Do Small Classes Reduce the Achievement Gap between Low and High Achievers?

Evidence from Project STAR

by

Spyros Konstantopoulos

Northwestern University

Amelie Constant

IZA

and

Georgetown University

Abstract

Previous research from randomized experiments has indicated positive immediate and long-term effects of small classes on average student achievement for all students. However, thus far the effects of small classes on the achievement gap between lower and higher achieving students, have not been fully examined. Given that previous findings on the social distribution of the small class effects are mixed and inconclusive, the present study attempts to shed some more light on the mechanism through which small classes impact the achievement of low and high achieving students. The authors used data from a four-year large-scale randomized experiment (Project STAR) to examine the effects of small classes on the achievement gap. The results obtained from different kinds of analyses consistently indicate that higher achieving students benefit even more from being in small classes in early grades than other students. Our findings also indicate that all students benefit from being in small classes. However, our findings do not indicate that small class reduction is a mechanism that can reduce the achievement gap.

The effects of class size on student achievement have been of great interest for educational researchers and policy makers the last two decades. Reducing class size to boost student achievement is a policy option that has gained considerable attention nationwide. Currently, many states and school districts have enacted or are considering class size reduction with the objective of improving academic achievement.

Previous studies that used high quality experimental data have consistently demonstrated the positive effects of small size classes on average student achievement for all students (e.g., Finn & Achilles, 1990; Krueger, 1999; Nye, Hedges, & Konstantopoulos, 2000). Specifically, these studies indicated that the average student achievement in small size classes is significantly higher than the average student achievement in regular size classes. These findings suggest that reducing class size is a promising intervention that increases academic achievement on average for all students.

However, it is tempting to imagine class size reduction as an educational intervention that increases academic achievement for all students and reduces the achievement gap between lower and higher achieving students by producing larger gains for lower achieving students. The important question of whether class size reduction can reduce the achievement gap and hence impact the academic achievement of low and high achieving students differently has not been fully answered thus far. The present study attempts to answer this question by examining differences in achievement variability between small and regular size classes using data from a four-year large-scale randomized experiment conducted in Tennessee in the mid 1980s. The present study also examines differences in academic achievement between students in small and in regular size classes in the upper and lower tails of the achievement distribution. Observed differences in achievement variability and in the tails of the achievement distribution between small and regular size classes would indicate that small classes have differential effects on different groups of students.

Examining 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. However, focusing on average differences of achievement distributions between small and regular size classes is only one way to evaluate the effects of class size. A more complete assessment of the effects of class size would also examine differences in the variability of student achievement between small and regular size classes as well as differences in the upper and lower tails of the achievement distribution. Specifically, differences in variability (in a specific outcome) between treatment and control groups in experimental studies provide important evidence about interactions between treatments and individuals' characteristics (see Bryk & Raudenbush, 1988). For example, differences in achievement variability between treatment and control groups may indicate that a treatment has differential effects on different groups of students, that is, some student groups may benefit more from being exposed to the treatment than others. This notion of interaction between treatments and individual characteristics goes back to the pioneering work of Cronbach and Snow (1977). The underlying idea is that a treatment may have one effect on some individuals and a different effect on others. In this study we follow Cronbach's and Snow's definition about interactions and examine whether different groups of students (such as lower and higher achieving students) benefit more or less from receiving a treatment (such as being in small classes).

Reducing class size can potentially impact the means as well as the variances of the achievement distributions of small and regular size classes. That is, class size reduction can also produce differences in the variability of student achievement between small and regular size classes. Class size reduction can affect student achievement variability in three ways: (a) the variability of student achievement in small size classes may be smaller than that in regular classes, (b) the variability of student achievement in small size classes may be larger than that in regular classes, and (c) the variability of student achievement in small size classes may be similar to that in regular classes. Smaller variability in student achievement in small size classes (than in regular size classes) suggests that the achievement gap between lower and higher achieving students is smaller in these types of classes. In contrast, larger variability in student achievement in small size classes (than in regular size classes) suggests that the achievement gap between lower and higher achieving students is larger in these types of classes. In the same vein, similar variability in student achievement in small and in regular size classes suggests that the achievement gap between lower and higher achieving students is similar in both types of classes. In addition, differences in achievement variability between small and regular size classes may indicate that achievement differences in the middle of the achievement distribution are qualitatively different that achievement differences in the tails. For example, achievement differences between lower achieving students in small and in regular size classes may be significantly smaller or larger than achievement differences between average or high achieving students in these classes.

Hypotheses About the Class Size Mechanism

Class size reduction can affect the achievement gap in three ways. First, if higher achieving students benefit more than lower achieving students from being in small size classes, then one would expect larger variability in student achievement in small size classes than in regular size classes. In this case the achievement distribution in small size classes will have a higher mean and a larger standard deviation (since it will be more spread out) than the achievement distribution in regular size classes. This may also indicate that achievement differences between students in small and in regular size classes are larger in the upper tail of the achievement distribution (higher achieving students) than in the lower tail (lower achieving students). If this hypothesis were true, then small size classes would not close the achievement gap between lower and higher achieving students. Second, if lower achieving students benefit more from being in small size classes, then one would expect smaller variability in student achievement in small size classes than in regular size classes. In this case the achievement distribution in small size classes will have a higher mean and a smaller standard deviation (since it will be less spread out) than the achievement distribution in regular size classes. This may also indicate that achievement differences between students in small and in regular size classes are larger in the lower tail of the achievement distribution (lower achieving students) than in the upper tail (higher achieving students). If this hypothesis were true, then small size classes would close the achievement gap between lower and higher achieving students since the lower achieving students would benefit more from being in small size classes than other students in the same types of classes. Third, if small and regular size classes help higher and lower achieving students similarly, then one would expect that the variability in student achievement in small size classes is comparable to the variability in student achievement in regular size classes. In this case the achievement distribution in small size classes will have a higher mean than the achievement distribution in regular size classes but a comparable standard deviation. That is the achievement distribution in small size classes is simply shifted to the right by about 1/5 of a standard deviation (SD), which is the average achievement benefit reported in previous studies (e.g., Nye et al., 2000). If this hypothesis were true, then small size classes would have no impact on the achievement gap between lower and higher achieving students since low, average, and high achieving students would benefit equally from being in small size classes.

There are different ways to investigate these hypotheses. First, one could examine the differential effects of small size classes on the achievement of low achieving, minority, and disadvantaged students. Some recent studies investigated this issue and found weak evidence that small size classes help low achieving, minority, or disadvantaged students more than other students (e.g., Nye, Hedges, & Konstantopoulos, 2001, 2002, 2004). Alternatively, one could examine differences in the variability of student achievement in small and regular size classes. Notice that differences in achievement variability between small and regular size classes indicate that the treatment has differential effects across different types of students (e.g., higher or lower achieving students). That is, the effects of the treatment may be different in the lower and upper tails of the achievement distribution. The present study examines differences in the variability of student achievement between small and regular size classes as well as differences in the upper and lower tails of the achievement distribution in an attempt to better understand the class size mechanism using data from a four-year large-scale randomized experiment conducted in Tennessee in the mid 1980s. This study also addresses issues related to the internal validity of Project STAR such as student switching among different types of classes and more than intended variability in actual class size within different types of classes.

Results from Previous Studies

There is some evidence from previous work that treatments can affect both the mean and the variance of a continuous outcome of interest. Earlier work on evaluating effects of educational interventions has indicated that such programs can change not only average student achievement, but the variability in student achievement as well. For example, research on resource allocation in schools argues that an important criterion variable for reform programs that aim to equalize school funding is not the average per pupil expenditure across school districts, but the variability of per pupil expenditure across school districts (Bowles & Levin, 1968; Monk, 1981). Reviews of research where students are assigned to different learning conditions (e.g., tutoring, mastery learning, and conventional) have also reported differences among the three conditions in average achievement as well as in achievement variability (Bloom, 1984). Studies of cross-national comparisons of student achievement have also found that countries with larger achievement gains in central tendency also had larger gains in achievement variability (Coleman, 1985). Specifically, Japan had not only the largest average achievement gain, but the largest gain in achievement variability as well. In addition, significant associations between school size, variation in mathematics course taking, sector, and variability in student achievement have also been reported using High School and Beyond data (Raudenbush & Bryk, 1987). Nonetheless, in the present study it is not obvious how class size reduction will affect achievement variability. Thus, our research hypothesis is that achievement variability in small size classes will be different than that in regular size classes (a non-directional hypothesis).

Related literature

The effects of class size reduction on student achievement have been empirically examined via various research designs over the past few decades. Numerous experimental and quasi-experimental studies have investigated the effects of class size on student achievement and have been reviewed by Glass and Smith (1979), Glass et al. (1982), Hedges and Stock (1983), and Mosteller, Light, and Sacks (1996). Overall, these reviews have indicated that class size reduction has positive effects on student achievement and these effects become larger as the class size becomes smaller. Nonetheless, the majority of the studies are small-scale and short term and even though their results may have high internal validity, the generality of their findings may be limited.

Another line of research examined the effects of class size reduction via education production function studies (see e.g., Hanushek, 1986). Typically such studies compute the association between class size and student achievement adjusting for important student variables such as race/ethnicity, social class, and previous achievement. The interpretation of the results of these econometric studies has been controversial. While some reviewers argue that the effects of class size are small and in many studies statistically insignificant (e.g., Hanushek, 1989), others contend that the magnitude of the estimates of the mean differences in student achievement are a better way to assess class size effects than statistical significance (e.g., Greenwald, Hedges, & Laine, 1996; Hedges, Laine, & Greenwald, 1994). Even though these studies are usually large-scale and hence their results may have higher external validity, their internal validity may be limited since it is not obvious that the association between class size and achievement is causal (that is class size may be endogenous). For example, it is likely that student achievement defines class membership. In addition, omitted variable bias is possible in these large scale observational studies and this can bias the estimates of the class size effects. Finally, the key independent variable (class size) is typically constructed using school size and number of teachers in the school and hence it is not an accurate but an aggregate measure of class size.

The Tennessee Class Size Experiment

The Tennessee class size experiment or Project STAR (Student-Teacher Achievement Ratio) is discussed in detail elsewhere (see, e.g., Krueger, 1999; Nye, Hedges, and Konstantopoulos, 2000). The experiment involved students in 79 elementary schools in 42 school districts in Tennessee. During the first year of the study, within each school, Kindergarten students were randomly assigned to classrooms in one of three treatment conditions: smaller classes (with 13 to 17 students), larger classes (with 22 to 26 students) or larger classes with a full-time classroom aide. Teachers were also randomly assigned to classes of different types. The assignments of students to different classroom types were maintained through the third grade for the students who remained in the study. Some students entered the study in the first grade and subsequent grades, and were randomly assigned to classes at that time. Teachers at each subsequent grade level were also randomly assigned to classes as the experimental cohort passed through the grades. Districts had to agree to participate for four years, allow site visitations for verification of class sizes, interviewing, and data collection, including extra student testing. They also had to allow the research staff to randomly assign pupils and teachers to class types and to maintain the assignment of students to class types from Kindergarten through grade 3.

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

Previous Findings from Project STAR

Early analyses of Project STAR data indicated that small size classes had positive effects on student achievement (Finn & Achilles, 1990). More recent analyses that considered validity threats (e.g., attrition, switching) also demonstrated that small size classes increase student achievement (Krueger, 1999; Nye et al., 2000). Other analyses have shown long-term positive effects of class size reduction on student performance (Finn, et. al., 2001; Krueger & Whitmore, 2001; Nye et al., 1999).

Project STAR data have also been used to examine the differential effects of class size on the achievement of low achieving, minority, and disadvantaged students. An early study reported some evidence that class size reduction has larger positive effects for minority students (Finn & Achilles, 1990). These average differences were significant for reading achievement for the first two years of the experiment. However, more recent studies that used modern and more appropriate statistical methods could not fully replicate the early findings. For example, Nye et al. (2001) found weak evidence that class size reduction has larger benefits for minority students. The gain was only observed in reading in one of the model specifications that the researchers examined. The differential effects of small size classes for disadvantaged students were statistically insignificant in all specifications. In a subsequent study Nye et al., (2002) examined the differential effects of small size classes for low achieving students and found no evidence of additional benefits for low achieving students. However, that study used samples of students who participated in Project STAR for two consecutive years, and thus did not include the new participants who joined the study the following year. Finally, a more recent study that used follow-up data from Project STAR indicated that being in small classes for four years may decrease the race/ethnic achievement gap in reading in subsequent grades 4 to 8 (Nye et al., 2004).

Nonetheless, overall there is weak evidence of differential effects of small size classes for low-achieving, minority, and disadvantaged students.

The present study examines how class size reduction impacts the achievement gap between lower and higher achieving students. To determine whether small size classes have differential effects on different types of students we compute differences in achievement variability between small and regular size classes, and differences in achievement in the upper and lower tails of the achievement distribution. Given the previous findings about the differential effects of small size classes one would expect that differences in achievement variability between smaller and larger classes should be small and insignificant and that the small class effect for lower achieving students would not be as important. This is actually the null hypothesis, which states that the variability in achievement between smaller and larger classes is zero and that all students benefit equally from being in small classes. However, given the results from studies that have evaluated educational interventions one would expect that class size reduction may increase the mean and the variability in student achievement. This would indicate that small size classes have higher variability in achievement than regular size classes and that high achieving students may benefit even more from being in small size classes than other students. If that were the case, then class size reduction would not reduce the achievement gap.

The Validity of Project STAR

In the STAR experiment, as in all longitudinal large-scale studies, the fidelity of implementation was somewhat compromised by three factors. First, there was some switching of students among class types in grades 1, 2, and 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 because of higher 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 these threats did not affect the outcome of the experiment in mean differences in achievement (see Krueger, 1999; Nye et al., 2000).

To ensure the validity of the experiment, it is also crucial that random assignment effectively eliminated preexisting differences between students and teachers assigned to different classrooms. First, the fact that the randomization of students and teachers to classes was carried out by the consortium of researchers who carried out the experiment, and not by school personnel, enhances its credibility. Second, the effectiveness of the randomization was examined by two recent studies that reported no differences on pre-existing characteristics of students or teachers among the assigned conditions (Krueger, 1999; Nye, Konstantopoulos, & Hedges, 2004). These results are consistent with what one would expect if randomization were successful. Note that these findings cannot prove that the groups did not differ in unobserved variables. However, confirming that differences in variables that were measured were not observed makes the probability that there are differences in unobserved variables smaller.

In randomized experiments such as Project STAR participants (e.g., students, 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 present study. That is, when randomization is successful, differences in central tendency and variability in achievement are entirely due to the treatment effect. Hence, the causal argument for differences in average achievement holds also for differences in achievement variability. In Project STAR this suggests that the only source of variance heterogeneity in achievement between smaller and larger size classes is the differential effect of the treatment (see Raudenbush & Bryk, 1987).

Statistical Analysis

The first part of the analysis involves differences in achievement variability, and hence the outcome variable is the variability in student achievement in each classroom. We computed variation in achievement for each classroom in the sample following the methods provided by Raudenbush and Bryk (1987). The first step involved the computation of the within classroom residuals. Since students are nested within classrooms we use a two-level model to compute the student level residuals. Specifically, the first level model for student i in classroom j is

$$Y_{ij} = \beta_{0j} + \beta_{1j} FEMALE_{ij} + \beta_{2j} MINORITY_{ij} + \beta_{3j} LOWSES_{ij} + e_{ij},$$

where Y represents mathematics or reading achievement for student i in classroom j, FEMALE is a dummy variable for gender, SES is a dummy variable for free or reduced price lunch eligibility, MINORITY is a dummy variable for minority group membership (indicating that the student was Black, Hispanic, or Asian), and e is a student and classroom-specific residual. The idea is to adjust for student characteristics in order to compute the residual variation in achievement in each classroom net of student effects. The second level model for the classroom specific intercept is

$$\beta_{0j} = \gamma_{00} + \eta_j ,$$

where γ_{00} is the average student achievement across all classrooms and η_j is a classroom-specific random effect. The remaining level one coefficients were treated as fixed at the second level. According to Raudenbush and Bryk (1987), the computation of achievement variability within each classroom involves the level-one residuals in each classroom and the degrees of freedom involved in the computation of the achievement variability in each classroom. Namely,

$$\hat{\theta}_j = \left(\sum e_{ij}^2\right) / v_j ,$$

where θ is the residual achievement variation in class j, e represents the student specific residuals in class j, and v_j indicates the degrees of freedom with which θ is estimated. In our case

$$v_j = n_j - 1 ,$$

where n is the number of students in classroom j. Further, Raudenbush and Bryk (1987) recommend the log transformation of θ and provide an unbiased estimator namely

$$d_j = 1/2(\log(\hat{\theta}_j) + v_j^{-1}) .$$

The term d_j represents now the residual variability in achievement in classroom j and has a known variance $v_j^{-1}/2$. Since the variance of each classroom specific outcome is known and these variances differ for different classrooms (heterogeneity of variance), the most appropriate method for analyzing these data is meta-analysis (Konstantopoulos & Hedges, 2004; Raudenbush & Bryk, 2002). Specifically, we ran a two-level (mixed effects) meta-analytic model which is expressed in a single level equation as

$$d_j = \delta_0 + \delta_1 SMALLCLASS_j + u_j + e_j ,$$

where δ_0 is the average variability in achievement across all classrooms, δ_1 is the average difference in achievement variability between small and regular classes that needs to be computed, SMALLCLASS is a dichotomous variable (1 if small class and 0 otherwise) that represents random assignment, u_j is a classroom specific random effect and e is the usual error term. The most important coefficient is δ_1 which represents the average difference in achievement variability between small and regular size classes. We conducted analyses for mathematics and reading achievement separately for each grade, that is, the analyses were repeated eight times. We also ran models adjusting for possible school effects, since students and teachers were not randomly assigned to schools. All analyses were repeated using fixed effects models for meta-analysis also. Our sample consisted of 325 classrooms in Kindergarten, 337 classrooms in the first grade, 324 classrooms in the second grade, and 326 classrooms in the third grade.

The second part of the analysis involved differences in achievement between students in small and in regular size classes in the upper and lower tails of the achievement distribution. Specifically, we used quantile regression to estimate the small class effect at various points of the achievement distribution (see Bushinsky, 1998; Koenker and Bassett, 1978). We ran quantile regressions for mathematics and reading test scores separately for each grade (grades K, 1, 2, and 3). In each grade mathematics and reading achievement scores were regressed on small class assignment (taking the value of one if a student is in small class and zero otherwise). Gender, race/ethnicity, and lower SES were included as covariates. We examined the small class effect at the lower tail (e.g., 10^{th} and 25^{th}), the middle (50^{th} quantile), and the upper tail (e.g., 75^{th} and top 90^{th}) of the achievement distribution.

Results

Do Small Classes Affect Achievement Variability?

The results reported here make use of the Stanford Achievement Test (SAT) reading and mathematics test scores collected from Kindergarten through grade 3 as part of Project STAR. First we computed the adjusted for student effects classroom achievement variability and then we regressed this variability on the small class binary variable to examine the small class effects. The results of this analysis are reported in Table 1. The left panel of Table 1 reports the results for mathematics and the right panel for reading achievement. The first column in each panel reports the regression coefficients, which are average differences in achievement variability between small and regular size classes. The second column in each panel reports the standard errors of these estimates, and column three reports the p-values for two-tailed t-tests. Within each grade the second row reports mean differences in achievement variability adjusted for school fixed effects.

Insert Table 1 Here

Although all 16 regression estimates are positive which indicates that achievement variability in small size classes is larger than that in regular size classes, only four estimates are significantly different from zero. This suggests that only 25 percent of the mean differences in achievement variability between small and regular size classes are statistically significant. In Kindergarten the differences in classroom achievement variability are significant at the 0.05 level in mathematics. In subsequent grades the differences in mathematics classroom achievement variability are not significant. In addition, the magnitude of the coefficients is smaller in grades 1, 2 and 3. This suggests that in mathematics, class size differences in variability are observed in the first year of the study.

The results in reading achievement are comparable. Specifically the coefficients in Kindergarten and first grade are larger than those in grades 2 and 3. In grade 1 the differences in classroom achievement variability are statistically significant at the 0.05 level in reading. In other grades the differences in reading achievement variability are not significant. This suggests that in reading, class size differences in variability are observed in the second year of the study. Thus, it appears that class size differences in achievement variability are observed for both mathematics and reading mainly during the first two years of the study (Kindergarten and first grade).

Intention to Treat Analysis

As in any large-scale long-term experiment the actual implementation of Project STAR deviated from the experimental design. One of the limitations was that in grades 1, 2, and 3 students who were initially assigned 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. From previous work we know that about 50 percent of the students assigned to either type of regular size classes in Kindergarten were reassigned to the other regular size class in the first grade (Krueger, 1999; Nye et al., 2000). With the exception of student switching between regular and regular size with aide classes in first grade, the non-random transition rates of students among treatment conditions ranged from two to nine percent across grades (see Nye et al., 2000). It is noteworthy that the transition rates from regular to small size classes were consistently eight to nine percent between grades, while the transition rates from small to regular size classes were much smaller (two to four percent).

Because the student transitions among types of classes were non-random it is possible that the estimates of the class size effects are biased. Previous work that examined mean differences in achievement between small and regular size classes found no evidence of bias (Krueger, 1999; Nye et al., 2000). In the present study we examine whether student switching among different types of classes affects differences in achievement variability between small and regular size classes. One way to examine the possible effects of switching among types of classrooms is to estimate the 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 that is typically used in clinical trials. Suppose a student is assigned to a regular size class in Kindergarten and switches to a small size class in the first grade. In the intention to treat analysis this student is assumed to be part of the regular size class in the first grade even though 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 qualitatively 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. For each grade (1, 2, and 3) we constructed a new variable that we call "original" assignment as a dichotomous variable taking the value of one if a student was originally assigned to a small size class and zero otherwise. Then, we reran the analysis discussed earlier in the statistical analysis section for mathematics and reading for grades 1, 2, and 3.

The results of this analysis are reported in Table 2. The structure of Table 2 is identical to that in Table 1. The results of the intention to treat analysis are qualitatively similar to and consistent with those reported in Table 1. Fifteen out of 16 regression estimates (about 94 percent) are positive, but only four of the estimates are statistically significant. In Kindergarten the differences in classroom variability are statistically significant at the 0.05 level in mathematics. As in Table 1 the magnitude of the coefficients is smaller in grades 1, 2 and 3. The results for reading are comparable. The coefficients in Kindergarten and first grade are larger than those in grades 2 and 3. In first grade the differences in classroom variability are statistically significant at the 0.05 level in reading. In other grades the differences in reading achievement variability are not significant. Overall these results also indicate that class size differences in achievement variability are observed for both mathematics and reading mainly during the first two years of the study (Kindergarten and first grade).

Insert Table 2 Here

The Association between Actual Class Size and Achievement Variability

Even though the experimental design had targeted a certain range of class size for each type of classroom (13 to 17 for smaller classes and 22 to 26 for larger classes), there was more than intended variation in small and regular size classes. That is, the actual class size ranged from 11-20 for small size classes and from 15-29 for regular size classes (see Table 3). As Table 3 shows there was a modest amount of 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 size classes may have affected the estimate of the treatment effect. Hence, a more complete analysis would examine the association between actual class size the classroom variability in achievement.

Insert Table 3 Here

In order to conduct this analysis one needs to construct actual class size and include it as the main independent variable in the meta-analysis regression. This approach however has the disadvantage that while target class is randomly assigned, actual class size is not and may be a result of non-random unobserved factors that may also be related to the outcome. That is, any relation between actual class size and achievement variability is not necessarily 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, Imbens, and Rubin, 1996; Nye et al., 2004). In the IV regression actual class size is regressed on random assignment and the predicted values of this regression are used in the meta-analysis 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 4. The structure of Table 4 is identical to that in Tables 1 and 2 and the results are qualitatively similar to those reported in Tables 1 and 2. Specifically, although all 16 regression estimates are negative as expected, only four estimates are significantly different from zero. In Kindergarten the differences in classroom variability are statistically significant at the 0.05 level in mathematics. Again, the magnitude of the coefficients is smaller in grades 1, 2 and 3. The results for reading are comparable. The coefficients in Kindergarten and first grade are larger than those in grades 2 and 3. In first grade the association between class size and classroom achievement variability is statistically significant at the 0.05 level in reading. In other grades the coefficients are not significant. Overall these results also indicate that class size differences in achievement variability are observed for both mathematics and reading mainly during the first two years of the study (Kindergarten and first grade). It should be noted that all analyses were replicated using also fixed effects models for meta-analysis and the results were qualitatively similar to those obtained using mixed effects models.

Insert Table 4 Here

Assessing Achievement Differences in the Upper and Lower Tails

The results from the previous analyses suggest that the small class effect may not be uniformly distributed across the achievement distribution. In fact, the results of the previous analyses indicate that the small class advantage may be larger in the upper tail of the achievement distribution. If high achieving students benefit more from being in small classes, then the small class advantage at the upper tail must be larger than that in the middle or in the lower tail of the achievement distribution.

The results of the quantile regression analyses are summarized in Table 5. All estimates are in standard deviation units. As expected all coefficients in the median (or robust) regression are positive, significantly different from zero, and range between 1/7 and 1/4 of a SD. These results are qualitatively similar to those reported in previous studies that estimated mean differences (see Nye et al., 2000). The estimates at the lower tail are also positive but smaller than the estimates in the median or in the upper tail of the achievement distribution. Nonetheless, 50 percent of the estimates at the 10^{th} quantile (grades 1 and 3) and all estimates at the 25^{th} quantile are statistically significant. This indicates that lower achieving students benefit from being in small classes. In the upper tail all estimates (75th and 90th quantile) are positive and statistically significant. The magnitude of the coefficients indicates that the small class effect is consistently larger for high achievers than for other students. All coefficients estimated in the upper tail of the achievement distribution are much larger than those obtained from the middle or the lower tail. In Kindergarten mathematics the coefficient at the 90th quantile is more than twice as large as the coefficient at the 50th quantile and nearly four times as large as the coefficient at the 10th quantile. In Kindergarten reading the difference in achievement between the median and the 90th quantile was much smaller (17 percent), but the difference between the 10th and the 90th quantile estimates was still large. In the first grade, the difference in achievement between the median and the 90th quantile was nearly 25 percent both in mathematics and reading achievement.

Insert Table 5 Here

Although these results seem to provide support of the notion that higher achieving students may benefit more from being in small size classes than other students in the same types of classes, one needs to examine whether the estimates across the different quantiles are statistically significant. Table 6 summarizes t-tests that examine whether the differences between quantile regression estimates across different quantiles are statistically significant. The first column of Table 6 indicates that in grades K and 2 the differences between the small class effect at the 10th and the 90th quantiles are statistically significant at the 0.05 level in mathematics. Also, in grade 1 the differences between the small class effects at the 10th and the 90th quantiles are statistically significant at the 0.05 level in reading. This indicates that in some grades the very high achievers benefit significantly more from being in small size classes than very low achievers. This finding partly replicates that from the previous analyses that pointed to significant differences in achievement variability in grade K in mathematics and grade 1 in reading. The results in column 2 indicate that 50 percent of the differences between the estimates at the 25th and the 75th quantiles are statistically significant. The results in columns 3 and 4 indicate that some of the differences between the estimates at the 50th and the 10th or 90th quantiles are statistically significant. Overall, these results provide some evidence that higher achievers benefit more from being in smaller classes than other students in the same types of classes. However, these results do not indicate that lower achieving students are better off in regular classes, that is, all types of students benefit from being in small size classes.

Insert Table 6 Here

Conclusion

Previous work that used Project STAR data provided consistent evidence that being in small size 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 think of the possibility that it could also reduce the achievement gap between higher and lower achieving students. However, previous research provided weak or no evidence that reducing class size benefits lower achieving students more than other students. The present study examined differences in achievement variability between smaller and larger size classes and differences in the upper and lower tails of the achievement distribution in an attempt to better understand the effects of class size reduction on the achievement gap.

The results reported here suggest that small size classes produce significantly higher variability in student achievement than regular size classes in Kindergarten in mathematics and in first grade in reading. The differences favoring small size classes were more pronounced and significant in the first two years of the experiment (Kindergarten and first grade) and smaller and insignificant in the last two years of the experiment (grades 2 and 3). Overall the results indicate that class size reduction does not only increase achievement for all students on average, but the variability in student achievement as well (at least in the first two grades). In addition, the results from the quantile regression analyses provided additional evidence that all types of students benefit form being in small classes, while high-achieving students may benefit even more.

These findings suggest differential effects of small classes across different types of students, that is, some types of students benefit more from being in small classes than other types of students. Specifically, due to the larger variability in achievement in small classes, the difference (or distance) in achievement between high and low achievers is greater in those classes than in regular size classes in grades K and 1. If the achievement distributions in small and regular size classes had the same mean but different variances (e.g., larger variances in small classes), then one would argue that high achievers may benefit more from being in small classes than in regular size classes while low achievers may benefit less. However, the achievement distributions differ in the means as well since smaller classes have a higher mean. This still indicates that higher achievers may benefit more from small classes, but given the considerable average difference in achievement (nearly 0.2 SD) low achievers would benefit at least as much as low achievers in regular size classes since the small class achievement distribution is shifted to the right. The results of the quantile regression analysis supported this notion. The results showed that high achievers benefit even more from being in small classes, but low achievers benefit as well. Nonetheless, the achievement gap between lower and higher achieving students is still larger in small classes than in regular classes in some grades. This suggests no evidence that manipulating class size can reduce the achievement gap between lower and higher achieving students.

Our analyses also addressed the possible effects of validity threats such as student switching between types of classrooms, and larger variability than intended by design in actual class size which resulted to some overlap in actual class size between smaller and larger size classes. The results of these analyses were consistent with those in the original analysis and further supported the notion that achievement variability is larger in small classes especially during Kindergarten and first grade. This again suggests that high achieving students may benefit even more from being small classes than in other types of classes at least in Kindergarten and first grade. However, we did not find any evidence of additional benefits of small classes for lower achieving students. This result should be interpreted with caution. It does not necessarily mean that lower achieving students are better off in larger size classes, since all students benefit from being in small classes.

These results are interesting since they shed some more light on the mechanism through which small classes may benefit students. One hypothesis is that in small classes teachers are more likely to identify lower achieving students and hence they are more likely to provide instructional strategies especially designed to benefit these students in the early grades. However, our findings did not support this hypothesis. Another hypothesis is that teachers are also more likely to identify higher achieving students in small classes and hence they are more likely to provide effective strategies that benefit high achieving students more. Alternatively, it is plausible that the instructional practices in small classes benefit higher achievers more. That is, high achieving students may be more engaged (or motivated) in learning in small classes than other students. Possibly high achieving students take more advantage of the opportunities or practices that take place in small classes or create more opportunities for learning in small classes than lower achieving students in the same classes or other students in other classes especially the first two years of school (Kindergarten and grade 1). One possibility is that the effects of small classes accrued mainly in the first and second year of the study. Some previous work has discussed that possibility and showed that the cumulative effects of small classes diminish over time in mathematics. Typically in the first two years of schooling students learn what behaviors are expected in school. This means that teachers spent a considerable amount of time on management and behavior related issues. It is likely that in smaller classes with fewer students these issues are addressed in a shorter time than in regular size classes, and this in turn means that in Kindergarten and first grade more time is spent on learning and instruction in small classes. It is possible that high achievers in small classes take advantage of this and engage more in learning than other students. Hence they may have steeper learning trajectories in the first two years of school (Kindergarten and grade 1) than other students or in later grades. By grade two students typically know what is expected in elementary school and hence it is likely that the time spent on management issues and learning and instruction is comparable in small and regular size classes. It is difficult to know exactly what the mechanism might be. Our results however, indicate that higher achieving students benefit even more from being in small classes mainly in Kindergarten and first grade.

Although this study helped us better understand the effects of small classes on student achievement, the mechanism is still not clearly defined. Unfortunately, data about classroom practices in different types of classrooms are not available. Such detailed observational data could have unveiled the mechanism of small class effects via information about instructional processes and interactions between students and teachers. A new randomized experiment with the objective of collecting high quality observational data in the classrooms would provide invaluable information about the small class effects.

References

Angrist, J., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91,* 444-472.

Bloom, B. S. (1984). The two sigma problem: The search fro methods of group-instruction as effective as one-to-one tutoring. *Educational Researcher, 13(6),* 4-16.

Bowles, S., & Levin, H. M. (1968). The determinants of scholastic achievement: An appraisal of some recent evidence. *Journal of Human Resources, 3,* 3-24.

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. Psychological Bulletin*, 104,* 396-404.

Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of Human Resources, 33,* 89-126.

Coleman, J. S. (1985). *Schools, families, and children* (Nora and Edward Ryerson Lecture). Chicago: University of Chicago.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods.* New York: Irvington Publishers.

Finn, J D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal,27,* 557-577.

Finn, J D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record,103,* 145-183.

Glass, G. V., Cahen, L.S., Smith, M. E., & Filby, N. N. (1982). *School class size: research and policy*. Beverly Hills: Sage.

Glass, G. V., & Smith, M. E. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1,* 2-16.

Greenwald, R., Hedges, L. V., & Lane, R. D. (1996). The effects of school resources on student achievement, *Review of Educational Research, 66*, 361-396.

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature 24*, 1141-77.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18,* 45-51.

Hedges, L. V., Laine, R., & Greenwald, R. (1994). Does Money Matter: A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. Educational Researcher, 23, 5-14.

Hedges, L. V., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal, 20*, 63-85.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica, 46,* 33-50.

Konstantopoulos, S., & Hedges, L.V. (2004). Meta-Analysis. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 281 -297). New York: Sage.

Konstantopoulos, S. (2005). *Asian-White differences in achievement: A quantile regression approach.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, CA.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114,* 497-532.

Krueger, A. B., & Whitmore, D. M. (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,* 1-28.

Monk, D. H. (1981). Toward a multilevel perspective on the allocation of Educational Resources. *Review of Educational Research, 51,* 215-236.

Mosteller, F., Light, R. J., Sachs, J. A. (1983). Sustained inquiry in education: Lessons learned from skill grouping and class size. *Harvard Educational Review, 66,* 797-842.

Nye, B., Hedges, L.V., & Konstantopoulos, S. (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*, 127-142.

Nye, B., Hedges, L.V., & Konstantopoulos, S. (2000). Effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37,* 123-151.

Nye, B., Hedges, L.V., & Konstantopoulos, S. (2000). Do the disadvantaged benefit from the small classes? Evidence from the Tennessee class size experiment. *American Journal of Education, 109,* 1-26.

Nye, B., Hedges, L.V., & Konstantopoulos, S. (2002). Do low achieving students benefit more from small classes? Evidence from the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 24*, 201-217.

Nye, B., Hedges, L.V., & Konstantopoulos, S. (2004). Do minorities experience larger lasting benefits from small classes? Evidence from a five-year follow -up of theTennessee class size experiment. *Journal of Educational Research, 98,* 94-100.

Nye, B., Konstantopoulos, S, & Hedges, L.V. (2004). How Large are Teacher Effects? *Educational Evaluation and Policy Analysis, 26,* 237-257.

Raudenbush, S. W., & Bryk, A. S. (1987). Examining correlates of diversity. *Journal of Educational Statistics, 12,* 241-269.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models*. Thousand Oaks, CA: Sage Publications.

Table 1. Mean Differences in Classroom Variability Between Small and Regular Classes for Mathematics and Reading

	Mathematics			Reading		
	Grade K			Grade K		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Small Class	0.069	0.029	0.019*	0.059	0.041	0.151
Including School Fixed Effects	0.069	0.028	0.011*	0.058	0.037	0.113
	Grade 1			Grade 1		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Small Class	0.033	0.028	0.236	0.069	0.029	0.017*
Including School Fixed Effects	0.033	0.024	0.168	0.061	0.022	0.006*
	Grade 2			Grade 2		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Small Class	0.025	0.030	0.397	0.012	0.030	0.700
Including School Fixed Effects	0.035	0.027	0.186	0.007	0.025	0.784
	Grade 3			Grade 3		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Small Class	0.011	0.026	0.662	0.032	0.025	0.199
Including School Fixed Effects	0.006	0.024	0.799	0.016	0.024	0.514

* $p < 0.05$

Table 2. Mean Differences in Classroom Variability Between Small and Regular Classes for Mathematics and Reading: Intention to Treat Analysis

	Mathematics			Reading		
	Grade K			Grade K		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Small Class	0.069	0.029	0.019*	0.059	0.041	0.151
Including School Fixed Effects	0.069	0.028	0.011*	0.058	0.037	0.113
	Grade 1			Grade 1		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Small Class	0.038	0.028	0.178	0.067	0.029	0.024*
Including School Fixed Effects	0.037	0.025	0.136	0.063	0.022	0.005*
	Grade 2			Grade 2		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Small Class	0.021	0.031	0.492	0.001	0.031	0.973
Including School Fixed Effects	0.024	0.028	0.395	-0.001	0.026	0.959
	Grade 3			Grade 3		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Small Class	0.011	0.027	0.687	0.034	0.026	0.194
Including School Fixed Effects	0.003	0.026	0.905	0.025	0.025	0.320

* $p < 0.05$

Table 3. Distribution of Actual Class Size Among Types of Classes and Grades

	Grade K			Grade 1			Grade 2			Grade 3		
	Classroom Type			Classroom Type			Classroom Type			Classroom Type		
Class Size	Small	Regular	Regular/Aide	Small	Regular	Regular/Aide	Small	Regular	Regular/Aide	Small	Regular	Regular/Aide
11										2		
12	8			2			3			2		
13	19			14			16			15		
14	22			18			27			17		
15	23		1	31			32			31		
16	31	1		16	1		29	1		31		1
17	24	4	1	33	1		19			27		
18		1	2	6	2		6			10	1	
19		7	6	3	4	3	1	3	3	5		4
20		6	6	1	10	6		2	1		9	3
21		14	12		18	18		7	11		11	12
22		20	20		27	15		23	21		13	16
23		16	21		19	20		20	21		10	14
24		19	14		16	11		22	25		15	14
25		6	6		7	9		9	15		16	15
26		4	3		5	9		6	7		5	12
27		1	6		2	4		4	1		5	8
28			1		1	2		1			2	6
29					1	2		2	2		2	2
Total	127	99	99	124	114	99	133	100	107	140	89	107
Average	14.96	22.16	22.54	15.52	22.47	23.20	15.16	23.29	23.32	15.53	23.42	23.77

Table 4. Effects of Actual Class Size on Classroom Variability for Mathematics and Reading

	Mathematics			Reading		
	Grade K			Grade K		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Class Size	-0.009	0.004	0.017*	-0.008	0.006	0.160
Including School Fixed Effects	-0.009	0.004	0.010*	-0.008	0.005	0.125
	Grade 1			Grade 1		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Class Size	-0.004	0.004	0.274	-0.009	0.004	0.021*
Including School Fixed Effects	-0.004	0.003	0.190	-0.008	0.003	0.006*
	Grade 2			Grade 2		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Class Size	-0.003	0.004	0.404	-0.001	0.004	0.708
Including School Fixed Effects	-0.004	0.003	0.193	-0.0008	0.003	0.795
	Grade 3			Grade 3		
	Coefficient	SE	Two-Tailed P-Value	Coefficient	SE	Two-Tailed P-Value
Class Size	-0.001	0.003	0.683	-0.004	0.003	0.220
Including School Fixed Effects	-0.0007	0.003	0.817	-0.002	0.003	0.555

* $p < 0.05$

Table 5. Achievement Differences Between Small and Regular Classes at Various Quantiles for Mathematics and Reading

	Mathematics				
Grade	10th Quantile	25th Quantile	50th Quantile	75th Quantile	90th Quantile
K	0.084	0.105*	0.147*	0.178*	0.356*
1	0.162*	0.209*	0.255*	0.302*	0.325*
2	0.090	0.112*	0.157*	0.236*	0.247*
3	0.100*	0.126*	0.138*	0.151*	0.201*

	Reading				
Grade	10th Quantile	25th Quantile	50th Quantile	75th Quantile	90th Quantile
K	0.063	0.158*	0.189*	0.252*	0.221*
1	0.145*	0.145*	0.236*	0.290*	0.299*
2	0.152*	0.152*	0.152*	0.217*	0.261*
3	0.078	0.104*	0.207*	0.182*	0.207*

* $p < 0.05$

Table 6. T-tests Indicating Differences in Quantile Regression Estimates

	Mathematics			
Grade	10th Vs 90th Quantile	25th Vs 75th Quantile	10th Vs 50th Quantile	90th Vs 50th Quantile
K	2.854*	1.105	1.038	2.313*
1	1.934	2.151*	2.541*	0.913
2	2.169*	2.687*	1.159	1.332
3	1.465	0.688	0.639	1.161

	Reading			
Grade	10th Vs 90th Quantile	25th Vs 75th Quantile	10th Vs 50th Quantile	90th Vs 50th Quantile
K	1.468	2.243*	2.737*	0.368
1	2.651*	3.452*	1.801	1.355
2	1.742	1.264	0	2.270*
3	1.835	1.465	2.638*	0

* $p < 0.05$