Instructions for authors, subscriptions and further details:

http://revistashipatia.com/index.php/rise/index

Modeling Class Size Effects Across the Achievement Distribution

Spyros Konstantopoulos
Wei Li

Michigan State University

Date of publication: February 25th, 2012.

To cite this article: Konstantopoulos, S., Li, W. (2012). Modeling Class Size Effects Across the Achievement Distribution. *International Journal of Sociology of Education 1*(1), 5-26. doi: 10.4471/rise.2012.01

To link this article: http://dx.doi.org/10.4471/rise.2012.01

RISE - International Journal of Sociology of Education, 1(1), *5-26*

Modeling Class Size Effects Across the Achievement Distribution

Spyros Konstantopoulos
Wei Li
Michigan State University

Abstract

Previous findings from Project STAR have highlighted the benefits of being in small classes in early grades on average. Here, we examined the differential "value-added" effects of small classes across the achievement distribution. We find that once previous grade achievement and small class membership are controlled for, small class effects are by and large insignificant. Although high-achievers benefited more from small classes in third grade reading, overall, the differential small class effects were not systematic. Findings from longitudinal analyses failed to show that being in small classes for two or more years is as beneficial as being in small classes for only one year. Although the cumulative small class effects are positive, significant, and meaningful in magnitude, they do not indicate which grade (or grades) are the most important.

Keywords: Small classes, student achievement, quantile regression

2011 Hipatia Press
ISSN 2014-3575
DOI: 10.4471/rise.2012.01

A fundamental objective of any education system is to allocate school resources most effectively in order to increase student performance. School policies are often designed to ensure the the best possible distribution of school resources that will result in higher levels of academic achievement for all students. Class size reduction has been identified by some researchers such as Glass and Smith (1979) and Finn and Achilles (1990) as a school mechanism that can increase student achievement. Class size reduction has been considered as an appealing school intervention in the U.S. because given adequate resources it is not difficult to implement. Specifically, if school buildings are well-equipped with adequate numbers of classrooms and qualified teachers class size reduction is easy to put into practice. A notable example of a study that was conducted to measure class size effects in the U.S. was the Tennessee class size experiment, where small classes had on average 15-16 students and regular size classes had on average 22-23 students.

Studies that have used high-quality experimental data from the Tennessee class size experiment have consistently demonstrated the positive effects of small classes on average student achievement for all students (e.g., Finn & Achilles, 1990; Krueger, 1999; Nye, Hedges, & Konstantopoulos, 2000a). Specifically, these studies indicated that the average student achievement in small classes was significantly higher than in regular classes. On average the class size difference between small and regular classes was seven students. The results of these studies suggested that reducing class size is a promising intervention that increases academic achievement on average for all students. As a result, in the U.S. some states have introduced class-size reduction programs. California, for example, introduced a class size reduction program that provided financial incentives to schools that reduce class size in the early grades to twenty or fewer students per classroom. Wisconsin adopted a program that reduced class size to fifteen students per classroom in early grades in schools with high percentages of students from disadvantaged backgrounds. ages of students from disadvantaged backgrounds.

Thus far, researchers have focused typically on the average effect of class size on student performance. Although the main interest in empirical studies that evaluate treatment effects lies in computing an average treatment effect, it is equally important to estimate class size effects across the distribution of achievement. For instance, it is unclear that class size reduction will have the same effects on the performance of low-, medium-, and high-achievers. Class size may interact with level of achievement to produce differential effects for different groups of students. The underlying hypothesis is that class size can affect student achievement on average, but can also affect the performance of low- and high-achievers differently. This is an important issue in the U.S. education system since school interventions are designed to increase student achievement for all, but at the same time to close the achievement gap between low- and high-achievers by boosting the performance of low-achievers (see Konstantopoulos & Chung, 2009).

Moreover, although researchers have investigated the effects of class size on student achievement cross-sectionally using scores from the end of the year assessments (e.g., Finn & Achilles, 1990; Krueger, 1999), the "value-added" effects of small classes have not been addressed adequately. In particular, cohort studies such as the Tennessee class size experiment allow researchers to probe whether class size effects are additive over time. That is, one can determine whether being in a small class in one grade has an effect on student achievement at the end of that grade given class size assignment and achievement in the previous grade. In this study we asked the question what are the "value-added" effects of class size reduction once prior class size assignment and achievement are taken into account? This question translates to whether there are additive effects of class size on achievement as students go through early grades. Some previous work has suggested that perhaps reducing class size for only one year is adequate to capture the small class effects and that additional years in small classes may not have additional meaningful effects (see Krueger, 1999). Under this hypothesis, the first grade small class effect on first grade achievement may be minimized for example, when small class membership and achievement in kindergarten are taken into account.

In the present study we examined the differential "value-added" effects of class size on student achievement across the distribution of achievement. Specifically, we investigated small class effects on low-, medium-, and high-achievers in one grade (e.g., first grade) controlling for achievement and small class membership in the previous grade (e.g., kindergarten). We hypothesized that the small class effect would be additive over time and that in each grade small class membership has independent effects on student achievement. We were particularly interested in the independent effects of small classes on low-achievers as students progressed through early grades (kindergarten through third grade). Since the 1960s an important goal of the U.S. education system has been to determine how school resources reduce the achievement gap between high- and low-achievers, who are over-represented by minority and economically disadvantaged students (see Coleman et al., 1966). To address this goal careful modeling that equates students' differences in background (e.g., prior achievement, small class membership) is necessary We used data from a 4-year, large-scale, randomized experiment, designed to gauge the effects of small classes on student achievement in Tennessee, mainly known as Project STAR (Student Teacher Achievement Ratio). Project STAR was a cohort study where a group of kindergarteners were followed for four years through third grade. These data are appropriate to use since they provide class size membership and achievement information in each grade.

Another objective of the study was to determine whether being in small classes in different grades had independent effects on achievement. That is, we also examined the effects of small classes on student achievement from kindergarten through third grade simultaneously. We asked the question, does being in small classes in kindergarten or first grade affect third grade achievement when second and third grade small class membership were taken into account? In other words, we were interested in identifying whether small class membership in more than one grade makes a difference and whether small class membership in later grades is more beneficial than in earlier grades. To address this objective we modeled third grade achievement as a function of student characteristics in third grade as well as small class indexes from kindergarten through third grade. Finally, a third objective of the study was to measure the effect of being in small classes for multiple years on third grade achievement of low-, medium-, and high-achievers. The hypothesis related with this objective was that students who were in small classes for four years would have higher achievement than students who were in regular size classes for four years. It is also likely that these effects would point to larger benefits for low-achievers, since as some re-

-searches have argued low-achievers may have more opportunities to learn in small classes (see Finn & Achilles, 1990).

Literature Review

The effects of class size reduction on student achievement have been examined empirically via various research designs and analyses over the past few decades. Numerous small-scale experimental and quasi-experimental studies have investigated the effects of class size on student achievement. Reviews of these studies have suggested that class-size reduction has positive effects on student achievement and that these effects become greater as class size becomes smaller (Glass, & Smith, 1979; Glass, Cahen, Smith, & Filby, 1982). The benefits of small classes appeared to be more pronounced in early elementary grades, in classes with less than twenty students, and among some minority and those of low socioeconomic status.

Previous studies have also examined the effects of class size reduction on student achievement using data from non-experimental studies. Typically, such studies compute the association between class size and student achievement, adjusting for important factors in student background such as gender, race and ethnicity, socioeconomic status, and previous achievement. The interpretation of these studies' results has been mixed; hence, this body of research has not yielded clear evidence about the effects of small classes. Some reviewers of this body of research have argued that the effects of class size on student achievement are small or nonexistent (Hanushek, 1989). However, others have suggested that reducing class size has positive effects on student achievement and that students may benefit from being in small classes (Greenwald, Hedges, & Laine, 1996). Still the findings about the effects of small classes do not seem to be systematic and at times they are open to different interpretation by different investigators.

Findings from primary studies during the last decade have also been mixed. For instance, Angrist and Lavy (1999) found that reducing class size increased fourth and fifth graders' scores significantly, and Pong and Pallas (2001) found positive small class effects on eighth-grade achievement. In contrast, Hoxby (2000) reported that smaller classes had little to no effect on student achievement, and Milesi and Gamoran (2006) found no evidence of class size effects on student achievement.

Project STAR data have also been used to examine the differential effects of class size on the achievement of low-achieving, minority, and economically disadvantaged students. An early study reported that class size reduction had 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, other studies could not fully replicate the early findings. For example, Nye, Hedges and Konstantopoulos (2000b) found weak evidence that class size reduction had 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 classes for disadvantaged students were statistically insignificant in all specifications.

Class size reduction could affect various levels of achievement differently. One possibility is that class size will impact the achievement of high, medium- and low-achieving students in similar ways. That is, small classes would have a uniform effect on student achievement in the middle and the tails of the distribution. Another possibility is that low-achievers benefit more from being in small classes than other students. In this case, small classes would interact with level of achievement and would help increase the performance of low-achievers. That would suggest differential effects of small classes. A third possibility would be that high-achievers benefit more from being in small classes than other students. In this case, small classes would also interact with level of achievement.

The effects of class size on low-achievers and the achievement gap has been discussed in the literature (e.g., Konstantopoulos, 2008). One study for example, examined the differential effects of small classes for students who were low-achievers in previous grades and found no evidence of additional benefits for these students (Nye Hedges, & Konstantopoulos, 2002). That study however, did not examine small class effects on low-achievers in the same grade and did not adequately control for class size membership in the previous grade. Another study indicated that being in small classes for four years may subsequently decrease the race/ethnic achievement gap in reading in grades 4 to 8 (Nye, Hedges, & Konstantopoulos, 2004). Although there is weak evidence of differential effects of small classes for low-achieving, minority, and economically disadvantaged students, previous work has not discussed the "value-added" class size effects across the achievement distribution. In addition, previous work has not discussed thoroughly the differences in class size effects on student achievement that are related with how the treatment was actually received/implemented or how it was intended to be received/implemented.

Method

Data

Project STAR was a four-year large-scale field experiment that included students in seventy-nine elementary schools in forty-two 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 (13 to 17 students), larger classes (22 to 26 students), or larger classes with a full-time teacher 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 different types of 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 three. Since the treatment was simply a function of class size and the regular size classes were not that large ethical considerations about the "negative" effects of the experiment on students should be minimized.

Overall, more than 11,000 students in 79 schools participated in the experiment over the four-year period. 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 convenient samples 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 over time.

Validity of Estimates in Project STAR

There were three potential threats to the validity of class size estimates in Project STAR. The first is related to whether random assignment of students to classes was successful. When random assignment is successful average differences in variables across treatment types are only due to chance and should not be systematic. Random assignment is a crucial aspect of the internal validity of Project STAR because it guarantees that preexisting differences between students assigned to different types of classrooms are virtually eliminated. The fact that random assignment of students to different types of classrooms was carried out by the consortium of researchers who carried out the experiment, enhances its credibility. Nonetheless, researchers have examined empirically whether random assignment was successful using observed characteristics such as gender, race, and SES (socioeconomics status), and have concluded that there was no evidence from observed characteristics that random assignment did not work (Krueger, 1999; Nye et al., 2000a). However, other researchers have had some concerns about random assignment (Hanushek, 1999; Konstantopoulos, 2011). For example, Konstantopoulos conducted within school analyses and found that for variables such as age and SES the observed significant differences were greater than 5 percent and in some grades greater than 10 percent. Thus, for these two variables the evidence is not so consistent with what one would expect had random assignment worked.

Attrition from grade to grade is also a potential threat for cohort studies such as Project STAR. Attrition can potentially affect the class size estimates if within small or regular size classes the students who drop out of the study are systematically different than those who remain in the study. This kind of differential attrition could introduce some selection bias in the estimates of class size if for example low-achievers drop out of small classes whereas high-achievers drop out of regular classes from year to year. Previous work has examined the effects of differential attrition on class size estimates with Project STAR data and has concluded that differential attrition did not seem to compromise the class size estimates (Krueger, 1999; Nye et al., 2000a). Although a more recent study showed that school attrition was related to school achievement and composition (e.g., minority or disadvantaged students) in some grades

(Konstantopoulos, 2011) , it is unclear that this potential selection can bias the treatment effects (i.e., small class effects). Thus, in the present study we will assume that Krueger's (1999) findings hold.

Variables of Interest

The dependent variables were the SAT-9 reading and mathematics test scores collected as part of Project STAR in kindergarten through third grade. SAT-9 is a widely used test that measures academic achievement of elementary and secondary school students in the U.S.. The main independent variable was class size. Specifically, regular size classes served as the reference group and small class was included in the model as a binary indicator (i.e., dummy). Student characteristics such as gender, race, and SES were also included in the models. All three variables were coded as binary indicators and modeled female, minority status, and low SES (took the value of 1 if student was eligible for free or reduced lunch) effects.

Statistical Analysis

The objective of the study is to examine the "value-added" effects of class size across the distribution of achievement, especially the effects in the upper and lower tails of the distribution. To that end, we used quantile regression, a technique that allows investigating small class effects at various points of the achievement distribution (Buchinsky 1998; Koenker and Bassett 1978). Frequently, education and social science researchers seek to determine the effects of school resources on low-achieving, minority, and disadvantaged students. For the purposes of this study, it is possible that class size effects are differential for average, lower, and higher-achieving students. The typical regression technique produces regression coefficients which are averages and, thus, it is inadequate to examine the effects of predictors at different points, called quantiles, of the achievement distribution. In contrast, quantile regression is appropriate because it helps researchers estimate the effects of interest at any point in the achievement distribution, not just the mean. (Hao & Naiman 2007).

Quantile regression is a natural extension of the typical linear regression because it estimates how predictors (e.g., class size) affect outcomes (e.g., achievement) not only in the middle, but in the tails of the outcome distribution as well. Hence, quantile regression estimates provide a more complete picture of the effects of predictors on the entire distribution of outcomes (Hao & Naiman 2007). Quantile regression is also 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 social sciences. We argue that this method can also be useful in education research that focuses on educational inequality and the academic prosperity of students especially in the lower tail of the achievement distribution. We believe that quantile regression serves the objective of our study well because it allows the estimation of differential effects of class size for various levels of achievement. In addition, covariate differential effects (e.g., race or SES) are also modeled across the achievement distribution. Finally, the same index (e.g., standard deviation units) can be computed for class size effects on achievement across the entire distribution, and hence, the results across different points (quantiles) of the achievement distribution are comparable.

We used three modeling strategies to examine small class effects on student achievement. First, we estimated small class effects using a treatment on the treated (TOT) approach. This approach estimates the effects of the treatment on student achievement as the treatment was actually received by the students. Such estimates are potentially biased however, since switching among class types may not have been random. The regression equation for grade g (g = 1, 2, or 3) in each quantile (i.e., tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth) is:

$$Y_g = \gamma_0 + \gamma_1 TOTSMALL_g + \gamma_2 FEMALE_g + \gamma_3 MINORITY_g + \gamma_4 LOWSES_g + \gamma_5 TOTSMALL_{g-1} + \gamma_6 Y_{g-1} + \varepsilon_g$$

where Yg indicates student achievement scores in grade g (e.g., first grade), g-1 indicates small class membership and achievement in the previous grade (e.g., kindergarten), FEMALE, MINORITY, and LOWSES are student characteristics and SMALL indicate type of classroom (TOT). The γ's are the regression coefficients that need to be estimated. The most important coefficient for our objective was $y1$ and represents the "value-added" of small classes on student achievement in one grade controlling for student characteristics and small class membership and achievement in the previous grade. The above model was used to model first, second, or third grade achievement (i.e., reading or mathematics scores).

The second approach we used was the intention to treat (ITT) approach. In Project STAR students switched among types of classes in the first, second, and third grades. That is, students who were assigned initially to a specific type of class in one year switched to other types of classes the next year. In particular, in the first grade only, students who were assigned to regular-size and regular-size-with-an-aide classes were re-randomized again to receive the other treatment condition (Krueger, 1999; Nye et al., 2000a). In the second and third grade no reassignments should have been made by design. In addition, switching to or from small classes was not part of the design. However, there was some transition of students to and from small classes. In first grade, nearly four percent of students who were in small classes in kindergarten and were present in first grade switched to regular classes. Another three percent of these students switched to regular classes with a full time aide. Nearly 15 percent of students who were in regular size classes in kindergarten moved to small classes in first grade. This pattern was repeated in the second the third grade, that is, a higher percentage of students (nearly 10 percent) who were in regular classes in one year moved to small classes the following year. The percentage of students who were in small classes in one year and moved to regular classes the following year was smaller in grades 2 and 3 (2 to 4 percent). If these transitions among classroom types were random or part of the study design the switching should not have biased the small class estimates.

The ITT analysis should provide unbiased estimates because it capitalizes on the principle of random assignment (see Friedman, 2006). With this approach the effects of class size reduction are modeled according to initial or original assignment. That is, treatment was modeled as it was originally assigned to a student the first year the student participated in the experiment, regardless of whether the student actually received the treatment in following grades. Therefore, this approach addresses the potential threat of non-random switching among classroom types. The regression equation for grade g (g = 1, 2, or 3) in each quantile (i.e., tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth) is

$$Y_g = \gamma_0 + \gamma_1 ITTSMALL_g + \gamma_2 FEMALE_g + \gamma_3 MINORITY_g + \gamma_4 LOWSES_g + \gamma_5 Y_{g-1} + \varepsilon_g ,$$

where ITTSMALL represents the intention or assignment by design. Notice that in this regression model previous class size membership is not included since it is not different from the assignment in the current grade (i.e., ITT estimate).

The third approach we used was an instrumental variables (IV) approach. Although 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 the size of small and regular classes. That is, the actual class size ranged from 11 to 20 for small classes and from 15 to 29 for regular classes. Hence, there was a modest overlap between the actual class sizes of the treatment conditions. This larger-than-intended variability in actual class size is likely non-random and perhaps the result of switching. This modest overlap in size between small and regular classes may have biased the estimate of the treatment effect. In other words, although target class size is assigned randomly, actual class size is not, and may be a result of non-random unobserved factors (e.g., parental or teacher pressure) that may also be related to the outcome. To address this issue the actual number of students in each class was regressed on the ITT small class variable (i.e., the instrument) and student characteristics such gender, race, and SES (see Angrist, Imberns, & Rubin, 1996; Krueger, 1999). The instrument is related to actual class size, but not related to achievement or any other variable by design. The predicted or fitted values of this regression were computed and used as the main predictor of achievement in the quantile regression. Specifically, the IV regression employed in the first stage of this analysis for grade g (g = 1, 2, and 3) is

$$CLSZ_g = \beta_0 + \beta_1 ITTSMALL_g + \beta_2 FEMALE_g + \beta_3 LOWSES_g + \beta_4 MINORITY_g + \varepsilon$$

where CLSZ is the actual class size in a classroom, and all other terms have been defined previously.

In the second stage, the predicted values of the regression model above were used to predict student achievement in the quantile regression. Specifically, the regression equation for grade g (g = 1, 2, or 3) in each quantile (i.e., tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth) is:

$$Y_g = \gamma_0 + \gamma_1 FITTED_g + \gamma_2 FEMALE_g + \gamma_3 MINORITY_g + \gamma_4 LOWSES_g + \gamma_6 Y_{g-1}$$

where FITTED represents the fitted values from equation (3) and all other terms have been already defined. Notice that in this regression model previous class size membership is not included since it is highly related with the fitted values.

To determine whether small class membership in more than one grade makes a difference we also conducted analyses using data across all grades to explore the effects of small classes on student achievement from kindergarten through third grade simultaneously. That is, we conducted analysis for students who were part of the experiment for four years. For this analysis we regressed third grade achievement on small class indicators (TOT) across all grades. The regression equation in each quantile is

$$Y_3 = \gamma_0 + \gamma_1 TOTSMALLK + \gamma_2 TOTSMALL1 + \gamma_3 TOTSMALL2 + \gamma_4 TOTSMALL3 + \gamma_5 FEMALE_3 + \gamma_6 MINORITY_3 + \gamma_7 LOWSES_3 + \varepsilon_3,$$

where subscript 3 indicates third grade.

Finally, in other analysis we investigated the cumulative effects of being in small classes (TOT) all four years on third grade achievement. In this analysis, the comparison group was being in regular classes all four years. The regression model for cumulative effects in each quantile is:

$$Y_3 = \gamma_0 + \gamma_1 FOURSMALL + \gamma_2 FEMALE_3 + \gamma_3 MINORITY_3 + \gamma_4 LOWSES_3 + \varepsilon_3,$$

where FOURSMALL represents the cumulative effects of small classes in all four grades.

For all models described above, we used STATA to run quantile regression and computed robust standard errors for the quantile regression estimates (via the cluster command). The robust standard errors we obtained take into account the clustering nature of the data (i.e., students nested within schools) as well as heteroscedasticity (i.e., non-constant variation).

Results

The descriptive statistics for variables of interest are summarized in Table 1. In kindergarten through third grade nearly 50 percent of the students were female and economically disadvantaged. Approximately one-third of the students were minority students. Nearly 25-30 percent of the students were in small classes across grades. The outcomes of interest were mathematics and reading scores that were standardized to have a mean of zero and a standard deviation of one. Hence, all estimates reported in the Tales 2 to 6 are in standard deviation units.

Table 1
Percentages of Variables Across Grades

	GRADE			
Variable	Kindergarten	First	Second	Third
Female	48.62	47.96	48.30	47.99
Minority	33.03	33.41	35.22	33.71
Low SES	48.44	51.35	51.61	50.54
Small Class	30.04	26.14	25.56	26.49
Number of Students	6,325	6,829	6,840	6,802
Number of Schools	79	76	75	75

Note.-SES = Socioeconomic Status

We discuss first the results of the TOT analysis. The class size estimates are summarized in Table 2. In particular, the estimates of small class effects and their standard errors are presented in Table 2. By and large, the small class estimates were insignificant with only a couple of exceptions. For example, in first grade mathematics the small class estimates in the middle and the lower tail of the distribution were positive and significant at 0.05. That is, in first grade middle- and low-achievers benefited more from being in small classes. The estimates in the upper tail were insignificant however indicating that high-achievers did equally well in both types of classes. In second grade mathematics only the small class estimate in the ninetieth quantile was significant. That is, in second grade, high-achievers in mathematics benefited more from being in small classes. In third grade mathematics none of the small class estimates were significantly different from zero. The results for reading achievement were not that different, only in reading all small class estimates were insignificant. Overall, these results suggested that once small class membership and achievement in the previous grade is controlled for current small class effects on achievement were not significant. Thus, it appears that the "value-added" effects of small classes on student achievement are trivial and not systematic.

Table 2

Small Class Effects Estimates in Mathematics and Reading at Various Quantiles Across Grades: TOT Analysis

	QUANTILE					
GRADE	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	N
Mathematics:						
1	0.232* (0.066)	0.156* (0.074)	0.175* (0.069)	0.114 (0.126)	0.158 (0.081)	4163
2	0.085 (0.125)	0.024 (0.098)	0.074 (0.104)	0.172 (0.123)	0.232* (0.115)	4656
3	-0.140 (0.113)	-0.194 (0.157)	-0.059 (0.158)	-0.057 (0.111)	-0.011 (0.101)	4685
Reading:						
1	0.142(.0.082)	0.121 (0.078)	0.040 (0.082)	0.105 (0.088)	0.052 (0.088)	4010
2	0.075 (0.096)	0.000 (0.107)	-0.053 (0.089)	0.004 (0.107)	0.067 (0.111)	4594
3	-0.035 (0.080)	-0.028 (0.091)	0.004 (0.061)	0.025 (0.093)	0.037 (0.094)	4710

Note.-Standard errors of estimates are in parenthesis

* $p < 0.05$

The estimates reported in Table 2 may be biased because parents, students, teachers and principals may have interfered with the fidelity of the experiment (e.g., switching among classes). As a result, we also conducted analysis that investigated the ITT small class effects. The class size estimates are summarized in Table 3. By and large, the small class estimates were insignificant. In first grade mathematics the small class estimates in the middle and the lower and upper quartiles of the distribution were positive and significant at 0.05. These results suggest that middle-, low-, and high-achievers benefited similarly from being in small classes. In second and third grade mathematics the small class estimates were significant, that is, in these two grades there was no evidence of the “value-added” of small classes, once small class membership and achievement in the previous grades were taken into account.

In first grade reading the small class estimate at the seventy-fifth quantile was significant, and in second grade reading the median estimate was significant. In third grade reading however, the estimates in the upper tail were both significantly different from zero. Hence, there is some weak evidence that in reading especially in grades 1 and 3 high-achievers seem to benefit more from small class membership. Overall, these results suggest that once achievement in the previous grade is controlled for the ITT small class effects are by and large not significant. In addition, the evidence of differential small class effects is weak.

Table 3

Small Class Effects Estimates at Various Quantiles Across Grades: ITT Analysis

	QUANTILE					
GRADE	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	N
Mathematics:						
1	0.047 (0.056)	0.089* (0.043)	0.143* (0.052)	0.183* (0.059)	0.161 (0.084)	4163
2	0.006 (0.072)	0.022 (0.038)	0.049 (0.051)	0.056 (0.066)	0.047 (0.090)	4656
3	0.066 (0.057)	0.034 (0.059)	0.025 (0.026)	0.052 (0.050)	-0.004 (0.065)	4685
Reading:						
1	0.067 (.0.057)	0.074 (.0.041)	0.102 (0.068)	0.163* (0.081)	0.144 (0.097)	4010
2	0.002 (0.049)	0.014 (0.035)	0.076* (0.034)	0.073 (0.045)	0.023 (0.060)	4594
3	0.010 (0.048)	0.038 (0.042)	0.081 (0.047)	0.160* (0.033)	0.160* (0.055)	4710

Note.-Standard errors of estimates are in parenthesis

* $p < 0.05$

The third set of estimates was obtained from the IV analysis. These estimates are summarized in Table 4. Intuitively, all estimates of Table 4 are negative since one would expect that as class size gets smaller achievement gets higher. However, the regression coefficients were typically not significant. In mathematics, the only significant estimates were observed in the first grade at the median and at the seventy-fifth quantile. In reading, class size estimates were significant in the first and third grades in the upper tail of the distribution, suggesting a small class advantage for high-achievers. Finally, the median estimate of class size was significant in the second grade. Again, as with estimates in Table 2 and 3 most small class effects were not different from zero.

Table 4

Small Class Effects Estimates at Various Quantiles Across Grades: IV Analysis

	QUANTILE					
GRADE	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	N
Mathematics:						
1	-0.007 (0.006)	-0.013 (0.011)	-0.022* (0.006)	-0.028* (0.012)	-0.024 (0.013)	4163
2	-0.001 (0.007)	-0.003 (0.003)	-0.007 (0.008)	-0.008 (0.007)	-0.007 (0.011)	4656
3	-0.010 (0.008)	-0.004 (0.006)	-0.004 (0.005)	-0.008 (0.008)	-0.001 (0.008)	4685
Reading:						
1	-0.010 (.0.006)	-0.011 (.0.009)	-0.015 (0.013)	-0.025* (0.011)	-0.022* (0.007)	4010
2	-0.000 (.0.008)	-0.002 (.0.005)	-0.011* (0.005)	-0.010 (0.062)	-0.003 (0.008)	4594
3	-0.001 (.0.008)	-0.005 (.0.004)	-0.012 (0.009)	-0.023* (0.007)	-0.023* (0.007)	4710

Note.-Standard errors of estimates are in parenthesis

* $p < 0.05$

Further, we estimated the conditional effects of being in small classes at various grades. In this analysis third grade achievement was modeled as a function of small class dummies for all four grades of the experiment. The results of this analysis are reported in Table 5. The estimates for mathematics achievement did not reach statistical significance. These results suggest that small class effects are not independent (i.e., there is confounding) of each other and that the effects likely persist from grade to grade. The results for reading scores are only slightly different. The overwhelming majority of estimates are still insignificant. However, the estimate at the fiftieth quantile in the second grade and the estimate at the seventy-fifth quantile in first grade were positive and significant.

Table 5
Small Class Effects Estimates on Third Grade Achievement at Various Quantiles

	QUANTILE					
	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	N
Third Grade Mathematics:						
Small Class in Kindergarten	0.075 (0.125)	0.050 (0.091)	0.100 (0.117)	0.100 (0.108)	0.000 (0.251)	2929
Small Class in First Grade	0.126 (0.193)	0.126 (0.172)	0.025 (0.171)	-0.038 (0.150)	0.000 (0.285)	2929
Small Class in Second Grade	-0.075 (0.254)	0.050 (0.235)	0.126 (0.155)	0.201 (0.209)	0.176 (0.288)	2929
Small Class in Third Grade	0.025 (0.172)	-0.075 (0.147)	-0.126 (0.144)	-0.063 (0.176)	0.000 (0.301)	2929
Third Grade Reading:						
Small Class in Kindergarten	0.078 (0.079)	0.104 (0.092)	0.156 (0.083)	0.233* (0.079)	0.233 (0.147)	2897
Small Class in First Grade	0.207 (0.151)	-0.026 (0.163)	-0.078 (0.073)	-0.233 (0.165)	0.000 (0.261)	2897
Small Class in Second Grade	-0.156 (0.166)	0.000 (0.196)	0.207* (0.100)	0.207 (0.191)	-0.078 (0.270)	2897
Small Class in Third Grade	0.052 (0.058)	0.078 (0.092)	-0.026 (0.086)	0.000 (0.154)	0.078 (0.212)	2897

Note.-Standard errors of estimates are in parenthesis
* $p < 0.05$

Finally, the cumulative effects of small classes on third grade achievement are presented in Table 6. The estimates indicate mean differences in achievement between students who were in small classes in all four grades and those who were in regular size classes in all four grades. All estimates are positive and significant indicating a cumulative small class effect. In mathematics the estimates in the upper tail were much larger than those in the lower tail, but this difference was not significant. In reading the estimates seemed more uniform across the distribution. Overall, the results suggest that being in small classes for four years is beneficial. However, these results do not identify the grade or grades that are most beneficial on student achievement.

Table 6
Cumulative Small Class Effects Estimates at Various Quantiles on Third Grade Achievement

	QUANTILE					
	Tenth	Twenty-fifth	Fiftieth	Seventy-fifth	Ninetieth	N
Third Grade Mathematics:						
Small Class in All Four Grades	0.151* (0.076)	0.100* (0.048)	0.126* (0.041)	0.201* (0.069)	0.276* (0.076)	2485
Third Grade Reading:						
Small Class in All Four Grades	0.207* (0.045)	0.156* (0.077)	0.259* (0.046)	0.207* (0.062)	0.194* (0.072)	2462

Note.-Standard errors of estimates are in parenthesis; Reference group is being in regular classes in all four grades
* $p < 0.05$

Discussion

The present study examined the differential "value-added" effects of small classes across the achievement distribution in an attempt to better understand the effects of class size reduction on student achievement. Specifically, we examined the small class effects on student achievement in one grade (e.g., first grade) controlling for small class effects and achievement in the previous grade (e.g., kindergarten).

The results suggest that once previous grade achievement and small class membership are taken into account the effects of being currently in a small class are typically not significantly different from zero. There were a couple of exceptions however. The TOT analysis showed a small class benefit in the middle and the lower tail of the first grade mathematics distribution. In second grade mathematics high-achievers benefited from being in small classes. The results of the ITT analysis also indicated a small class advantage for various levels of achievement in first grade mathematics. In addition, high-achievers in third grade benefited more from being in small classes in reading. Finally, the results of the IV analysis were similar to those reported in the ITT analysis. That is, there is some weak evidence (e.g., nearly 20 percent of the estimates are significant) that small class effects are additive from grade to grade. The small class benefit however, is sometimes evident for low-achievers and in other cases it is evident for medium- or high-achievers. Thus, although there is some weak evidence for differential effects of small class and level of achievement, the effects do not seem to be systematic or consistent.

The results of the longitudinal analysis that examined simultaneously the effects of small classes in kindergarten, first, second, and third grade on third grade achievement were similar to the results reported above. That is, once all small class indicators across grades were included in the same model only two of the 40 estimates (i.e., 5 percent) were significant. That is, the estimates may have been significant simply by chance. These results underline the notion that there is a spillover effect of small classes from grade to grade and that the small class effects in one grade are not independent from the effects of small classes in another grade. One could also argue that the small class effects are perhaps most important in one year and that additional years do not seem to have independent significant effects. The proportions of students switching among small and regular classes in each grade was not high (rarely greater than 10 percent) and, as a result, the small class indicators from grade to grade were strongly related to each other making it difficult to detect independent effects.

The longitudinal analysis examined the effects of the length of exposure to small classes in a four-year period. For example, nearly one-fourth of the students stayed in small classes in all four grades and it is plausible that this longer exposure to the treatment would produce larger cumulative small class effects. As expected the results of the longitudinal analysis that investigated the cumulative effects of being in small classes in all four grades (compared to being in regular size classes in all four grades) pointed to positive and significant effects. This finding is consistent with results from previous studies (Konstantopoulos & Chung, 2009). In mathematics the small class advantage was larger for high-achievers and consistently greater than one-fifth of a standard deviation, which is a considerable effect. One plausible hypothsis is that high achievers who are consistently in small classes for all four years create more learning opportunities in mathematics for themselves than other students. That is, high-achievers may seek to interact more with teachers in smaller classrooms about mathematics terms and tasks. Alternatively, teachers may create mathematics related activities in small classes that serve high-achievers more than other students. The cumulative estimates were overall larger in reading especially in the middle of the distribution, that is, in reading the average student gained more than other students from being in small classes consistently. One plausible hypothesis is that reading is taught by teachers more uniformly and targets the average student. These effects could be seen as causal if switching among classroom types and attrition within classroom types from grade to grade were random.

One potential limitation of this study is that we were unable to control for differences among schools. Typically, this is accomplished by including school fixed effects (i.e., dummies) in the regression equations. In principle, differences among schools could affect small class effects, since in Project STAR random assignment was conducted within schools. Technically, it was difficult for us to control for school effects in the quantile regression models since the estimation became unstable once the school dummies were included in the models (i.e., nearly 80 dummies to capture school effects in each quantile). Another potential limitation is that we were unable to model the possible inconsistency of small class effects across schools (see Konstantopoulos, 2011). This is typically captured via interaction terms between schools and the treatment. Again, this estimation was not possible because there were too many terms in the regression that needed to be computed. Further, it is unclear how different curricula may have affected the class size estimates reported here. If the curricula enacted by teachers were the same or very similar the class size estimates would have been affected similarly. If however, the curricula were different among schools then some interaction between curricula and class size is possible. Regardless, it is unfortunate that this kind of information was not available to us and therefore it was not possible to test this hypothesis. This is also a limitation of the study.

To conclude, the findings of this study did not provide consistent evidence about the differential "value-added" or independent effects of small classes in early grades. That is, the small class effects in specific grades fade when small class effects and achievement from previous grades are controlled for. Moreover, when the effects of small classes are conditioned on one another across a four-year period the small class advantage is not evident. It is unclear therefore, that being in small classes for two or more years is as beneficial as being in small classes for only one year. Although the cumulative small class effects are positive, significant, and meaningful in magnitude, they do not indicate which grade or grades are the most important. However, the results from the analysis that examined the small class effects across all grades simultaneously most likely suggest that being in small classes in one grade is probably enough and that the small class effects do not seem to accrue over time. Finally, the findings of the present study suggest no evidence that class size reduction can affect low-achievers more than high-achievers. Hence, it is not likely that class size reduction can close the achievement gap.

References

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect Of Class Size On Scholastic Achievement. Quarterly Journal of Economics, 114, 533–575.

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.

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

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., & Weinfeld, F. D., et al., (1966). Equality of educational opportunity. Washington, DC: U.S. Government Printing Office.

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, CA: 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., & Laine, R. D. (1996). The effects of school resources on student achievement, Review of Educational Research, 66, 361-396.

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

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. Educational Evaluation and Policy Analysis, 21, 143-163.

Hao, L., & Naiman, D. Q. (2007). Quantile Regression. Thousand Oakes, CA: Sage.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. Quarterly Journal of Economics, 115, 1239-1285.

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

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR. Elementary School Journal, 108, 275-291.

Konstantopoulos, S. (2011). How consistent are class size effects? Evaluation Review, 35, 71-92.

Konstantopoulos S, & Chung, V. (2009). What are the long-term effects of small classes on the achievement gap? Evidence from the Lasting Benefits Study. American Journal of Education, 116, 125-154

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

Milesi, C., & Gamoran, A.(2006). Effects of Class Size and Instruction on Kindergarten Achievement. Educational Evaluation and Policy Analysis, 28, 287–313.

Nye, B., Hedges, L.V., & Konstantopoulos, S. (2000a). 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. (2000b). 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.

Pong, S., & Pallas, A. (2001). Class Size and Eighth-Grade Math Achievement in The United States and Abroad. Educational Evaluation and Policy Analysis, 23, 251–273.

Spyros Konstantopoulos is associate professor and program coordinator of measurement and quantitative methods at the department of counseling, educational psychology, and special education at the college of education at Michigan State University

Contact address: Michigan State University, Erickson Hall, Room 450, East Lansing, MI 48824, United States of America. E-mail address: spyros@msu.edu