed by those who advocate smaller educational contexts. Why do some small schools work better than others? Some schools—both public and private—have small enrollments because they wish to (and are able to) consciously limit the number of students they serve (and frequently also the type of students they enroll). However, the vast majority of small schools are public, and many are in rural areas that must enroll all students in their catchment area. Even with the powerful trend toward consolidation, many schools have small enrollments because there are simply few students in the community (especially in rural areas and communities with declining populations). It seems quite inappropriate to confuse these two types of small schools. Some are "small by design"; others are "small by default." The first group of schools inherently possesses many advantages not shared by the latter group. Interesting as they are, small schools such as Central Park East Elementary School are incredibly different from the majority of rural and small-town small schools, even though some enroll economically disadvantaged students.[40]

Our own and others' research has led us to wonder whether "smallness" by itself is an inherently valuable characteristic, as many advocates claim. Smallness accompanied by the ability to organize a school around a special theme or ideology, to enroll only students, families, and faculty to whom this theme appeals, and to select among applicants is a special kind of smallness. This is very different from smallness experienced by the large majority of "small by default" schools. Indeed, many small elementary schools would prefer to be larger, partly because resources flow to most public schools based on student enrollment. In the context of this study, it is impossible to establish whether small classes and small schools are the product of conscious efforts to limit the size of educational contexts or simply the result of low enrollments. However, in reality such distinctions are crucial when making policy about enrollment or class size.

An Alternative to Small

Policymakers and school practitioners regularly make decisions about the size of elementary schools and classrooms, what grades to include in their school, and the total number of students in each grade. Although school professionals are often required to make such decisions based on local funding, available personnel, or demographic and enrollment projections, ideally they would also base such important decisions on high-quality empirical evidence. Although this study does not meet the current call for randomized studies that would allow very strong causal inferences, we suggest that the empirical results we have drawn from these multiwave longitudinal data and sophisticated statistical models provide a very strong base from which to extract direct policy implications.

In light of our findings, the policy-relevant question may not be whether *small* contexts are more beneficial for student learning than *large* contexts, but whether *medium*-size environments are preferable to large environments, at least in relation to class size. Earlier in this chapter we describe several problems and unforeseen consequences that arose from policies that sought to reduce class sizes in California, even though such decisions were based on very solid empirical evidence from the experimental Tennessee class-size study. With these unintended consequences of California's policy in mind—as well as ever-present concerns about funding—"small" may be unattainable or even undesirable. For example, in districts and schools where large classrooms are a reality, fiscal questions might lead decisionmakers to wonder whether moving from large to even medium-size classrooms would produce equally favorable (and less costly) results. In general, our results suggest that such a move would offer comparable learning benefits. However, our findings about school size are restricted to the extremes, rather than the middle of the distribution. First graders' literacy learning is lower in the largest schools; first graders' mathematics learning is higher in the smallest schools.

Our purpose in this study has been to provide evidence about the potentially confounding elements of elementary school size based on solid data and appropriate methodology. We hope that people who work in schools—and those who make decisions affecting them—would seriously consider how the various elements of size work together, rather than simply accept the increasingly common ideology that "small is good." The findings reported in this paper lead us away from an unquestioning allegiance to small size. Rather than the constant mantra of "small is good," our results lead us to a different proclamation: "large is bad."

Notes

1. Jeremy D. Finn and Charles M. Achilles, "Tennessee's Class Size Study: Findings, Implications, Misconceptions," *Educational Evaluation and Policy Analysis* 21, no. 2 (1999): 97–109; Alan B. Krueger, "Economic Considerations and Class Size," Working Paper 447 (Princeton University, Industrial Relations Section, 2000); Alan B. Krueger and Diane M. Whitmore, "Would Smaller Classes Help Close the Black-White Achievement Gap?" in *Bridging the Achievement Gap*, edited by John E. Chubb and Tom Loveless (Brookings, 2002), pp. 11–46; Barbara Nye and others, "Do Low-Achieving Students Benefit More from Small Classes? Evidence from the Tennessee Class Size Experiment," *Educational Evaluation and Policy Analysis* 24, no. 3 (2002): 201–18.

2. Eric A. Hanushek, "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects," *Educational Evaluation and Policy Analysis* 21, no. 2 (1999): 143–63; Eric A. Hanushek, "The Evidence on Class Size," Occasional Paper 98-1 (Rochester, N.Y.: W. Allen Wallis Institute of Political Economy, 1998); Caroline M. Hoxby, "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics* 115, no. 4 (2000): 1239–85.

3. Hanushek, "Some Findings from an Independent Investigation of the Tennessee STAR Experiment."

4. Ibid. See also Hoxby, "The Effects of Class Size on Student Achievement."

5. See William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (New York: Houghton Mifflin, 2002).

6. Alex Molnar and others, "Evaluating the SAGE Program: A Pilot Program in Targeted Pupil-Teacher Reduction in Wisconsin," *Educational Evaluation and Policy Analysis* 21, no. 2 (1999): 165–77; Alex Molnar and others, "Wisconsin's SAGE Program and Achievement through Small Classes," in *Bridging the Achievement Gap*, edited by Chubb and Loveless, pp. 91–108.

7. Molnar and others, "Evaluating the SAGE Program."

8. CSR Research Consortium, *What We Have Learned about Class Size Reduction in California* (Sacramento: California Department of Education, 2002).

9. Hoxby, "The Effects of Class Size on Student Achievement"; Krueger and Whitmore, "Would Smaller Classes Help Close the Black-White Achievement Gap?"

10. CSR Research Consortium, *What We Have Learned about Class Size Reduction in California*; Christopher Jepsen and Steven Rivkin, *Class Size Reduction, Teacher Quality, and Academic Achievement in California Public Elementary Schools* (San Francisco: Public Policy Institute of California, 2002).

11. CSR Research Consortium, *What We Have Learned about Class Size Reduction in California*.

12. U.S. Department of Education, *A Descriptive Evaluation of the Federal Class-Size Reduction Program: Final Report* (Washington, 2004).

13. Valerie E. Lee and Julia B. Smith, "High School Size: Which Works Best and for Whom?" *Educational Evaluation and Policy Analysis* 19, no. 3 (1997): 205–27.

14. As we found in ibid.

15. Ibid.

16. Valerie E. Lee and Susanna Loeb, "School Size in Chicago Elementary Schools: Effects on Teachers' Attitudes and Students' Achievement," *American Educational Research Journal* 37, no. 1 (2000): 3–31.

17. Carnegie Council of America, *Turning Points: Preparing American Youth for the 21st Century* (New York: Carnegie Corporation, 1989); Joyce L. Epstein, "What Matters in the Middle Grades: Grade Span or Practices?" *Phi Delta Kappan* 71, no. 6 (1990): 438–44.

18. Epstein, "What Matters in the Middle Grades"; David Marshank, "From Teachers' Perspectives: The Social and Psychological Benefits of Multiage Elementary Classrooms," paper presented at the conference Emerging Images of Learning: World Perspectives for the New Millennium, Chicago, March 1994.

19. Epstein, "What Matters in the Middle Grades."

20. David T. Burkam, Deborah L. Michaels, and Valerie E. Lee, "School Grade Span and Academic Achievement: How Important Is the Presence of a First Grade to a Kindergartner's School Performance?" *Elementary School Journal* 107, no. 3 (2007): 287–303.

21. See, for example, Valerie E. Lee and Julia B. Smith, "Effects of School Restructuring on the Achievement and Engagement of Middle-Grade School Students," *Sociology of Education* 66, no. 3 (1993): 163–87.

22. Robert Barr and Rebecca Dreeben, *How Schools Work* (University of Chicago Press, 1983).

23. Burkam, Michaels, and Lee, "School Grade Span and Academic Achievement"; Valerie E. Lee and others, "Full-Day vs. Half-Day Kindergarten: In Which Program Do Children Learn More?" *American Journal of Education* 11, no. 2 (1996): 163–208; Valerie E. Lee, Douglas D. Ready, and Laura LoGerfo, "School Effects on Learning Rates over the Kindergarten and First-Grade Years," paper presented at the annual conference of the American Educational Research Association, San Francisco, April 2006.

24. National Center for Education Statistics (NCES), *Early Childhood Longitudinal Study: User's Manual for the ECLS-K First Grade Restricted-Use Data Files and Electronic Codebook* (Department of Education, Office of Educational Research and Improvement, 2002).

25. Stephen W. Raudenbush and Anthony S. Bryk, *Hierarchical Linear Models,* 2nd ed. (Thousand Oaks, Calif.: Sage, 2002); Judith D. Singer and John B. Willett, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* (New York: Oxford University Press, 2003).

26. See Raudenbush and Bryk, *Hierarchical Linear Models*; Michael H. Seltzer, Ken A. Frank, and Anthony S. Bryk, "The Metric Matters: The Sensitivity of Conclusions about Growth in Student Achievement to Choice of Metric," *Educational Evaluation and Policy Analysis* 16 (Spring 1994): 41–49; David R. Rogosa and John B. Willett, "Understanding Correlates of Change by Modeling Individual Differences in Growth," *Psychometrika* 50, no. 2 (1985): 203–08; John B. Willett, "Questions and Answers in the Measurement of Change," *Review of Research in Education* 5 (1988): 345–422.

27. John B. Willett, "Questions and Answers in the Measurement of Change," p. 377.

28. Ibid.

29. Rogosa and Willett, “Understanding Correlates of Change by Modeling Individual Differences in Growth.”

30. Willett, “Questions and Answers in the Measurement of Change.”

31. David T. Burkam and others, “Social-Class Differences in Summer Learning between Kindergarten and First Grade: Model Specification and Estimation,” *Sociology of Education* 77, no. 1 (2004): 1–31; Lee, Ready, and LoGerfo, “School Effects on Learning Rates over the Kindergarten and First-Grade Years.”

32. Our initial unconditional HLM analyses included a traditional third-degree polynomial model. This model revealed a nonlinear growth pattern between the start of kindergarten and the end of first grade in both reading and mathematics, with increasing learning in kindergarten, decreasing learning over the summer months, and increasing learning in first grade. For two reasons, however, this paper does not develop the polynomial model further. First, the complexity of such models makes them rather difficult to interpret. Second, traditional growth models assume that the temporal distance between repeated measures is constant across individuals—an assumption violated by the data structure of ECLS-K. Rather, we employ piecewise linear models, which permit us to explore the differential kindergarten and first-grade growth rates.

33. We selected only public schools and students, as we felt that decisions about school and class size, and the costs and benefits flowing from such decisions, would be fundamentally different in public and private schools.

34. See Lee and Loeb, “School Size in Chicago Elementary Schools”; Lee and Smith, “High School Size.”

35. NCES, *Early Childhood Longitudinal Study*.

36. As suggested by Seltzer, Frank, and Bryk, “The Metric Matters.”

37. Lee and others, “Full-Day vs. Half-Day Kindergarten”; Douglas D. Ready and others, “Explaining Girls’ Advantage in Kindergarten Literacy Learning: Do Classroom Behaviors Make a Difference?” *Elementary School Journal* 106, no. 1 (2005): 21–38.

38. See Valerie E. Lee and David T. Burkam, *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School* (Washington: Economic Policy Institute, 2002).

39. The results in figures 3–5 are computed as follows. The intercept is first added to the monthly effect for each size category, and then the sum is multiplied by nine. This is because all results in table 4 are in the points-per-month learning metric.

40. See Deborah Meier, *The Power of Their Ideas: Lessons for America from a Small School in Harlem* (Boston: Beacon Press, 1995).