What Are the Long-Term Effects of Small Classes on the Achievement Gap? Evidence from the Lasting Benefits Study

SPYROS KONSTANTOPOULOS
Michigan State University

VICKI CHUNG
Northwestern University

The findings on the social distribution of the immediate and lasting benefits of small classes have been mixed. We used data from Project STAR and the Lasting Benefits Study to examine the long-term effects of small classes on the achievement gap in mathematics, reading, and science scores (Stanford Achievement Test). The results consistently indicated that all types of students benefit more in later grades from being in small classes in early grades. These positive effects are significant through grade 8. Longer periods in small classes produced higher increases in achievement in later grades for all types of students. For certain grades, in reading and science, low achievers seem to benefit more from being in small classes for longer periods. It appears that the lasting benefits of the cumulative effects of small classes may reduce the achievement gap in reading and science in some of the later grades.

A central issue in educational research and policy is how to allocate school-related resources most effectively in order to increase achievement for all student groups. Many of the policies designed to ensure that successful distribution of school resources will result in high levels of academic achievement for all students have, in fact, a dual objective. In particular, such educational policies aim to increase achievement for all students and simultaneously close the achievement gap between low- and high-achieving students. That is, an important objective is to produce larger gains for low-achieving students or to have all students perform as well as the high-achieving students. The urgency with which these objectives are pursued has been increased with the passage of the No Child Left Behind Act, which provides incentives for making

adequate annual progress (or penalties for the lack of it) toward closing the achievement gap. Ideally the reduction in the gap would mean larger increases in achievement for low achievers who are at risk of school failure. Naturally, numerous school interventions have been designed to meet the needs of low achievers who are at risk for school failure and to close the achievement gap between low and high achievers. For example, several strategies of compensatory education target low-achieving students and aim to increase their achievement.

Although such strategies are important, an ideal intervention would increase the achievement levels for all students, while having additional positive effects on the achievement levels of low achievers. Some researchers have argued that class size reduction may be such an intervention (e.g., Finn and Achilles 1990; Krueger and Whitmore 2001). Studies that used high-quality experimental data from Project STAR (student teacher achievement ratio) have consistently demonstrated the positive effects of small classes on average achievement for all students (e.g., Finn and Achilles 1990; Krueger 1999; Nye et al. 2000b). Specifically, these studies demonstrated that the average student achievement in small classes (15 students on average) was significantly higher than that in regular classes (22 students on average), and these findings suggest that reducing class size is a promising intervention that increases academic achievement, on average, for all students. In addition, it is tempting to imagine class size reduction as an educational intervention that increases academic achievement for all students and simultaneously reduces the achievement gap between low- and high-achieving students by producing larger gains for low achievers. Several studies have examined the differential effects of small classes on achievement for minority and disadvantaged students (Finn and Achilles 1990; Krueger 1999; Krueger and Whitmore 2001; Nye et al. 2000a, 2002), however, and the findings have been mixed. For example, Nye and colleagues found weak or no evidence for differential effects of small classes on minority, disadvantaged, and low-achieving students, while Finn and Achilles found statistically significant differential effects favoring minority students in reading in grade 1.

SPYROS KONSTANTOPOULOS is associate professor of measurement and quantitative methods in the Department of Counseling, Educational Psychology, and Special Education in the College of Education at Michigan State University. His empirical work focuses on class size, school, and teacher effects and the achievement gap. His methodological work focuses on power computations in nested designs. VICKI CHUNG is a PhD candidate in human development and social policy at Northwestern University. Her research interests include immigration, acculturation, and how aspects of acculturation affect the peer relationships, school outcomes, and mental and physiological health of immigrant youth.

The effects of class size on student achievement have been of great interest to educational researchers and policy makers during the last two decades. In fact, reducing class size to boost student achievement is a policy option that has gained considerable attention nationwide. Class size reduction is an appealing school intervention because it is easy to implement and does not necessarily require changes in teaching methods or instructional practices. Currently, many states and school districts have enacted or are considering class size reduction with the objective of improving academic achievement for all students and especially for low-achieving students. For instance, Wisconsin has adopted a class size reduction program in early grades to examine the effects of small classes on students of low socioeconomic status (SES). It is imperative, then, to examine whether class size reduction will indeed achieve the intended objective and increase achievement for low-achieving students more than other students (i.e., reduce the achievement gap).

The critical question of whether class size reduction can reduce the achievement gap and hence affect the academic achievement of low- and high-achieving students differently has not been fully answered thus far. In particular, the previous work has not addressed the question of whether the differential effects of small classes on the achievement of low-achieving students persist or weaken over time. The objective of this article is to answer this question by examining differences in achievement, at various percentiles of the achievement distribution, between students who were in small or regular-size classes in early grades (e.g., in grade 3). Specifically, we examined the effects of being in small classes in grade 3 on achievement in subsequent grades 4–8. We used data from a four-year, large-scale, randomized experiment known as the Tennessee Class Size Experiment, or Project STAR, and the long-term follow-up study known as the Lasting Benefits Study (LBS) that followed participants of Project STAR through grade 8 (see Finn et al. 2001; Nye et al. 1999).

The Effects of Class Size Reduction on the Achievement Gap

Previous work on the effects of class size has focused exclusively on estimating mean differences in student achievement between small and regular-size classes (e.g., Krueger 1999; Nye et al. 2000b). However, focusing on average differences of achievement distributions between these types of classes is only one way to evaluate the effects of class size. A more complete assessment would also examine small-class effects in the upper and lower tails of the achievement distribution. Studies have shown (e.g., Hedges and Nowell 1995) that achievement differences between student groups in the middle of the achievement distribution may be qualitatively different from achievement differences in the upper and lower tails. For example, achievement differences between low

achievers in small and in regular classes may be significantly smaller (or larger) than achievement differences between average- or high-achieving students in these classes.

Class size reduction can affect the achievement gap in three ways. First, if small and regular classes affect the achievement of high- and low-achieving students in similar ways, then one would expect that the achievement distributions in small and regular classes would differ only in central tendency— that is, small classes would have a higher mean than regular classes but a comparable standard deviation. This suggests that the achievement distribution in small classes is simply shifted to the right by about one-fifth of a standard deviation, which is the average achievement benefit reported in previous work using Project STAR data (e.g., Nye et al. 2000b). In this case, small classes would have no effect on the achievement gap (between low and high achievers) because all students would benefit equally from being in small classes. Second, if low achievers benefit more from being in small classes, then one would expect that the achievement distribution in small classes would have a higher mean and a smaller standard deviation than the achievement distribution in regular classes. In this case, small classes would close the achievement gap because achievement differences between low and high achievers would be smaller. Third, if high achievers benefit more than low achievers from being in small classes, then one would expect that the achievement distribution in small classes would have a higher mean and a larger standard deviation than the achievement distribution in regular classes. In this case, small classes would not close the achievement gap between low and high achievers.

There are a variety of ways to investigate these hypotheses. First, one could examine the differential effects of small classes on the achievement of low-achieving students. A recent study investigated this issue and found virtually no evidence that small classes help these students more than other students (e.g., Nye et al. 2002). Specifically, Nye and colleagues examined whether low achievers in kindergarten benefited from being in small classes in the first, second, or third grade and found that small classes did not affect low achievers' performance. In addition, a more recent study measured the effects of class size on the within-classroom variation in achievement and found larger variation in achievement in small classes in the first two years of Project STAR (Konstantopoulos 2008). Alternatively, one could examine differences between small and regular classes in the upper and lower tails of the achievement distribution. The idea is that, if the average achievement gain for low-achieving students in small or regular classes is larger than the average achievement gain for medium- or high-achieving students in small or regular classes, then there is evidence that low achievers benefit from small classes more than other students. This would suggest a reduction in the achievement gap. In this case, one would observe larger small-class effects in the low end of the achievement

distribution (low-achieving students) than in the upper end (high-achieving students) or the middle. In addition, if small-class effects were significantly larger (in a statistical sense) in the lower tail, then this would indicate that low achievers indeed benefit more than medium or high achievers from being in small classes. For example, suppose that the small-class effect at the tenth percentile of the achievement distribution is one-third of a standard deviation, while the small-class effect at the ninetieth percentile is one-fifth of a standard deviation. One could argue that this difference is meaningful since it is nearly one-tenth of a standard deviation, which is a considerable effect in education. This finding would indicate that small classes close the achievement gap. In addition, one could construct a *t*-test to examine whether the difference between these two estimates is statistically significant (while taking into account the dependency in the estimates).

The Current Study

The purpose of our study was to examine the effects of being in small classes in early grades on the achievement gap in later grades 4–8. We hypothesized that being in small classes in early grades would have lasting benefits for low achievers in later grades. The mechanism through which low achievers could benefit more than other students from being in small classes has not been explicitly described thus far. A caveat of Project STAR is that information about classroom practices is not available. However, there are some plausible hypotheses about the ways that small classes can affect low achievers. One plausible hypothesis is that teachers in small classes systematically modify their instruction to promote student achievement. That is, it is likely that the majority of teachers in small classes are more likely to identify low achievers and design their instruction to benefit these students. Alternatively, it is also plausible that the instructional practices that take place in small classes benefit low achievers more than other students. That is, in smaller classes, low achievers may have more opportunities to interact with their teachers and may receive more personal attention from their teachers. As a result, such students may engage more frequently in learning, which could result in higher achievement levels. Note that the second hypothesis does not require teachers to modify their classroom practices. Simply put, in small classes students receive more attention from teachers, which may result in larger benefits for low achievers than for other students. In addition, it is likely that small classes boost teacher effectiveness since teachers in small classes are more likely to spend less time in disciplining students and more time in engaging students in learning activities. This is crucial for poorer schools since recent work has shown that teacher effectiveness matters more in schools with higher pro-

portions of disadvantaged (and low-achieving) students (Nye, Konstantopoulos, and Hedges 2004). Regardless, if these hypotheses were true, small classes in early grades would be a method to close the achievement gap between low and high achievers in later grades because low achievers would benefit more than other students in early grades, and this gain is more likely to last for some years.

We used data from a four-year, large-scale, randomized experiment conducted in Tennessee in the mid-1980s (Project STAR). The students who were involved in the experiment were followed for five years after the experiment (grades 4–8). To examine the small-class effects carefully, we carried out several analyses and used quantile regression to estimate class size effects in the tails of the achievement distribution. First, we examined the effects of small classes in grade 3 (the last year of the Project STAR experiment) on the achievement gap in subsequent grades 4–8. Second, we examined the cumulative effects of small classes in grades kindergarten to 3 on the achievement gap in subsequent grades 4–8. Students who participated in Project STAR could have been in small classes for four years or less because some students left (while others joined) the experiment each year or switched treatment groups in some grades. A plausible hypothesis is that students who were in small classes for longer periods may have higher achievement in later grades than do other students. Hence, we determined the effects of being in small classes for different time periods (e.g., zero, one, two, three, or four years) on the achievement gap in grades 4–8. Again we defined the achievement gap as the difference in small-class effects in the upper and lower tails of the achievement distribution. We also conducted analyses that addressed issues related to the internal validity of Project STAR, such as students' switching among different types of classes and more-than-intended variability in actual class size within different types of classes (see Krueger 1999; Nye et al. 1999, 2000b).

In addition, we examined the gender, race, and SES gap across the entire distribution of achievement for grades 4–8. For example, it is likely that the achievement gap is not uniformly distributed across the achievement distribution and that it varies by achievement level (e.g., average, low, or high achievers). Finally, we estimated interaction effects between class size and gender, race, and SES across the distribution of achievement in grades 4–8.

The Tennessee Class Size Experiment

The Tennessee Class Size Experiment, or Project STAR, is discussed in detail elsewhere (e.g., Krueger 1999; Nye et al. 2000b). The experiment involved students in 79 elementary schools in 42 districts in Tennessee. During the first year of the study, within each school, kindergarten students were assigned

randomly to classrooms in one of three treatment conditions: smaller classes (with 13–17 students), larger classes (with 22–26 students), or larger classes with a full-time classroom aide. Teachers were also assigned randomly to classes of different types. Some students entered the study in the first grade or subsequent grades and were assigned randomly to classes at that time. Teachers at each subsequent grade level were also assigned randomly to classes as the experimental cohort passed through the grades. Districts had to agree to participate for four years and allow school visits for verification of class sizes, interviewing, and data collection, including extra student testing. They also had to allow research staff to assign pupils and teachers randomly to class types and to maintain the assignment of students to class types from kindergarten through grade 3. The students who participated in Project STAR for at least one year were followed for five years (grades 4–8) as part of the LBS (see Nye et al. 1994).

Project STAR has high internal validity because, within each school, students and teachers were assigned randomly to classes of different sizes. In addition, because Project STAR is a large-scale randomized experiment that includes a broad range of schools and districts (urban, rural, wealthy, and poor), it has higher external validity than smaller-scale studies. Moreover, the study was part of the everyday operation of the schools that participated, and hence there is a lower likelihood that novelty effects affected the class size estimates.

Previous Findings about Small-Class Effects

Project STAR data have been used to examine the differential effects of class size on the achievement of low-achieving, minority, and disadvantaged students. Although initial findings reported that class size reduction had larger positive effects for minority students (see Finn and Achilles 1990), more recent studies that used modern and more appropriate statistical methods could not fully replicate the early findings. For example, Nye and colleagues (2000a) found weak evidence that class size reduction had larger immediate benefits for minority students. In a subsequent study, Nye and colleagues (2002) examined the differential effects of small classes for low-achieving students and found no evidence of additional benefits for these students. However, a more recent study that used follow-up data from Project STAR indicated that being in small classes for four years may subsequently decrease the race/ethnic achievement gap in reading in grades 4–8 (Nye, Hedges, and Konstantopoulos 2004). Nonetheless, overall there is weak evidence of differential effects of small classes for low-achieving, minority, and disadvantaged students.

Although Project STAR provides the highest quality data about class size

effects that we have had thus far, the study was conducted nearly 20 years ago. During the last decade researchers have examined class size effects analyzing more recent data and have reported mixed findings. For instance, Angrist and Lavy (1999) used nonexperimental data from Israeli elementary schools and found that reducing class size increased fourth and fifth graders' scores significantly. In contrast, Hoxby (2000) analyzed nonexperimental data from elementary schools in Connecticut and reported that smaller classes had little to no effect on student achievement. Other researchers used large data sets such as the Third International Math and Science Study to examine class size effects on math achievement in several countries (see Pong and Pallas 2001; Wossmann and West 2006). For example, Pong and Pallas found beneficial small-class effects on eighth-grade math achievement in the United States but not in other countries. More recently, Milesi and Gamoran (2006) used data from the Early Childhood Longitudinal Study and found no evidence of main class size effects on mathematics and reading achievement or differential class size effects for students from different backgrounds. Overall, the findings from recent work on class size effects have been mixed and inconclusive. Part of this inconsistency is due to the variability in implementation of the class size reduction programs. For example, in Project STAR there was no shortage of well-trained teachers or classrooms, and there were ample classroom facilities and commitment from the school leadership. In contrast, evidence from Florida and California indicates that inadequate funding resulted in a shortage of qualified teachers and typical classroom facilities. Milesi and Gamoran (2006) have provided an excellent discussion about the differences in class size effects between Project STAR and other studies.

Given the mixed previous findings about the differential effects of small classes, one would expect that the small-class effect for low achievers would not be as evident as one would hope. However, it is likely that the cumulative effects of small classes point to larger benefits for low achievers. In the current study we examined how class size reduction affected the achievement gap between low and high achievers in grades 4–8. To determine whether small classes had differential effects on different types of students, we computed differences in achievement at the upper and lower tails of the achievement distribution.

Method

Validity of Project STAR

In Project STAR, as in all longitudinal large-scale studies, fidelity of implementation was compromised somewhat by three factors. First, there was some

switching of students among class types in grades 1–3. Second, there was student attrition between kindergarten and grade 3. Third, there was some overlap in the actual sizes among different types of classes due to larger-than-designed variability in sample sizes within classes. The effects of these threats to the validity of the experiment were investigated by other researchers who concluded that the threats did not affect the outcome of the experiment in mean differences in achievement (see Krueger 1999; Nye et al. 2000b). In addition, another study examined the effects of attrition and switching among classes using follow-up data from the LBS and found that attrition and switching did not affect the small-class mean estimates (Nye et al. 1999).

To ensure the validity of the experiment, it was also crucial that random assignment effectively eliminated preexisting differences between students and teachers assigned to different classrooms. Two recent studies checked whether randomization worked and reported no differences on preexisting characteristics of students or teachers among the assigned conditions (Krueger 1999; Nye, Konstantopoulos, and Hedges 2004). These results are consistent with what one would expect if randomization were successful. In randomized experiments such as Project STAR, participants (e.g., students and teachers) have an equal probability of being assigned to treatment groups (e.g., small classes, regular classes, and regular classes with a classroom aide). This suggests that the students (and teachers) assigned to different class types have similar observed and unobserved characteristics. In turn, this indicates that random assignment is orthogonal to observed and unobserved characteristics. The fact that there is no evidence that randomization was not successful facilitates the causal argument in the current study. That is, when randomization is successful, differences across the achievement distribution should be due entirely to the treatment effect. Hence, the causal argument for differences in average achievement holds also for differences in the tails.

Statistical Analysis

The analysis involved computing differences in achievement between students in small and in regular classes at the upper and lower tails of the achievement distribution. Specifically, we used quantile regression (see Buchinsky 1998; Koenker and Bassett 1978) to estimate the small-class effect at various points on the achievement distribution in grades 4–8. Ordinary least squares regression focuses on average associations between an outcome and a predictor or average differences in an outcome between two groups. However, social scientists are also interested in estimating associations or group differences in outcomes in the tails of the outcome distribution (see Hao and Naiman 2007). Oftentimes, social science researchers focus on inequality and examine race

or gender differences in achievement or wages across the entire distribution to obtain a clearer picture of the race or gender gap. For example, it is likely that the race or gender gap for average achievers is different (i.e., smaller or larger) than the race or gender gap for low or high achievers. Similarly, education researchers frequently examine the effects of school resources or school interventions on low-achieving, minority, and disadvantaged students. It is likely that certain school interventions (such as class size reduction) have differential effects on average, low, and high achievers. Examining the effects of school interventions across the entire achievement distribution provides crucial information about reducing the achievement gap. The typical regression model is inadequate to examine the effects of predictors at different points (called quantiles) of the outcome distribution.

Quantile regression is a natural extension of the typical linear regression because it estimates how changes in the predictors (e.g., school interventions) affect changes in the outcome (e.g., achievement) not only in the middle but in the tails of the outcome distribution as well. Hence, quantile regression provides a more compete understanding of the effects of predictors on the entire outcome distribution (see Hao and Naiman 2007). In addition, quantile regression is a more robust method (compared to typical regression) for analyzing skewed distributions with outliers. Currently, quantile regression is a widely used method in economics and the social sciences, and we argue that this method can also be useful in education research that focuses on educational inequality and the academic prosperity of disadvantaged students, who are typically overrepresented in the lower tail of the achievement distribution. Two recent studies employed quantile regression to examine small-class effects on low and high achievers in early grades (Konstantopoulos 2008) and the persistence of school effects on low- and high-wage earners (Konstantopoulos and Constant 2008). Note that the purpose of the current study is to determine whether small classes produce additional benefits in achievement for low achievers in grades 4–8. Quantile regression is well suited for this purpose because it shows how class size reduction affects the achievement of low, average, and high achievers and allows comparisons among the differential effects measured for the entire achievement distribution. An advantage of quantile regression estimates is that the same index (e.g., standard deviation units) can be computed for small-class effects on achievement across the entire distribution, and hence, the results across different points (quantiles) of the achievement distribution are on the same scale. If small classes produced additional benefits for low achievers, then class size reduction would be an intervention that could reduce the achievement gap.

We ran quantile regressions for mathematics, reading, and science test scores separately for each grade (4–8). In each grade, mathematics, reading, and science scores were regressed first on small-class assignment (taking the value

of one if a student was in a small class in grade 3 and zero otherwise). Gender, race/ethnicity, and lower SES were also included in the equation as covariates. We examined the small-class effect at the lower tail (e.g., tenth and twenty-fifth percentiles), the middle (fiftieth percentile), and the upper tail (e.g., seventy-fifth and top ninetieth percentiles) of the achievement distribution. Then, we tested whether the small-class effects in the upper and lower tails were significantly different from one another. The second part of the analysis determined the cumulative effects of small classes on the achievement gap in grades 4–8. In particular, we created four dummies for zero, one, two, or three years of being in small classes (four years in small classes was the reference group) and modeled these nonlinear effects for mathematics, reading, and science in grades 4–8. Notice, however, that the cumulative effects of small classes were not part of the experimental design, and hence these effects are likely to be related to unobserved variables (e.g., school leadership or parental influences), which indicates a possible endogeneity problem and complicates the interpretation of these estimates as causal. Because our data have a nesting structure, since students are nested within classrooms and classrooms are nested within schools, it is important to take into account this nesting structure when computing the standard errors of the regression coefficients. We used STATA to run quantile regression and computed bootstrapped standard errors for the quantile regression estimates. The bootstrapped standard errors we obtained are robust and take into account the clustering nature of the data as well as heteroscedastic errors.

Results

Small Class in Grade 3

The first series of analyses modeled the treatment effect in grade 3 as a binary indicator. The results of this analysis are summarized in table 1. Note that because the outcome variable was standardized to a mean of zero and a standard deviation of one and the treatment effect is a dummy variable, the regression estimates indicate average differences in achievement between small and regular classes expressed in standard deviation units. The median small-class effects were all positive, two-thirds of them were statistically significant, and the remaining coefficients were significant at the .10 level, with the exception of fifth-grade mathematics. Across achievement tests, the overall small-class effect was slightly smaller than one-tenth of a standard deviation in the middle of the distribution. The estimates in the tails were comparable overall— all coefficients were positive, and more than 50 percent of them were significant at the .05 level. These results indicate that all types of students (low, medium,

TABLE 1

Estimated Third-Grade Small-Class Effects on Mathematics, Reading, and Science at Various Quantiles

| | QUANTILE | | | | | |
GRADE	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	N
Mathematics:						
4	.129*	.086[+]	.086*	.129*	.129*	4,305
5	.166	.095	.071	.095*	.166*	2,505
6	.140*	.059	.070[+]	.070	.117[+]	3,450
7	.023	.046	.116*	.092*	.116[+]	3,896
8	.031	.082*	.072*	.062[+]	.103	4,411
Reading:						
4	.040	.101[+]	.101*	.121*	.152*	4,225
5	.265*	.111*	.089*	.111*	.132[+]	2,507
6	.152	.119*	.065*	.087*	.065	3,454
7	.138*	.138*	.161*	.092*	.138*	3,902
8	.112*	.090*	.067*	.045[+]	.090	4,413
Science:						
4	.113	.113*	.094*	.075*	.094*	4,298
5	.188*	.104[+]	.063[+]	.104*	.167*	2,502
6	.214*	.117*	.058[+]	.078[+]	.078	3,443
7	.180*	.090*	.090*	.045	.158[+]	3,882
8	.133*	.089*	.066[+]	.044	.078[+]	4,402

NOTE.—Class size variable indicates small versus regular size membership at grade 3.
[+] $p < .10$.
* $p < .05$.

and high achievers) in grades 4–8 benefited from having been in small classes in grade 3. While it is remarkable that the effects endure through grade 8 across the achievement distribution, it should be noted that they decline over time (i.e., the class size effects in grade 8 are typically smaller than in previous grades). In certain cases (e.g., reading in grade 5 and science in grade 6), the small-class effects at the tenth percentile were more than twice as large as these at the ninetieth percentile. We should note that these effects were slightly larger than one-fifth of a standard deviation, which is a considerable effect in education. However, these estimates at the tenth percentile were not significantly larger than those at the ninetieth percentile. The estimates at the tenth percentile, however, were significantly different from those in the middle at the .10 level. Although this provides some evidence that in some grades and achievement tests low achievers benefit more from being in small classes than do median achievers, generally the effects of small classes seem to positively affect all types of students.

Intention-to-Treat Analysis in Grade 3

As in any large-scale, longitudinal study, the implementation of Project STAR deviated from the experimental design. One limitation was that in grades 1–3 students who were assigned initially to a specific type of class in one year switched to other types of classes the next year. For example, in the first grade students who were assigned to regular-size and regular-size-with-an-aide classes were randomized again to receive the other treatment condition. Studies have shown that about 50 percent of the students assigned to one type of regular class in kindergarten were reassigned to the other type of regular-size class in the first grade (Krueger 1999; Nye et al. 2000b). With the exception of students' switching between regular classes and regular classes with an aide in first grade, the nonrandom transition rates of students among treatment conditions ranged from 2 to 9 percent across grades (see Nye et al. 2000b). It is noteworthy that the transition rates from regular to small classes were consistently 8–9 percent between grades, whereas transition rates from small to regular classes were much lower (2–4 percent).

Because student transitions among types of classes were nonrandom, it is possible that the estimates of the class size effects are biased. Research that examined mean differences in achievement between small and regular classes, however, showed no evidence of bias (Krueger 1999; Nye et al. 2000b). In the current study we examined whether students' switching among different types of classes affected the small-class effects across the achievement distribution. One way to examine the possible effects of this switching is to estimate effects of the treatment as it was originally assigned the first year a student entered the study. This is equivalent to the intention-to-treat analysis typically used in clinical trials. Suppose, for example, a student is assigned to a regular class in kindergarten and switches to a small class in first grade. In the intention-to-treat analysis, this student is assumed to be part of the regular-size class in the first grade, although he or she actually received a different type of treatment in that grade. The idea is that, if the intention-to-treat analysis produces estimates of the treatment effect that are similar to the estimates obtained from the analysis that defines treatment as it was actually received, switching between classrooms would not compromise the internal validity of the experiment. Hence, for grade 3 we constructed a new dichotomous variable called "original" assignment that took the value of one if a student was originally assigned to a small class in grade 3 (or in any previous grade) and zero otherwise. Then, we reran the analysis discussed earlier for mathematics, reading, and science in grades 4–8.

The results of this analysis are reported in table 2. The structure of table 2 is identical to that of table 1. The results of the intention-to-treat analysis

Estimated Third-Grade Small-Class Effects on Mathematics, Reading, and Science at Various Quantiles: Intention-to-Treat Analyses

| | QUANTILE | | | | | |
GRADE	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	N
Mathematics:						
4	.108*	.086*	.065$^+$	.129*	.086*	4,305
5	.095	.166*	.103*	.119*	.214*	2,505
6	.094	.070	.070	.094*	.199*	3,450
7	.069	.046	.162*	.162*	.116*	3,896
8	.041	.082*	.062*	.082*	.103	4,411
Reading:						
4	.081	.121*	.081*	.101*	.121*	4,225
5	.244*	.089	.066$^+$	.089*	.111	2,507
6	.174*	.109$^+$	.065$^+$	.087*	.065	3,454
7	.138$^+$	.161*	.161*	.115*	.138*	3,902
8	.180*	.157*	.090*	.056	.090	4,413
Science:						
4	.131$^+$	.113*	.075*	.075*	.075*	4,298
5	.063	.021	.042	.083	.104	2,502
6	.117	.078	.058$^+$	.078*	.078	3,443
7	.203*	.135*	.113*	.068	.135$^+$	3,882
8	.144*	.111*	.066*	.044	.066	4,402

NOTE.—Class size variable indicates small versus regular size membership at grade 3 as originally assigned.

$^+$ $p < .10$.

* $p < .05$.

are qualitatively similar to and consistent with those reported in table 1. Again, at the median, all coefficients were positive, and two-thirds of them were statistically significant. These estimates are somewhat smaller than those reported in table 1 overall, and they decline over time; that is, typically the effects are smaller in grade 8. Nonetheless, there is some evidence that the effects persist through grade 8. The estimates in the tails were comparable overall, in that all coefficients were positive, and more than 50 percent of them were significant at the .05 level. As in table 1, these results indicate that all types of students (low, medium, and high achievers) in grades 4–8 benefit from being in small classes in earlier grades. As in table 1, in certain cases such as reading in grades 5 and 6 and science in grade 8, the small-class effects at the tenth percentile were nearly twice as large as those at the ninetieth percentile. Again, these effects were slightly larger than one-fifth of a standard deviation. However, the estimates at the tenth percentile were not significantly larger than those at the ninetieth percentile at the .05 level. The estimate at

the tenth percentile in reading, however, was significantly different from that in the middle at the .10 level. Although this provides some evidence that in some grades and achievement tests low achievers benefit more from being in small classes than do median achievers, generally the effects of small classes seem to have positively affected all types of students. In addition, theses results indicate that switching did not affect the small-class effects much across the entire achievement distribution.

Actual Class Size in Grade 3

Although the experimental design had targeted a certain range of class size for each type of classroom (13–17 for smaller classes and 22–26 for larger classes), variation was more than intended in small and regular classes. For example, in grade 3 the actual class size ranged from 11 to 19 for small classes and from 16 to 29 for regular classes (see Konstantopoulos 2008). That is, there was some overlap between the actual class sizes of the three treatment conditions. This larger-than-intended variability in actual class size for each type of classroom and the modest overlap between small and regular classes may have affected the estimate of the treatment effect. Hence, a more complete analysis would examine the effects of actual class size in the upper and lower tails of the achievement distribution.

To conduct this analysis, one needs to construct a variable for actual class size and include it as the main independent variable in the quantile regression. This approach, however, has the disadvantage that, although target class size is assigned randomly, actual class size is not and may be a result of nonrandom unobserved factors that may also be related to the outcome (i.e., the variable may be endogenous). In this case it is difficult to explain the relationship between actual class size and achievement as a causal effect. A common way to overcome this problem is to use random assignment as an instrumental variable (IV) for actual class size (see, e.g., Angrist et al. 1996; Nye, Konstantopoulos, and Hedges 2004). In the IV regression, actual class size is regressed on random assignment, and the predicted values of this regression are used in the quantile regression as the main independent variable. The advantage of this procedure is that it yields estimates of the causal effects of actual class size.

The results of this analysis are reported in table 3. The structure of table 3 is identical to that of tables 1 and 2, and the results are similar to those reported in tables 1 and 2. Notice that in table 3 all coefficients are negative, as expected, because the predictor is continuous, and one would expect that the smaller the classroom, the larger the effect on student achievement. More than 60 percent of the median estimates were significant at the .05 level.

TABLE 3

Estimated Effects of Third-Grade Actual Class Size on Mathematics, Reading, and Science at Various Quantiles

	QUANTILE					
GRADE	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	N
Mathematics:						
4	−.016*	−.010*	−.011*	−.019*	−.016*	4,305
5	−.020	−.014	−.009	−.013*	−.020*	2,505
6	−.018$^+$	−.007	−.009	−.009$^+$	−.014*	3,450
7	−.003	−.006	−.014*	−.013*	−.015$^+$	3,896
8	−.005	−.010*	−.009	−.007	−.012*	4,411
Reading:						
4	−.005	−.013*	−.012*	−.015*	−.019*	4,225
5	−.032*	−.014$^+$	−.011*	−.013*	−.017	2,507
6	−.019*	−.015*	−.008*	−.013*	−.011	3,454
7	−.017*	−.017*	−.019*	−.011*	−.017*	3,902
8	−.013	−.011$^+$	−.008$^+$	−.005$^+$	−.011	4,413
Science:						
4	−.014	−.014*	−.011*	−.009*	−.011*	4,298
5	−.023*	−.013*	−.008*	−.013*	−.020*	2,502
6	−.027*	−.014*	−.007$^+$	−.010*	−.010$^+$	3,443
7	−.022*	−.011*	−.011*	−.005	−.019*	3,882
8	−.016*	−.011*	−.008*	−.008$^+$	−.010	4,402

Note.—Class size variable indicates actual class size in grade 3.
$^+$ $p < .10$.
* $p < .05$.

Overall, these estimates were qualitatively similar to these reported in tables 1 and 2 and suggested that small-class effects in grade 3 persisted through grade 8 across the entire achievement distribution. The estimates in the tails were comparable overall; all coefficients were negative, and more than 50 percent of them were significant at the .05 level. As in tables 1 and 2, these results clearly indicate that all types of students (low, medium, and high achievers) in grades 4–8 benefit from being in small classes in earlier grades. These results also suggest that the small-class effects (across the entire achievement distribution) were most likely unaffected by the larger-than-intended variability in actual class size.

The Gender, Race, and SES Gap

We also examined the gender, race, and SES gap in achievement. The results of this analysis are summarized in table 4. We first discuss the results for

mathematics, then for reading, and finally for science. In mathematics, female students outperformed their male peers by nearly one-tenth of a standard deviation in the middle of the distribution. The female advantage was more pronounced for lower achievers (nearly one-third of a standard deviation) but less pronounced for higher achievers. White students outperformed minority students in the middle of the distribution by one-fourth of a standard deviation in grades 4 and 5 and by more than one-third of a standard deviation in grades 7 and 8. The race gap was comparable for low and high achievers in grades 4 and 5, whereas in grades 7 and 8 the achievement gap was more pronounced for high achievers, with whites performing nearly one-half of a standard deviation higher. Low-SES students achieved significantly lower than high-SES students across the entire distribution of mathematics achievement. The SES gap was larger than the race and the gender gap overall.

The results for reading achievement were comparable. Again, female students outperformed their male peers by nearly one-tenth of a standard deviation in the middle of the distribution, and the female advantage was more pronounced for lower achievers (nearly one-third of a standard deviation). Also, white students outperformed minority students in the middle of the distribution by more than one-third of a standard deviation in grades 4 and 5 and by nearly one-half of a standard deviation in grade 7. The race gap was comparable overall for low and high achievers across grades. High-SES students outperformed their low-SES peers across the entire distribution of mathematics achievement. The SES gap was typically more pronounced for low achievers and in grades 6 and 7 for high achievers. The SES gap was larger than the race and the gender gap overall.

In science the gender gap was reversed favoring male students. In particular, male students outperformed their female peers by more than one-tenth of a standard deviation in the middle of the distribution in grades 4, 7, and 8. The female advantage was more pronounced for higher achievers, especially in grades 7 and 8. The race gap was larger in science compared to that in mathematics and reading. The race gap was also more pronounced for low achievers in grades 4–8. It is noteworthy that in science the SES gap was smaller than the race gap. Still, high-SES students outperformed their low-SES peers across the entire distribution of science achievement. The SES gap was more pronounced for low achievers.

In addition, we estimated interaction effects between class size and gender, race, and SES across the distribution of achievement. Overall, across the entire distribution of achievement tests and across grades, the interaction estimates were small and insignificant. Only 12 of the 225 interactions we examined were significant, which indicates that these results could have occurred by chance (the probability is approximately 5 percent). Most of the interaction effects were significant in science and fewer in reading and mathematics. There

TABLE 4

Estimated Gender, Minority, and SES Effects on Mathematics, Reading, and Science at Various Quantiles

	QUANTILE				
	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth
Mathematics:					
Grade 4:					
Gender	.323*	.215*	.108*	.108*	.086*
Minority	−.215*	−.301*	−.258*	−.280*	−.237*
Low SES	−.688*	−.495*	−.452*	−.409*	−.430*
Grade 5:					
Gender	.285*	.143*	.071$^+$	.048	−.024
Minority	−.214*	−.356*	−.238*	−.214*	−.238*
Low SES	−.618*	−.404*	−.428*	−.404*	−.475*
Grade 6:					
Gender	.375*	.234*	.140*	.117*	.070
Minority	.023	−.152	−.140	−.140	.000
Low SES	−.632*	−.457*	−.421*	−.398*	−.538*
Grade 7:					
Gender	.439*	.208*	.069*	−.023	−.116
Minority	−.116*	−.277*	−.439*	−.439*	−.462*
Low SES	−.532*	−.370*	−.393*	−.416*	−.555*
Grade 8:					
Gender	.267*	.165*	.113*	.021	−.021
Minority	−.216*	−.288*	−.381*	−.453*	−.555*
Low SES	−.627	−.473*	−.473*	−.473*	−.555*
Reading:					
Grade 4:					
Gender	.243*	.142*	.101*	.061$^+$	.091*
Minority	−.385*	−.364*	−.405*	−.385*	−.445*
Low SES	−.952*	−.587*	−.445*	−.466*	−.496*
Grade 5:					
Gender	.288*	.177*	.133*	.022	−.022
Minority	−.687*	−.332*	−.399*	−.465*	−.465*
Low SES	−.709*	−.510*	−.421*	−.399*	−.554*
Grade 6:					
Gender	.369*	.250*	.152*	.065	.022
Minority	−.521*	−.315*	−.326*	−.326*	−.282*
Low SES	−.803*	−.488*	−.412*	−.456*	−.586*
Grade 7:					
Gender	.345*	.184*	.115*	.069*	.023
Minority	−.482*	−.505*	−.528*	−.482*	−.505*
Low SES	−.597*	−.459*	−.436*	−.436*	−.620*
Grade 8:					
Gender	.292*	.157*	.067$^+$	.022	−.022
Minority	−.337*	−.404*	−.472*	−.517*	−.629*
Low SES	−.719*	−.562*	−.472*	−.427*	−.427*

TABLE 4 (*Continued*)

	QUANTILE				
	Tenth	Twenty-fifth	Fiftieth	Seventy-five	Ninetieth
Science:					
Grade 4:					
Gender	.150*	−.056	−.150*	−.150*	−.150*
Minority	−.356*	−.469*	−.460*	−.469*	−.375*
Low SES	−1.050*	−.563*	−.375*	−.319*	−.394*
Grade 5:					
Gender	.063	−.063	−.063*	−.167*	−.250*
Minority	−.792*	−.542*	−.438*	−.396*	−.458*
Low SES	−.729	−.479*	−.375*	−.333*	−.375*
Grade 6:					
Gender	.117	−.019	−.039	−.078*	−.156*
Minority	−.623*	−.584*	−.486*	−.350*	−.370*
Low SES	−.876*	−.506*	−.409*	−.370*	−.409*
Grade 7:					
Gender	.090*	.000	−.113*	−.135*	−.248*
Minority	−.653*	−.608*	−.540*	−.585*	−.585*
Low SES	−.630*	−.383*	−.360*	−.450*	−.473*
Grade 8:					
Gender	−.044	−.133*	−.155*	−.288*	−.310*
Minority	−.620*	−.509*	−.443*	−.443*	−.498*
Low SES	−.642*	−.443*	−.354*	−.399*	−.454*

NOTE.—SES = socioeconomic status.
$^+$ $p < .10$.
* $p < .05$.

were a few exceptions, however. For example, in grade 6, female students in the ninetieth percentile of the mathematics distribution benefited less from being in small classes. Similarly, in grade 7, female students in the tenth percentile of the mathematics distribution benefited less from being in small classes. In grade 8, low-SES students in the tenth percentile of the reading or science distribution benefited much from being in small classes. Finally, low-SES students in the ninetieth percentile of the science distribution benefited from being in small classes in grade 5, but the effect was reversed in grades 4 and 7. Nonetheless, the results suggested little evidence of interaction effects.

Cumulative Effects of Small Classes

Project STAR is a longitudinal study, and the students who participated in it experienced different lengths of exposure to the treatment. For example, stu-

dents could be in small classes for zero, one, two, three, or four years. An interesting question, then, is whether a "dose-response" relationship was observed and persisted over time. That is, it is plausible that longer exposure to the treatment would produce larger (cumulative) small-class effects that persist over time. A related issue is whether the additive exposure to the treatment would indeed increase student achievement meaningfully or whether a one-time exposure (e.g., one year) is sufficient. To model the cumulative effects of small classes, we created four binary indicators that examined the differences between different degrees of exposure in small classes (e.g., zero, one, two, and three vs. four years in small classes). The estimates of these analyses are summarized in tables 5–7. Note that the overwhelming majority (more than 95 percent) of students who participated in Project Star were in small classes in successive years (two, three, or four years). The coefficients reported in tables 5–7 indicate the small-class advantage of being in those classes for all four years of the experiment.

The results for mathematics achievement are summarized in table 5. Notice that all estimates are positive and large. In addition, the median differences are overwhelmingly significantly different from zero. The results in the median indicated that the effects of small classes are cumulative and that multiple years of exposure produce larger effects in mathematics. The results were more pronounced in grade 3 and became smaller over time, as one would expect. However, they persisted through grade 8, and the cumulative class size effects were on average at least one-fifth of a standard deviation in grades 7 and 8. The cumulative effects in the tails also lasted through grade 8, and the results in the tails were comparable qualitatively and followed patterns similar to those in the middle of the achievement distribution. This suggested that both low and high achievers benefit equally from longer exposure to small classes. However, in one occasion in grade 4, the estimated difference between four versus one year in small classes was nearly one-half of a standard deviation at the tenth percentile, which is a large effect in education. In fact, the estimate at the tenth percentile was significantly larger than that at the ninetieth percentile (see table 8). Generally, however, the cumulative effects of small classes seem to have positively affected all types of students. It should also be noted that overall, the achievement differences between students who were in small classes for one year and those who were in small classes all four years were similar to those between students who were never in small classes and those who were in small classes all four years. This finding indicates that perhaps two or more years in small classes in early grades are needed to obtain larger cumulative effects.

The results for reading achievement are summarized in table 6, which has the same structure as table 5. Again, all estimates are positive and large. The median differences are overwhelmingly significantly different from zero. The

Estimated Cumulative Small-Class Effects on Mathematics at Various Quantiles

CUMULATIVE SMALL-CLASS EFFECT (Years in Small Classes)	QUANTILE					N
	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	
Grade 3:						6,025
Four versus zero	.301*	.213*	.326*	.301*	.326*	
Four versus one	.301*	.226*	.326*	.351*	.251*	
Four versus two	.113	.050	.276*	.226*	.226$^+$	
Four versus three	.188*	.075	.276*	.251*	.326*	
Grade 4:						4,305
Four versus zero	.258*	.188*	.172*	.194*	.172*	
Four versus one	.473*	.263$^+$	.215*	.194*	.065	
Four versus two	.129$^+$	.215*	.086$^+$	.108	.065	
Four versus three	.065	.011	.108	.086	.194*	
Grade 5:						2,505
Four versus zero	.238*	.261*	.214*	.238*	.261*	
Four versus one	.190$^+$	.261$^+$	.261*	.261*	.285	
Four versus two	.214	.261*	.119	.190	.238$^+$	
Four versus three	.000	.048	.166$^+$	.190$^+$	.166*	
Grade 6:						3,450
Four versus zero	.234*	.140$^+$	.187*	.164$^+$	.304*	
Four versus one	.117	.117	.164*	.047	.070	
Four versus two	.281$^+$	.117	.117	.140	.211$^+$	
Four versus three	.047	.117	.211$^+$	.164	.211$^+$	
Grade 7:						3,896
Four versus zero	.231*	.231*	.254*	.208*	.231$^+$	
Four versus one	.231*	.254*	.231*	.162*	.162	
Four versus two	.370*	.277*	.200*	.185*	.162	
Four versus three	.162$^+$	.208*	.231*	.162*	.046	
Grade 8:						4,411
Four versus zero	.237*	.247*	.206*	.144*	.247*	
Four versus one	.432*	.309*	.309*	.288*	.432*	
Four versus two	.216*	.288*	.206*	.082	.226	
Four versus three	.206$^+$	.288*	.165*	.144$^+$	.309*	

$^+$ $p < .10$.
* $p < .05$.

results in the median indicate that the effects of small classes are cumulative and that multiple years of exposure produce larger effects in reading. In addition, although the effects become smaller over time, they persist through grade 8. The largest estimates were obtained in grade 3 and were larger than those in mathematics in table 5. The results in the tails are somewhat similar and point to larger benefits for low achievers than for high achievers in grades 4–8. The estimates at the tenth percentile are typically larger than those in

TABLE 6

Estimated Cumulative Small-Class Effects on Reading at Various Quantiles

CUMULATIVE SMALL-CLASS EFFECT (Years in Small Classes)	QUANTILE					N
	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	
Grade 3:						5,947
Four versus zero	.311*	.311*	.415*	.363*	.389*	
Four versus one	.441*	.415*	.415*	.428*	.207*	
Four versus two	.311*	.233*	.233*	.259*	.233$^+$	
Four versus three	.233*	.233*	.311*	.272*	.285*	
Grade 4:						4,225
Four versus zero	.263*	.263*	.182*	.202*	.182*	
Four versus one	.445*	.344*	.233*	.243*	.162	
Four versus two	.162	.182*	.152*	.182$^+$	.121	
Four versus three	.162	.223*	.081	.081	.040	
Grade 5:						2,507
Four versus zero	.399*	.199*	.199*	.222*	.199	
Four versus one	.244	.177$^+$	.199*	.288*	.199	
Four versus two	.377	.133$^+$	.111	.177*	.199	
Four versus three	.222	.111	.111	.155	.066	
Grade 6:						3,454
Four versus zero	.369*	.217*	.152*	.195*	.152	
Four versus one	.412*	.109	.065	.141	.130	
Four versus two	.369*	.152*	.130	.152*	.174	
Four versus three	.174$^+$	.239*	.130	.184$^+$	.043	
Grade 7:						3,902
Four versus zero	.322*	.253*	.207*	.161*	.207*	
Four versus one	.276*	.184*	.161*	.115$^+$	.184	
Four versus two	.253$^+$	.184*	.138	.161$^+$	.276	
Four versus three	.230$^+$	.253$^+$	.092	.069	.000	
Grade 8:						4,413
Four versus zero	.292*	.247*	.157*	.135*	.157$^+$	
Four versus one	.314*	.314*	.225*	.180*	.157	
Four versus two	.225*	.314*	.180*	.202*	.225	
Four versus three	.067	.135	.112	.135	.090	

$^+$ $p < .10$.
* $p < .05$.

the ninetieth percentile and also more likely to be significant at the .05 level. For example, in two occasions in grades 4 and 6 the estimated differences between four versus one year in small classes were slightly larger than two-fifths of a standard deviation at the tenth percentile, a large effect in education. In fact, these estimates at the tenth percentile were significantly larger than those at the fiftieth percentile, and in grade 4 the estimate was also significantly larger than that at the ninetieth percentile. As in mathematics, the cumulative

TABLE 7

Estimated Cumulative Small-Class Effects on Science at Various Quantiles

CUMULATIVE SMALL-CLASS EFFECT (Years in Small Classes)	QUANTILE					N
	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	
Grade 3:						6,266
Four versus zero	.311*	.311*	.311*	.324*	.311*	
Four versus one	.207*	.311*	.182*	.233*	.311*	
Four versus two	.207*	.233*	.182*	.233*	.311*	
Four versus three	.104	.207*	.182*	.169[+]	.311*	
Grade 4:						4,298
Four versus zero	.263*	.188*	.131*	.150*	.150*	
Four versus one	.300	.150*	.103	.188*	.131	
Four versus two	.131	.056	.028	.075	.075	
Four versus three	.225	.113	.075	.075	.000	
Grade 5:						2,502
Four versus zero	.313*	.167*	.146*	.146*	.188*	
Four versus one	.188	.146	.156*	.208*	.208[+]	
Four versus two	.188	.042	.073	.042	.063	
Four versus three	.125	.188[+]	.146[+]	.188*	.083	
Grade 6:						3,443
Four versus zero	.331*	.195*	.136*	.156*	.097	
Four versus one	.039	.058	.097[+]	.078	.039	
Four versus two	.156	.097	.078	.039	.058	
Four versus three	.331*	.156	.156*	.195*	.136	
Grade 7:						3,882
Four versus zero	.338*	.248*	.203*	.225*	.259*	
Four versus one	.450*	.248*	.135	.248*	.203	
Four versus two	.248	.270*	.090	.248*	.248[+]	
Four versus three	.135[+]	.203*	.180*	.203*	.079	
Grade 8:						4,402
Four versus zero	.199*	.155*	.155*	.089*	.111	
Four versus one	.221	.221*	.177*	.133*	.111	
Four versus two	.133	.133[+]	.133[+]	.022	.044	
Four versus three	.022	.066	.044	.000	.000	

[+] $p < .10$.
* $p < .05$.

effects in the tails lasted through grade 8. Overall, as in mathematics, the achievement differences between students who were in small classes for one year and those who were in small classes all four years were similar to those between students who were never in small classes and those who were in small classes all four years.

The results for science achievement are summarized in table 7, which has the same structure as tables 5 and 6. Again, all estimates are positive but not

T-Tests Indicating Differences in Quantile Regression Estimates: Cumulative Small-Class Effects

	QUANTILE	
YEARS IN SMALL CLASSES	Tenth versus Fiftieth	Tenth versus Ninetieth
Grade 4:		
Mathematics:		
Four versus one	1.429	2.415*
Reading:		
Four versus one	2.936*	3.543*
Grade 6:		
Reading:		
Four versus one	2.370*	1.050
Science:		
Four versus zero	2.699*	2.412*

* $p < .05$.

as large as those reported in tables 5 and 6. The median differences are typically significantly different from zero. The results in the median indicate that the effects of small classes are cumulative and that multiple years of exposure produce larger effects in science. As in mathematics and reading, the cumulative effects in science become smaller over time but persist through grade 8. The results in the tails are somewhat similar and point to larger benefits for low achievers than for high achievers. As in reading, the estimates at the tenth percentile are typically larger than those in the ninetieth percentile (except grade 3). In one occasion, in grade 6, the estimated difference between four versus zero years in small classes was nearly one-third of a standard deviation at the tenth percentile. In fact, the estimate at the tenth percentile was significantly larger than that at the ninetieth percentile. As in mathematics and reading, the cumulative effects in the tails lasted through grade 8. Nonetheless, the class size estimates in science were smaller than those in mathematics and reading, indicating that the cumulative effects of small classes matter somewhat less in science in later grades.

As mentioned above, we examined whether the cumulative small-class estimates varied across the different percentiles and whether these differences were statistically significant. Table 8 summarizes some of the *t*-tests we used to examine this question in certain grades. Most of the *t*-tests in the table showed significant differences in achievement between different percentiles. This indicates that in some later grades the very low achievers benefited significantly more from being in small classes for four years in early grades than did medium or very high achievers. However, this evidence is weak since

small-class effects were similar for all types of students generally, and the majority of the *t*-tests we constructed (not presented in this table) were insignificant.

Conclusion

Previous work that used data from Project STAR demonstrated consistently that being in small classes in early grades leads to higher student achievement on average. Given that class size reduction is an educational intervention that benefits all students by increasing their achievement, it is tempting to expect that it could also reduce the achievement gap between high and low achievers. However, previous findings provided weak or no evidence that reducing class size has immediate benefits for low-achieving students more than other students (Nye et al. 2002). More recent work provided weak evidence that the cumulative effects of small classes may have long-term benefits for minority students in grades 4–8 (see Nye, Hedges, and Konstantopoulos 2004). The current study examined the durability of the small-class effects in early grades through grade 8 across the entire achievement distribution in an attempt to better understand the effects of class size reduction on the achievement gap.

The results from the quantile regression analyses provided convincing evidence that all types of students (e.g., low, medium, and high achievers) benefit from being in small classes (in early grades) across all achievement tests. Specifically, being in small classes in grade 3 resulted in considerable increases in student achievement that last to grade 8. The results are consistent over time and suggest that the effects of small classes in early grades have similar lasting benefits for different types of students, such as low, medium, and high achievers. These findings do not point to differential effects of small classes across different types of students—that is, it appears that all types of students benefit equally from being in small classes in early grades. This study shows virtually no evidence that manipulating class size in the third grade can reduce the achievement gap between low- and high-achieving students in later grades. These findings are consistent with those reported by Nye and colleagues (2002).

Our analyses also addressed the possible effects of validity threats, such as students' switching between types of classrooms and larger variability than intended by design in actual class size, which resulted in overlap in actual class size between smaller and larger classes. The results of these analyses were consistent with those in the original analysis and further supported the notion that small-class effects persist over time for all types of students. Thus, these validity threats appear to have virtually no effect on the small-class estimates across the achievement distribution. Again, however, we did not find any evidence of additional lasting benefits of small classes for low achievers or

other students. This result should be interpreted with caution. It does not mean that low-achieving students are better off in larger classes because all students benefit from being in small classes. In fact, the estimates in tables 1–3 suggest that low achievers in small classes are better off than their peers in regular classes.

This study also demonstrates that longer exposure to small classes in grades kindergarten to 3 produces higher levels of achievement for low, medium, and high achievers in grades 4–8. This shows that additional years in small classes in early grades result in benefits in achievement in later grades for all students. This finding is consistent with that reported by Nye and colleagues (1999). Our results also suggested that one year of exposure in small classes in not enough and that students benefit more from additional years in small classes. It is noteworthy that in certain grades, in reading and science, the cumulative effects of small classes for low achievers are substantial in magnitude and significantly different from those for high achievers. Hence, there is some weak evidence that the cumulative effects of small classes have larger lasting benefits for low achievers, and this finding is in congruence with that reported by Nye, Hedges, and Konstantopoulos (2004) about minority students. This is not surprising since in our samples, on average, 50 and 70 percent of low achievers (i.e., the tenth percentile) were minority and low-SES students, respectively. That is, as expected, there is a higher proportion of minority and low-SES students in the low-achieving group.

One hypothesis is that in small classes teachers are more likely to identify low achievers and hence are more likely to provide instruction designed to benefit these students in early grades. Alternatively, in small classes there is a higher likelihood for low achievers to interact with teachers and be more engaged in learning. When this kind of classroom practice takes place for several years in early grades (K–3), it appears that it helps low achievers more than other students in later grades (4–8). It must be noted, however, that these results should not be interpreted as casual effects because it is not obvious that the mechanism that assigned students to longer or shorter periods in small classes in early grades was random. That is, these results simply suggest (strongly nonetheless) that longer exposure to small classes in early grades increases student achievement in later grades for all students and for low achievers in particular.

Because students and teachers are not assigned randomly to schools but rather select their schools, it is likely that estimates of class size are affected by differences in schools (e.g., school effects). Previous analyses of Project STAR data that computed class size effects adjusted for school fixed effects via school dummies (see, e.g., Konstantopoulos 2008; Krueger 1999). Similarly, it was important to control for school effects in the current study to adjust adequately for possible differences across schools. As a result, we replicated

some of the analyses described earlier, including school fixed effects (as dummies) in the quantile regressions. The results were generally similar to those reported in tables 1–7, which indicates that schools did not affect the small-class estimates across the achievement distribution, and, hence, our results are robust.

Results suggested that the effects of small classes were smaller for science achievement. Unfortunately, observational data from classrooms were not available for Project STAR, and hence it is difficult to know exactly how science was taught in small and regular classes. Our results suggest that in grades 4–8 science teaching is equally effective in different types of classrooms. One plausible hypothesis is that the structure of instruction or classroom practices in science is comparable across the different types of classrooms (e.g., small and regular classes). For example, if science activities typically take place in small groups (across all types of classrooms) in which the majority of students (low, average, and high achievers) engage actively in science learning, then it is likely that there would not be any significant differences in science achievement between small and regular classes. If this hypothesis were true, the small-class advantage would dissipate because of the way teachers enact science activities in the classroom, so that the students in different types of classrooms receive similar instruction and help for learning. Another possibility is that elementary school teachers may have minimal background in science (e.g., a small number of science courses in college and low levels of professional development) and minimal experience in using instructional technology to teach science and hence may not feel adequately prepared for teaching science effectively. If only a small percentage of teachers have undergraduate degrees in science or science education and the majority of teachers feel that they are not well qualified to teach science, then they will be less likely to put much emphasis on science teaching. If this hypothesis were true, science teaching would not be as effective in elementary school grades for all types of classrooms, and effectiveness in science teaching would not vary much by class size. In that case, small-class effects would not necessarily last in later grades (4–8) for science achievement.

Our results show that, overall, small classes are beneficial to student achievement. However, we caution the reader that the effectiveness of class size reduction efforts is closely linked to the proper implementation of such programs. Indeed, evidence from Project STAR has consistently suggested that small-class membership benefits all kinds of students. These promising findings encouraged class size reduction policies and implementations in many districts and states. California and Florida are some of the states that implemented class size reduction programs. Although the evidence from the Tennessee Class Size Experiment has been strong regarding the benefits of being in small classes, in California and Florida evidence has not demonstrated major ad-

vantages in achievement for students in small classes. This inconsistency in the findings is partly due to the lack of support to allocate the necessary resources that would make class size reduction programs successful. That is, in order to implement class size reduction programs on a large scale (e.g., throughout a state), many necessary conditions need to be satisfied. For example, to implement such programs successfully, more classrooms and teachers are needed, which indicates that states and districts need to commit themselves to allocating sufficient funds. In other words, schools need to be able to afford having new classrooms that are comparable to those already in use and hiring new well-qualified teachers. This is more likely to take place gradually, assuming the funding for new classrooms and teachers is available for longer periods, and requires adequate planning and long-term commitment and investment to such programs. Class size reduction is likely a more effective school intervention when school organizational and pedagogical conditions are intact. In the short run, however, the lack of space in schools and the shortage of well-trained teachers will most likely hinder the successful implementation of such programs, which was witnessed in California and Florida (see Biddle and Berliner 2002). It is also likely that inadequate funding may be more detrimental for poorer districts because such districts may not be able to attract and keep well-trained teachers or have access to more classrooms, and this further promotes educational inequality.

This study helped us further understand the effects of small classes on student achievement. However, the mechanism through which small classes affect student achievement in later grades is still not clearly defined. Nonetheless, it is remarkable that an intervention that is easily defined and implemented can have important lasting benefits at least to the end of middle school for all students. It is even more encouraging that the cumulative benefits seem (for some achievement tests and in certain grades) more pronounced for low achievers. Thus, class size reduction appears to be an intervention that increases the achievement levels for all students while simultaneously reducing the achievement gap.

References

Angrist, Joshua, Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–72.

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (2): 533–75.

Biddle, Bruce J., and David C. Berliner. 2002. "Small Class Size and Its Effects." *Educational Leadership* 59 (8): 12–23.

Buchinsky, Moshe. 1998. "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research." *Journal of Human Resources* 33 (1): 89–126.

Finn, Jeremy D., and Charles M. Achilles. 1990. "Answers and Questions about Class Size: A Statewide Experiment." *American Educational Research Journal* 27 (3): 557–77.

Finn, Jeremy D., Susan Gerber, Charles Achilles, and Jayne Boyd-Zaharias. 2001. "The Enduring Effects of Small Classes." *Teachers College Record* 103 (2): 145–83.

Hao, Lingxin, and Daniel Q. Naiman. 2007. *Quantile Regression*. Thousand Oaks, CA: Sage.

Hedges, Larry V., and Amy Nowell. 1995. "Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals." *Science* 269 (5220): 41–45.

Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115 (4): 1239–85.

Koenker, Roger, and Gilbert Bassett. 1978. "Regression Quantiles." *Econometrica* 46 (1): 33–50.

Konstantopoulos, Spyros. 2008. "Do Small Classes Reduce the Achievement Gap between Low and High Achievers? Evidence from Project STAR." *Elementary School Journal* 108 (4): 275–91.

Konstantopoulos, Spyros, and Amelie Constant. 2008. "The Gender Gap Reloaded: Are School Characteristics Linked to Labor Market Performance?" *Social Science Research* 37 (2): 374–85.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114 (2): 497–532.

Krueger, Alan B., and Diane M. Whitmore. 2001. "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." *Economic Journal* 111 (468): 1–28.

Milesi, Carolina, and Adam Gamoran. 2006. "Effects of Class Size and Instruction on Kindergarten Achievement." *Educational Evaluation and Policy Analysis* 28 (4): 287–313.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 1999. "The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment." *Educational Evaluation and Policy Analysis* 21 (2): 127–42.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 2000a. "Do the Disadvantaged Benefit from the Small Classes? Evidence from the Tennessee Class Size Experiment." *American Journal of Education* 109 (1): 1–26.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 2000b. "Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class Size Experiment." *American Educational Research Journal* 37 (1): 123–51.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 2002. "Do Low-Achieving Students Benefit More from Small Classes? Evidence from the Tennessee Class Size Experiment." *Educational Evaluation and Policy Analysis* 24 (3): 201–17.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 2004. "Do Minorities Experience Larger Lasting Benefits from Small Classes? Evidence from a Five-Year Follow-Up of the Tennessee Class Size Experiment." *Journal of Educational Research* 98 (2): 94–100.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26 (3): 237–57.

Nye, Barbara A., Jayne B. Zacharias, B. D. Fulton, Charles M. Achilles, V. A. Cain, and D. A. Tollett. 1994. "The Lasting Benefits Study: A Continuing Analysis of the Effect of Small Class Size in Kindergarten through Third Grade on Student Achievement Test Scores in Subsequent Grade Levels; Seventh Grade." Technical report, Center of Excellence for Research in Basic Skills, Tennessee State University.

Pong, Suet-ling, and Aaron Pallas. 2001. "Class Size and Eighth-Grade Math Achievement in the United States and Abroad." *Educational Evaluation and Policy Analysis* 23 (3): 251–73.

Wossmann, Ludgar, and Martin West. 2006. "Class-Size Effects in School Systems around the World: Evidence from between-Grade Variation in TIMSS." *European Economic Review* 50 (3): 695–736.