

# **Disruption, Skill and Heterogeneous Benefits of Smaller Classes**

Graham McKee, Steven Rivkin, and Katharine Sims

January 2010

## Abstract

With few exceptions, empirical research investigating the possibility of heterogeneous benefits of class size reduction lacks a conceptual framework about specific dimensions of potential heterogeneity. In this paper we develop a model of education production that incorporates disruption, curriculum targeting, and student heterogeneity in the rate of learning and explore the possibility that the benefits of smaller classes vary along multiple dimensions using the Tennessee STAR data. The estimates show that students in higher poverty schools and those higher up the school achievement distribution realize larger benefits from smaller classes. This heterogeneity along the school poverty dimension is consistent with the belief that class size reduction has a larger effect on time available for learning in higher poverty schools, and evidence from the ECLS that shows higher rates of disruptive behavior in higher poverty schools supports this interpretation. The heterogeneity by initial achievement is consistent with both the belief that smaller classes enable teachers to focus more attention on higher achievers and the belief that higher achieving students derive larger benefits for improvements to instruction. Although these explanations are not mutually exclusive, observations taken as part of Project STAR tend to support the latter explanation, is there is little or no evidence that smaller classes led to changes in teacher behavior or pedagogy.

## **I. Introduction**

Research generally shows that low income students and students of color derive larger benefits from class size reduction, but the rather ad-hoc specification of sub-groups is typically not based on underlying conceptual arguments regarding the pathways through which class size reduction affects what transpires in the classroom. Lazear (2001) provides a possible explanation for the observed pattern of class size estimates, focusing on the benefits of smaller classes in terms of a reduction in time lost to disruptive behavior. Yet benefits of smaller classes may also emanate from additional opportunities to tailor instruction to individual students or to use innovative curricula, a hypothesis that finds some support in Rice (1999).

In this paper we develop a model with two potential dimensions along which benefits of smaller classes might vary and use data from the Tennessee STAR random assignment experiment to investigate the pattern of class size effects and examine whether it is consistent with the existence of one or both of these channels. The first channel builds on the work of Lazear (2001) on the link between class size, family income, and time lost to disruption. Specifically, either because of less enriching pre-school experiences, less stability at home, less effective teachers, or some other factor, the probability of disrupting the class is hypothesized to be higher on average for lower income children. This higher initial propensity to disrupt causes class size reduction to have a larger effect on the share of class time in which there is no disruptive behavior, thereby leading to larger benefits of smaller classes in high poverty schools with more low income students.

The second potential channel works through an expansion of opportunities to differentiate instruction in order to meet the needs of students with varying levels of academic preparation. Although one might expect the curriculum to target the median student in terms of academic preparation, accountability regimes are based on pass rates and thereby provide an incentive to focus on students lower down the skill distribution. By permitting a manageable number of small groups and more time for individual instruction, class size reduction might expand the use of differentiated curricula within classes that would bestow particular benefit to higher achieving students who might otherwise gain little from class time.

Finally, the third potential channel recognizes that the benefits of class size reduction may vary by initial achievement due to differences in effort, ability to focus, or cognitive skills. On the one hand, struggling students may benefit more from improvements in the quality of instructional time or exert greater effort and focus better in smaller classes. On the other hand, high achievers may derive greater benefit because higher levels of effort or focus or more rapid processing of new information translates into higher returns from improvements in school quality.

We investigate variation in the benefits of class size by both income and achievement using quantile treatment effect (QTE) methods. The QTE estimates show a consistent pattern of larger class size effects for students at the upper end of the achievement distribution regardless of whether the sample includes all students or a subsample of students determined by income or the share of low income students in a school. The heterogeneity by initial achievement is consistent with both the belief that smaller classes enable teachers to focus more attention on higher achievers and the belief that

higher achieving students derive larger benefits for improvements to instruction.

Although these explanations are not mutually exclusive, observations taken as part of Project STAR tend to support the latter explanation, as there is little or no evidence that smaller classes led to changes in teacher behavior or pedagogy.

The QTE and OLS estimates also show that students in higher poverty schools realize larger benefits from smaller classes. This heterogeneity is consistent with the belief that class size reduction has a larger effect on time available for learning in higher poverty schools, and evidence from the ECLS that shows higher rates of disruptive behavior in higher poverty schools supports this interpretation. The fact that heterogeneity on the basis of individual income is much less pronounced lends additional support for the importance of disruption, as all students and not just lower income students benefit from the increase in time available for instruction.

The remainder of the paper is structured as follows. Section II reviews selected literature on the estimation of class size and quantile treatment effects. Section III develops the model through which the effects of class size on achievement vary along two dimensions. Section IV describes the Tennessee STAR data and the empirical model. Section V describes the estimates and brings additional evidence to bear on the causal channels, and Section VI summarizes the analysis and discusses policy implications.

## **II. Selected Research on Class Size Effects**

Since the release of the Coleman Report in 1966 a great deal of research on the effects of smaller classes has been completed, most of which focuses on the average treatment effect but some of which investigates heterogeneous effects along various

dimensions. The purposeful sorting of families into schools and allocation of resources among schools complicates identification of class size effects, because it is difficult to identify exogenous variation in class size. In most cases those currently in large classes do not provide a valid counterfactual for those currently in small classes, requiring statistical methods to account for unobserved heterogeneity among students and schools in order to identify class size effects.

Recent research utilizes innovative statistical methods and a large scale random assignment experiment to identify the causal effect of smaller classes and examine heterogeneity along several dimensions. Studies by Angrist and Lavy (1999), Hoxby (2000), Jepsen and Rivkin (2008), and Rivkin, Hanushek, and Kain (2005) identify plausibly exogenous variation in class size in order to learn more about the effects of class size on achievement (in Israel, Connecticut, California, and Texas, respectively). The studies generally find that smaller classes significantly increase achievement in the early grades (grade 5 and below), and the effects tend to be larger for lower-income students.<sup>1</sup>

Babcock and Betts (2009) focus on lower achieving children and consider both behavior and academic factors in their analysis of the effects of class size. Using information from first grade report cards they attempt to disentangle the source of the benefit of smaller classes by seeing whether it is the students graded poorly on effort or those graded poorly on academic achievement who derive larger benefits from smaller classes. The results suggest that it is the students with lower effort who benefit more from smaller classes. However, there remain some concerns that the difference-in-differences

---

<sup>1</sup> The exception is Hoxby (2000), but data deficiencies introduce downward bias which might account for the small and insignificant estimates.

framework may not control for all confounding factors, particularly given the likelihood that improvements in effort and behavior are likely to benefit all students by reducing the amount of class time lost to disruption.

Three recent papers use quantile treatment effect methods on non-experimental data to investigate the possibility that the impact of smaller classes varies along the achievement distribution. Eide and Showalter (1998) find no significant benefit of smaller classes at any quantile using US data. However, the treatment of class size as exogenous raises concerns about the possibility that unobserved factors bias the estimates. Levin (2001), and Ma and Koenker (2006) use two-stage quantile treatment effect methods with the same Dutch school data to account for unobserved factors. Levin groups classes into six size categories and finds that little evidence of a significant relationship between class size and achievement across most of the distribution in grades 4, 6, and 8. Ma and Koenker analyze the benefits of marginal class size reductions rather than creating a small number of class size categories. Their results on continuous treatment effects are similar to those of Levin.

The finding of little or no benefit to smaller classes in Dutch schools contradicts much recent research on US schools, raising concerns that methods used to account for unobserved heterogeneity do not control for all confounding factors. This leaves open the possibility that the benefits of smaller classes do vary systematically across the achievement distribution.

Importantly, the Tennessee STAR experiment, where students were randomly assigned to small classes or larger classes, provides a good arena for identifying both the

average benefits of small classes and any heterogeneity in class size reduction effects.<sup>2</sup> Word et al., (1990), Finn and Achilles (1990), Krueger (1999), Finn, Gerber, Achilles, and Boyd-Zaharias (2001), and Krueger and Whitmore (2001) analyze the STAR experiment and find that smaller class sizes have a significant and lasting impact on academic achievement and educational attainment. These studies also find larger benefits for low income students and students of color.

Several more recent studies use QTE methods estimated over the STAR data to investigate heterogeneity in the benefits of smaller classes. Both Ding and Lehrer (2005) and Konstantopoulos (2008) find that higher ability students derive larger benefits from class size reduction. In contrast to much existing research, Ding and Lehrer (2005) also find little or no evidence that minority or low income students derive additional benefits from smaller classes.

### **III. Conceptual Framework**

This section develops a model of the relationship between class size and the quality of instruction with three sources of variation for the benefits of smaller classes. The first, building on Lazear (2001), comes from the relationship between class size and time lost to disruption for students in schools that differ according to the share of children in poverty. The second potential source of heterogeneity emanates from the relationship between class size and the relevance of the curriculum for students in different parts of the classroom achievement distribution,

---

<sup>2</sup> A large-scale class size reduction would require a substantial expansion of the number of teachers which could adversely affect the quality of instruction and the benefits of the program. Jepsen and Rivkin (2008) examine the effects of the California class size reduction on the quality of teachers.

and the third source comes from differences due to student effort, ability to focus, and initial skills.

Equation (1) models learning for student  $i$  in classroom  $c$  in school  $j$  as a function of the amount of classroom time available for learning and the value of that time in terms of learning, plus all other student, community, and school factors:

$$(1) \quad \textit{learning}_{icj} = \rho(d)_{cj}^n q(n, a)_{icj} + X_{icj}$$

where

$\rho$  is the proportion of time a student is not disrupting the class;

$d$  is the classroom average propensity to disrupt the class;

$q$  is the value of a unit of instructional time;

$n$  is class size;

$a$  is an index of academic preparation;

And  $X$  is a vector of other student, community and school factors.

The term  $\rho(d)_{cj}^n$  is drawn from Lazear (2001) and represents the share of class time not lost to disruption by any of the  $n$  students in the room, while the term  $q(n, a)_{icj}$  models the value of a unit of instructional time as function of both class size and academic preparation. Importantly, all students in a classroom experience the same amount of instructional time, but the value of instructional time may vary by academic preparation.

Variation in classroom behavior,  $d$ , and academic preparation,  $a$ , provide two dimensions of potential heterogeneity in the benefit of smaller classes. In order to illustrate the ways in which disruption and academic preparation may affect the benefits

from class size reduction, we take the derivative of equation 1 with respect to n and then again with respect to d (equation 2) and a (equation 3 below):

$$(2) \quad \frac{\partial^2 \text{learning}}{\partial n \partial d} = \{[\rho(d)]^{n-1} \frac{\partial \rho(d)}{\partial d}\} * \{(\ln(\rho(d)) + 1) * q(n, a) + n \frac{\partial q(n, a)}{\partial n}\}$$

Equation (2) illustrates the relationship between the propensity to disrupt class and the benefit of class size reduction.<sup>3</sup> The product of the two relationships in squiggly brackets determines the sign of the cross-partial derivative of learning with respect to n and d. The first is negative, as the derivative of  $\rho$  with respect to d is assumed to be negative (a higher average propensity to disrupt reduces the share of time available for learning).

The sign of the second is ambiguous and depends on the magnitudes of the various terms.  $n \frac{\partial q(n, a)}{\partial n}$  is assumed to be negative: the quality of instructional time is assumed to decrease as class size increases.  $q(n, a)$  is positive and the product of  $\ln(\rho)$  and n lies between 0 and -1, so  $n \ln(\rho) + 1$  is also positive.<sup>4</sup>

Thus the relationship between the benefits of class size reduction and the degree of disruption (d) thus depends upon the magnitudes of two counteracting effects. First, as Lazear (2001) points out, at lower values of  $\rho$  reduced class size has a larger effect on the share of time available for learning and thus a larger effect on achievement. Second, at lower values of  $\rho$  any improvement in the quality of instruction time due to smaller classes has a lower overall impact, because classrooms with lower values of  $\rho$  have less

---

<sup>3</sup> The derivative of learning with respect to class size (n) equals

$[\rho(d)]^n \ln(\rho(d))q(n, a) + [\rho(d)]^n \frac{\partial q(n, a)}{\partial n}$ . Taking the derivative of this relationship with respect to

d produces Equation (2) and with respect to a produces equation 3.

<sup>4</sup> At a value of  $\rho$  below 0.95,  $n \ln(\rho) + 1$  becomes negative, but at such a low value of  $\rho$  the share of class time available for instruction is well below 50 percent.

time for learning. This effect, not considered by Lazear, may partially, fully, or even more than fully offset the higher benefit resulting from a larger decrease in the amount of time lost to disruption.

Equation (3) illustrates the relationship between initial achievement and the benefit of class size reduction:

$$(3) \quad \frac{\partial^2 \text{learning}}{\partial n \partial a} = [\rho(d)]^n * \left\{ (\ln(\rho(d))) + \frac{\frac{\partial^2 q(n, a)}{\partial n \partial a}}{\frac{\partial q(n, a)}{\partial a}} \right\}$$

As is the case with disruption, the relationship between the benefit of class size reduction and initial achievement cannot be signed a priori in this framework. Here the sum of the two relationships in squiggly brackets determines the sign of the cross-partial derivative of learning with respect to  $n$  and  $a$ . The term  $\ln \rho(d)$ , roughly the average disruption of a

single student, is negative, while the ratio  $\frac{\frac{\partial^2 q(n, a)}{\partial n \partial a}}{\frac{\partial q(n, a)}{\partial a}}$  is positive if class size

induced increases in the quality of instructional time rise with achievement.

As discussed above, this pattern could result from changes in pedagogy, distribution of teacher effort, or student characteristics. Consider first the link between teacher behavior and the distribution of the returns to smaller classes. Higher achieving students will tend to benefit more from smaller classes if 1) teachers tend to target the curriculum and their efforts toward less well-prepared students  $\left( \frac{\partial q(n, a)}{\partial a} < 0 \right)$  and 2) the emphasis on the lower portion of the distribution increases with class size,

$\left( \frac{\partial^2 q(n, a)}{\partial n \partial a} < 0 \right)$  (The increasing emphasis on less well-prepared students would occur if

individualized instruction or the division of the classroom into small groups becomes more difficult in larger classes.) Thus the sign of the cross-partial derivative of learning with respect to  $n$  and  $a$  will be negative if the proportional increase in the degree of emphasis on less-well prepared students as class size increases exceeds the “average” disruption of a student. Note that the hypothesis that  $\frac{\partial q(n, a)}{\partial a}$  is negative is based on the belief that schools, particularly in the accountability era that focuses on pass rates, target the curriculum toward those at or near the bottom of the distribution.

Now consider the link between student characteristics and the benefits of smaller classes. If students higher up the achievement distribution tend to learn more per unit of instructional time,  $\frac{\partial q(n, a)}{\partial a} > 0$ . Babcock and Betts (2009) focus on effort, but ability to focus may also underlay any heterogeneity in the rate of learning. If in addition, higher achievers tend to realize larger benefits from class size reduction induced improvements in the quality of instruction,  $\frac{\partial^2 q(n, a)}{\partial n \partial a} < 0$ , the cross partial of derivative of learning with respect to  $n$  and  $a$  would be unambiguously negative: higher achieving children would be expected to derive a larger benefit from smaller classes.

Clearly, multiple channels could separately or together produce the finding that higher achievers benefit disproportionately from smaller classes, and other evidence would have to be brought to bear to provide information on causal mechanisms. And there remains the possibility that lower achievers realize larger class size reduction induced improvements in the quality of instruction and  $\frac{\partial^2 q(n, a)}{\partial n \partial a} > 0$ . In combination

with higher achievers learning more per unit of instructional time, this would lead to larger effects of class size reduction at the lower end of the achievement distribution.

In sum, this model of the link between learning and class size suggests disruption and initial achievement as two dimensions over which the return to class size reduction may vary. However, the pattern of heterogeneous effects cannot be predicted a priori. Not only does it become an empirical question, but differences in district, state, or federal policies may produce variation in the distribution of treatment effects across each of these dimensions.

#### **IV. Data and Empirical Model**

This section describes the Tennessee STAR data and the QTE specifications used to investigate the pattern of heterogeneity in the effects of smaller classes on achievement. The data description focuses on potential deviations from an ideal experiment by comparing the observed characteristics of students in large and small classes, and the discussion of the empirical models discusses the proxy used to measure the school average propensity to disrupt and the interpretability of the QTE estimates.

##### *IV.a. Data*

Project STAR was legislated by the Tennessee government and carried out by researchers from the state's four universities (Tennessee State, Memphis State, the University of Tennessee, and Vanderbilt). In the experiment, students and teachers were randomly assigned to small classes (13-17 students), regular classes (22-25), or regular classes with teacher assistants from kindergarten through 3<sup>rd</sup> grade. Previous research

finds no significant effect of having an aide, and we group students in the two types of large classes together.<sup>5</sup>

Following Lehrer and Ding (2005), we also limit the study to kindergarten in order to mitigate problems introduced by non-random attrition and deficiencies in the assignment to classrooms in later grades necessitated in large part by student mobility. Deviations from the ideal experiment including non-random attrition from both the small and large classes and non-random assignment of teachers to classes potentially introduces bias. Hanushek (2001) documents the extensive attrition and classroom switching that took place following kindergarten in the STAR experiment and highlights the limited information on teacher quality available in the study.

Seventy-nine schools participated in the experiment. These schools are not a random sample, as they had to meet criteria such as size (large enough to have three classes in grades K-3) and location (the legislation required representation of inner city, urban, suburban, and rural schools). Because of this, inner city schools are overrepresented, and the share of minority students in the sample exceeds the state average.

Table 1 presents summary statistics for students in large and small classes. There are no pre-experiment test scores, so it is impossible to determine if the kindergarten groups are similar along this dimension. However, the two kindergarten groups are similar with respect to other characteristics. Only a slightly higher percent (1.5%) of regular class or regular class with TA students received free school lunch and only a slightly higher percent (1.6%) of small class students are white. Age is also very similar.

---

<sup>5</sup> The inclusion of an indicator based on initial assignment to a large class with an aide has virtually no effect on the magnitude or significance of the coefficient on small class.

The t-tests in the right-most column confirm that these differences are not significant, suggesting that random assignment is a reasonable assumption for kindergarten. Note that the mathematics and reading achievement test scores are transformed to have means of zero and standard deviations of one.

#### *IV.b. Empirical Model*

Identification of heterogeneity in class size effects by level of disruption and underlying skill requires a full accounting for confounding factors, the identification of a valid proxy for disruption that is unrelated to other determinants of the benefit of smaller classes, and satisfaction of the assumption that student rank in the test score distribution is invariant to class size. Following the description of the empirical model we discuss the use of the share of students eligible for a subsidized lunch as a proxy for the level of disruption and then consider issues related to the interpretability of the QTE estimates.

Consider the following equation in which achievement for student  $i$  is a function of class size status and an error that captures student heterogeneity, differences in teacher quality among classrooms, and random noise.

$$(3) \quad A_i = \beta_{\text{smallclass}_i} + \varepsilon_i$$

In the ideal experiment with random assignment of students and teachers and no switching or attrition, achievement in the large classes would provide a valid counterfactual for what students in small classes would have achieved in the absence of the treatment. Therefore  $\hat{\beta}$  would provide a consistent estimate of the average treatment effect, and QTE estimates would provide consistent estimates of the distance between the small class and large class achievement distributions at a given quantile.

Given the balance between small and large classes in the observed characteristics it appears that any deviations from the ideal experiment in kindergarten are likely to be small, and the primary reason for including student and teacher characteristics is to reduce sampling error and increase the precision of the estimates. We also include school fixed effects in some specifications based on the fact that the randomization occurred within schools.

QTE estimates provide information on heterogeneity by achievement, but there is no direct measure of disruption. Although many factors contribute to the probability of disruptive behavior, the higher rate of family instability, more limited opportunities to attend enriching pre-school, more difficult time attracting teachers, and more limited school resources likely lead to more disruption on average in schools with a higher share of lower income children. Therefore we use the share of students eligible for a subsidized lunch as a proxy for the level of disruptive behavior.

Of course the school share of low income students may have a direct relationship with achievement for a number of reasons including those outlined above, making it necessary to account for these other linkages in order to draw inferences about any variation in class size effects. Importantly, estimation of differences in class size effects by school subsidized lunch eligibility focuses on the interaction between class size and the low income share and not on the direct effect of the low income share on achievement. Therefore school fixed effects can be included as controls for differences among schools including those related to subsidized lunch eligibility.

We recognize as Barlock and Betts (2009) point out that other factors related to low income might also affect both achievement directly and the benefit of class size

reduction. Therefore we include indicators for free lunch eligibility in all specifications and also estimate separate class size effects by student eligibility for a subsidized lunch rather than the school low income share in order to provide additional information on the merits of the disruption interpretation.

A final consideration is the interpretability of the QTE estimates. The QTE estimates are difficult to interpret because the marginal achievement distributions of those in small and those in large classes provide limited information about the joint distribution. The difference between the  $q$ th quantiles of the small class and large class distributions provides a valid estimate of the treatment effect at the  $q$ th quantile only if a student would be in the  $q$ th quantile regardless of treatment status. Satisfaction of this rank preservation assumption cannot be demonstrated empirically, and therefore the results are most cautiously interpreted as reflecting differences in the marginal distributions.

## **V. Results**

The first portion of this section reports the empirical analysis of the STAR data. Following this discussion of results, we provide evidence on the likely causal mechanisms that contribute to any observed heterogeneity in class size effects. This evidence includes classroom observations on teacher and student behavior in a subsample of Project STAR classrooms and teacher reports on student misbehavior from the Early Childhood Longitudinal Survey-Kindergarten Cohort.

### *Va. Analysis of Class Size Effects*

The investigation of class size effects begins with an analysis of the entire sample that ignores variation due to disruption and focuses solely on heterogeneity along the achievement distribution.<sup>6</sup> The sample is then divided in half on the basis of the school share of students eligible for a subsidized lunch, and the patterns of class size effects are examined separately for the two groups of schools. In order to provide additional information on the source of any heterogeneity by income, the sample is also divided on the basis of individual student eligibility for a subsidized lunch. All specifications include indicator for teachers with fewer than three years of experience and teachers with an advanced degree, student demographic information including race-ethnicity, gender, and subsidized lunch eligibility indicators. Specifications that do not include school fixed effects also include indicators for community type (suburban, rural, urban, and inner-city).

Figure 1 shows kernel density plots by class size and subject. Neither of the plots suggests that a smaller class shifts out the distribution more at lower levels of achievement. If anything, the plots suggest a larger effect at higher levels of achievement, as the horizontal distances between the small and large class plots tend to increase with achievement. Note that the plots do not adjust for differences in demographic characteristics or limit the identification of the class size effect to within school differences only. Nonetheless, the random assignment experimental design and observed balance between small and large classes suggests that the inclusion of such controls should not lead to substantial changes in the achievement differences by class size.

Table 2 reports estimates of the average benefit of a small class and the magnitudes of the shifts in the distributions at the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 90<sup>th</sup> percentiles

---

<sup>6</sup> See Hao and Naiman (2007) and Koenker (2001) for further information on quantile regression.

for specifications with and without school fixed effects, and the estimates produce a pattern of class size effects similar to that shown in Figure 1. Regardless of whether school fixed effects are included estimates of class size effects increase monotonically as you move up achievement distribution in both mathematics and reading. Not surprisingly given the previous findings on the STAR experiment, the estimated effect of small classes is positive and highly significant.

The 90<sup>th</sup>-10<sup>th</sup> and 75<sup>th</sup>-25<sup>th</sup> percentile differences in effect magnitudes differ only slightly by specification, and we focus on the models without school fixed effects. In mathematics, the 90<sup>th</sup>-10<sup>th</sup> percentile difference exceeds 0.21 standard deviations and the 75<sup>th</sup>-25<sup>th</sup> percentile difference exceeds 0.11 standard deviations; the corresponding figures for reading are 0.16 standard deviations and 0.08 standard deviations.

This finding that small class size leads to a greater outward shift at the upper end of the achievement distribution is consistent with previous QTE estimates discussed above. We now separate the sample by low income share in order to explore disruption as another possible source of heterogeneity in the return to smaller classes.

Table 3 reports QTE and OLS estimates for schools above and schools below the median share of students eligible for a subsidized lunch, and the results provide strong support for the hypothesis that the benefits of class size reduction are larger in high poverty schools. Differences in the OLS estimates of average class size effects exceed 0.08 standard deviations in math and 0.10 standard deviations in reading regardless of whether school fixed effects are included

In terms of heterogeneity by underlying skill, the QTE estimates in Table 3 show that the pattern of heterogeneous returns to smaller classes found for the sample as a

whole also holds for both higher income and lower income schools. This pattern is particularly pronounced for schools with higher shares of low income students regardless of whether the specification includes school fixed effects. In mathematics the 90-10 difference exceeds 0.19 standard deviations and the 75-25 difference exceeds 0.14 standard deviations regardless of whether school fixed effects are included. The differences in reading are slightly smaller: the 90-10 difference equals 0.17 standard deviations and the 75-25 difference equals 0.13 standard deviations in the specifications with school fixed effects, and those from the specification without fixed effects are only slightly smaller.

As noted above, the pattern of increasing magnitudes as achievement rises appears less pronounced for the sample of schools with a smaller fraction of students eligible for a subsidized lunch, though for each subject and specification the 90-10 differential exceeds 0.11 and for all but the reading specification with school fixed effects the 75-25 differential exceeds 0.05.

Although the results in Table 3 provide strong evidence of a larger return to class size reduction in lower income schools, the lack of an explicit measure of disruption raises questions about the source of this heterogeneity. A plausible alternative hypothesis is that lower income children realize larger benefits due to individual circumstances rather than the classroom environment. If this is in fact the case we would expect a higher effect for lower income students regardless of school demographic composition.

Table 4 reports class size estimates produced from samples divided by subsidized lunch eligibility rather than the share of eligible students in a school, and the difference by individual income is far smaller than the difference by the school share of low income

students.<sup>7</sup> This provides additional support for the disruption interpretation, as it is something about the school rather than the individual that drives the differential class size effect by low income share. Note also that the pattern of increasing class size effects as you move up the achievement distribution is present for both samples, providing additional evidence that this finding holds across schools and student populations.<sup>8</sup>

*Vb. Additional Evidence on Causal Mechanisms*

The findings reveal heterogeneous class size effects by both the school share of low income students and achievement, but to interpret these results as supporting specific causal channels would require additional evidence. In this section we present additional evidence supporting 1) the use of school share of low income students as a proxy for the level of disruption; and 2) the hypothesis that the higher benefit realized by higher achieving students results from student heterogeneity rather than changes in teacher behavior in response to smaller classes.

Table 5 reports the distribution of kindergarten teacher responses to the statement: “The level of child misbehavior in this school interferes with my teaching,” by the school share of students eligible for a subsidized lunch. The responses reveal sharp differences in the probability that teachers experience disruptive behavior. In both the fall and spring surveys teachers in schools with a subsidized lunch eligible share above the median are more than two and one half times as likely to agree or strongly agree with the statement than teachers in schools with a below median share of subsidized lunch eligible students.

---

<sup>7</sup> Notice that the specifications used in this paper permit the effects of all explanatory variables to vary by income, not just class size. This may account for the much smaller differential by individual income than is reported in other work using the STAR data.

<sup>8</sup> School fixed effect estimates (not reported) produce similar findings, though the fixed effects must be constrained to be the same for both income groups because of small sample numbers of children of a given income type in many schools.

Although these responses come from the nationally representative Early Childhood Longitudinal Survey-Kindergarten cohort and not from the same set of Tennessee schools as the STAR data, there is no reason to believe that this relationship does not hold for Tennessee.

Of course there could be differences in other school characteristics related to the share of low income students that could also affect the return to smaller classes, and it is not possible to rule them all out. However, there is some evidence that differences in teacher quality do not account for the variation in class size effects. Specific, OLS estimates (not reported) that permitted the benefit of smaller classes to vary by teacher experience found little or no evidence of such variation, as the interaction term between an indicator for being in the first or second year and class size was insignificant at any conventional level. Certainly there exist other dimensions by which teacher quality can vary, but the strong evidence of substantial quality differences between rookie and more experienced teachers suggests that the coefficient on the interaction between the indicators for teacher inexperience and small class should have been significant had quality differences accounted for the variation in class size effects.

The evidence on the source of heterogeneity by achievement comes from observations of STAR classrooms. These observations provide little or no support for the belief that teacher behavior or pedagogy differed significantly by class size (Finn et al, 2007), raising doubts that changes in the distribution of effort, increased differentiation, or other pedagogical changes underlay the higher return for higher achievers. These observations suggest instead that differences among students account directly for the observed heterogeneity in returns.

## **Conclusion**

We investigate variation in the benefits of class size by both income and achievement in order to understand better the sources of any heterogeneity in the benefits of smaller classes. The QTE estimates show a consistent pattern of larger class size effects for students at the upper end of the achievement distribution regardless of whether the sample includes all students or a sub-sample of students determined by income or the share of low income students in a school. In addition, both OLS and QTE estimates show that students in high poverty schools realize much larger benefits of smaller classes than students in schools with a lower share of low income students.

The pattern of estimates and supporting evidence provides support for the hypothesis that children in more disruptive environments realize larger benefits from smaller classes. Not only are class size effects larger in higher poverty schools, but the class size effect differential is much smaller when calculated by individual student income. The survey evidence that misbehavior is a much larger problem in schools with high shares of lower income students supports the disruption interpretation.

The QTE estimates and supporting observational evidence of little or no changes in teacher behavior or pedagogy following class size reduction supports the hypothesis that student differences directly affect the return to smaller classes. Whether it is greater effort, better ability to focus, or skills that facilitate more rapid learning, it appears that student differences account for the positive relationship between achievement and the benefit of smaller classes. The finding that the ordering by achievement is more pronounced in the schools with higher low income shares is consistent with the hypothesis that higher achievers realize larger benefits from smaller class size induced

improvements in the quality of instruction, as these schools appear to have experienced larger increases in the amount of instructional time.

Regardless of the precise channels through which class size affects achievement, the evidence indicates that the benefits of class size reduction differ systematically by level of poverty and achievement, even ignoring any accompanying changes in the quality of teachers or other general equilibrium effects. Importantly, the findings provide mixed evidence on the likely impact of class size reduction on the achievement gap. On the one hand smaller classes appear to increase the achievement variance within schools and thus provide disproportionate benefit to the more successful students. On the other hand, the higher benefits to children in higher poverty schools contributes to the closing of the achievement gap, particularly since subsidized lunch eligible students in these schools are likely to be poorer on average than similarly classified students in lower poverty schools. Perhaps more important, smaller classes appear to significantly increase achievement even for low achievers in high poverty schools. If raising the achievement of the lowest performers is a primary goal this would appear to be more important than any impact on dispersion in the consideration of whether the benefits of class size reduction exceed the costs.

## References

- Angrist, Joshua, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2): 533-76.
- Babcock, Philip and Julian Betts. 2009. "Reduced-Class Distinctions: Effort, Ability, and the Education Production Function. NBER Working Paper number 14,777.
- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Ding, Weili & Steven Lehrer, 2005. "Class Size and Student Achievement: Experimental Estimates of Who Benefits and Who Loses from Reductions," Working paper.
- Djebbari, Habiba; Smith, Jeffrey. 2008. "Heterogeneous impacts in PROGRESA." *Journal of Econometrics*, 145 (June): 64-80.
- Finn, J. D., & Achilles, C. A. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27, 557-577.
- Finn, J. D., Boyd-Zaharias, J., Fish, R. M., & Gerber, S. B. (2007). *Project STAR and Beyond: Database Users Guide*. Lebanon, TN: HEROS, Inc. Retrieved from <http://www.heros-inc.org/data.htm>.
- Finn, J. D., A. Suriani, and C. A. Achilles. 2007. "Small Classes in the Early Grades: One Policy – Multiple Outcomes" unpublished manuscript. (November).
- Firpo, Sergio. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica*, 75:1 (January): 259-276.
- Hanushek, Eric. 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: 2 (Summer): 143-163
- Hao, Lingxin; Naiman, Daniel. 2007. *Quantile Regression*. Thousand Oaks, CA : SAGE Publications, Inc.
- Hoxby, Caroline Minter. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115(4): 1239-86.
- Koenker, Roger; Hallock, Kevin. 2001. "Quantile Regression." *Journal of Economic Perspectives*, 15:4 (Fall): 143-156.

- Konstantopoulos, Spyros, "Do Small Classes Reduce the Achievement Gap Between Low and High Achievers, Evidence from Project STAR?" *Elementary School Journal*, 108: 275-291.
- Krueger, Alan. 2003. "Economic Considerations and Class Size." *The Economic Journal*, 113 (February): 34-63.
- Krueger, Alan. 1999. "Experimental Estimates of Education Production Functions." *The Quarterly Journal of Economics*, 114:2 (May): 497-532.
- Krueger, Alan ; Whitmore, Diane. 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." *The Economic Journal*, 111 (January): 1-28.
- Lazear, Edward. 2001. "Educational Production." *Quarterly Journal of Economics*. 116: 3 (August) 777-803.
- Levin, Jesse. 2001. "For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement." *Empirical Economics*, 26: 221-246.
- Ma, Lingjie; Koenker, Roger. 2006. "Quantile regression methods for recursive structural equation models." *Journal of Econometrics*, 134 (September): 471-506.
- Maitra, Pushkar and Farshid Vahid. 2006. "The Effect of Household Characteristics on Living Standards in South Africa 1993-1998: A Quantile Regression Analysis with Sample Attrition." *Journal of Applied Econometrics*. 21: 7 (November) 999-1018.
- Rice, Jennifer. 1999. "The Impact of Class Size on Instructional Strategies and the Use of Time in High School Mathematics and Science Courses." *Educational Evaluation and Policy Analysis*, 21: 2 (Summer): 215-229.
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73(2): 417-58.
- Word, E., Johnston, J., Bain, H. P., Fulton, D. B., Boyd-Zaharias, J., Lintz, M. N., Achilles, C. M., Folger, J., & Breda, C. (1990). *Final Report: Student/teacher achievement ratio (STAR): Tennessee's K-3 class-size study*. Nashville, TN: Tennessee State Department of Education. Retrieved from <http://www.heros-inc.org>.

Figure 1. Kernel Density Plots of Kindergarten Standardized Achievement by Class Size and Subject

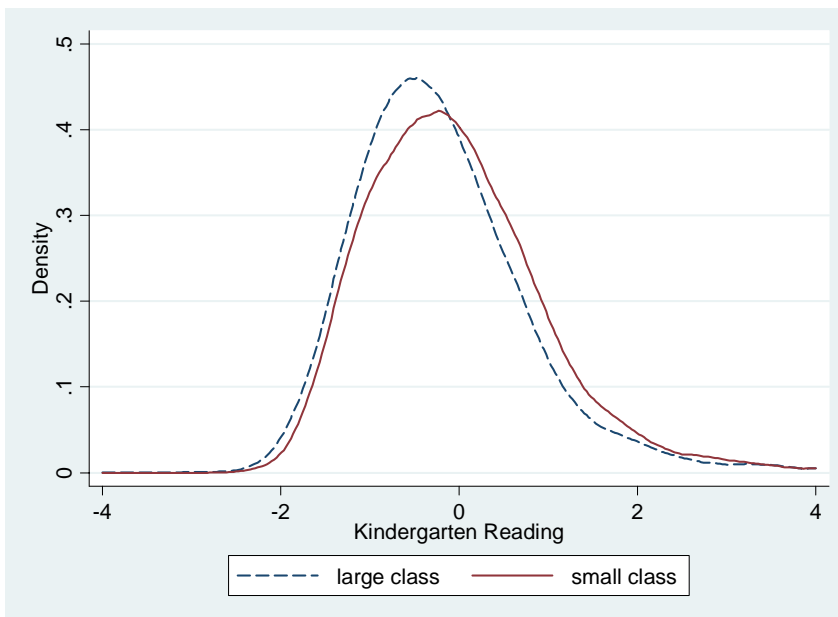
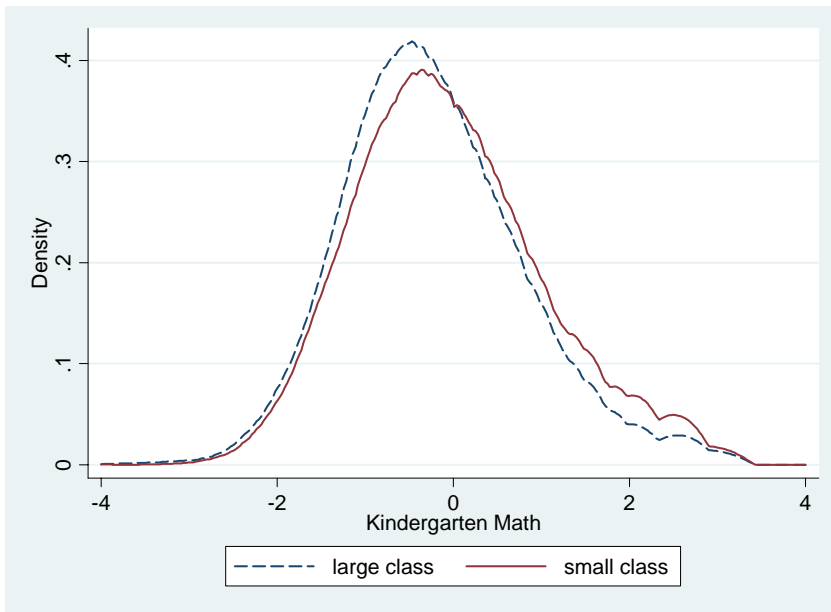
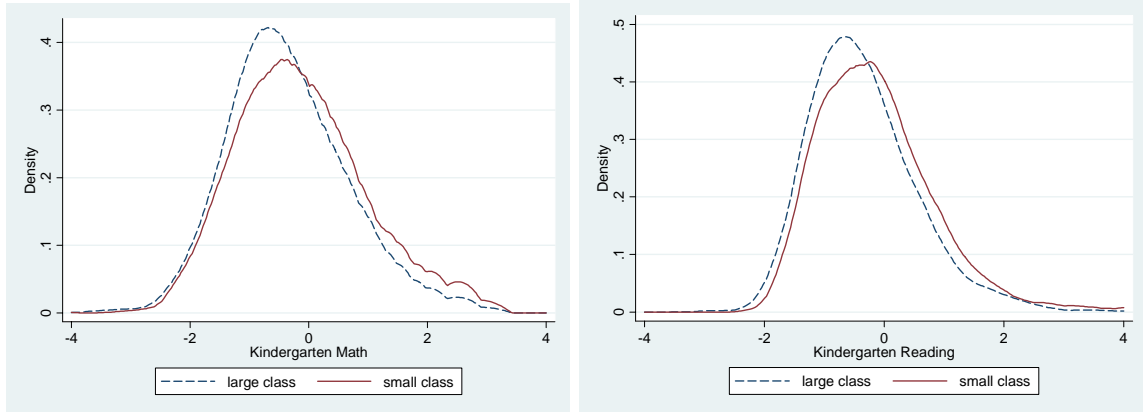


Figure 2. Kernel Density Plots of Kindergarten Standardized Achievement by Share of Students Eligible for a Subsidized Lunch, Class Size, and Subject

### High Share of Subsidized Lunch Eligible Students



### Low Share of Subsidized Lunch Eligible Students

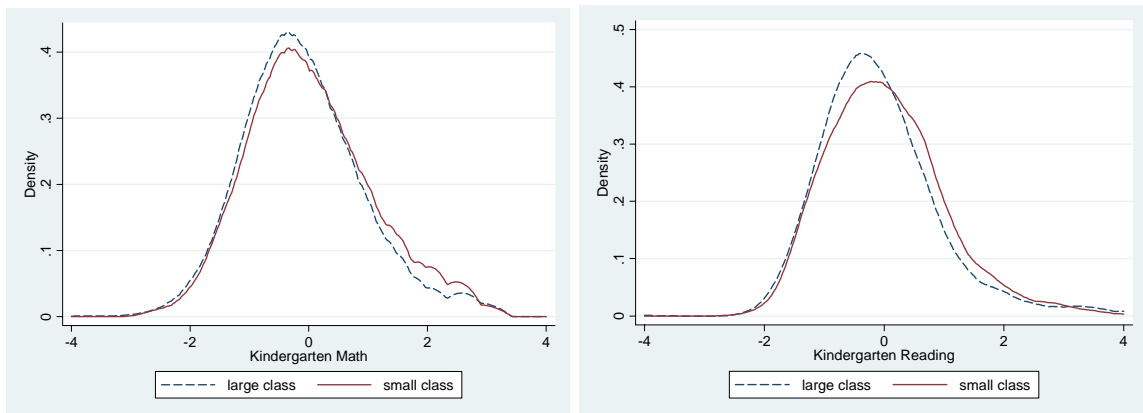


Table 1  
 Summary statistics by kindergarten class size

	small			regular			pr( T >t)
	Obs	Mean	SD	Obs	Mean	SD	
Grade K							
Free lunch	1754	47.1%	0.499	4089	48.6%	0.5	0.28
Non-white	1754	31.6%	0.465	4089	33.2%	0.471	0.26
Age	1754	5.47	0.34	4089	5.46	0.34	0.30
Class size	1754	15.12	1.50	4089	22.53	2.21	0.00
Math SAT	1754	-0.05	1.07	4089	-0.23	1.01	0.00
Reading SAT	1730	-0.02	1.04	4028	-0.20	0.99	0.00

The right-most column contains p-values for t-tests of the null hypothesis that the small and regular class averages are equal.

Table 2. Estimated Effect on Being in a Small Class on Kindergarten Mathematics and Reading Achievement (standard errors in parentheses)

	Math		Reading	
quantile				
.10	<b>.080</b> (.039)	<b>.100</b> (.032)	<b>.087</b> (.030)	<b>.114</b> (.029)
.25	<b>.093</b> (.028)	<b>.115</b> (.032)	<b>.127</b> (.026)	<b>.139</b> (.020)
.50	<b>.151</b> (.029)	<b>.170</b> (.027)	<b>.162</b> (.030)	<b>.182</b> (.027)
.75	<b>.205</b> (.044)	<b>.208</b> (.034)	<b>.211</b> (.043)	<b>.212</b> (.029)
.90	<b>.293</b> (.061)	<b>.248</b> (.055)	<b>.247</b> (.055)	<b>.294</b> (.057)
OLS	<b>.166</b> (.029)	<b>.189</b> (.026)	<b>.165</b> (.028)	<b>.191</b> (.026)
Observations	5843	5843	5761	5761
School f.e.	no	yes	no	yes

Table 3. Estimated Effect on Being in a Small Class on Kindergarten Mathematics and Reading Achievement, by Share of Students Eligible for Subsidized Lunch (standard errors in parentheses)

Share eligible for subsidized lunch	Math				Reading			
	low	high	low	high	low	high	low	high
1. kindergarten QTE estimates								
.10	<b>.089</b> (.045)	<b>.081</b> (.062)	<b>.059</b> (.046)	<b>.144</b> (.054)	<b>.039</b> (.038)	<b>.151</b> (.038)	<b>.098</b> (.041)	<b>.138</b> (.035)
.25	<b>.073</b> (.040)	<b>.107</b> (.038)	<b>.074</b> (.033)	<b>.126</b> (.035)	<b>.108</b> (.037)	<b>.143</b> (.036)	<b>.143</b> (.033)	<b>.128</b> (.032)
.50	<b>.120</b> (.045)	<b>.178</b> (.038)	<b>.174</b> (.049)	<b>.147</b> (.043)	<b>.136</b> (.036)	<b>.214</b> (.035)	<b>.190</b> (.034)	<b>.184</b> (.043)
.75	<b>.183</b> (.063)	<b>.254</b> (.062)	<b>.164</b> (.055)	<b>.295</b> (.048)	<b>.159</b> (.035)	<b>.231</b> (.053)	<b>.138</b> (.042)	<b>.255</b> (.034)
.90	<b>.259</b> (.084)	<b>.274</b> (.083)	<b>.169</b> (.067)	<b>.352</b> (.069)	<b>.216</b> (.083)	<b>.294</b> (.073)	<b>.245</b> (.083)	<b>.312</b> (.062)
OLS estimates	0.124 (0.039)	0.210 (0.042)	0.147 (0.036)	0.233 (0.038)	0.103 (0.040)	0.224 (0.039)	0.139 (0.038)	0.245 (0.037)
Observations	2960	2883	2960	2883	2909	2852	2909	2852
School f.e.	no	no	yes	yes	no	no	yes	yes

Table 4. Estimated Effect on Being in a Small Class on Kindergarten Mathematics and Reading Achievement, by Subsidized Lunch Eligibility (no school fixed effects; standard errors in parentheses)

Eligible for subsidized lunch	Kindergarten			
	Math		Reading	
	no	yes	no	Yes
.10	<b>.038</b> (.059)	<b>.116</b> (.052)	<b>.052</b> (.041)	<b>.110</b> (.045)
.25	<b>.088</b> (.043)	<b>.083</b> (.037)	<b>.145</b> (.040)	<b>.129</b> (.030)
.50	<b>.171</b> (.040)	<b>.158</b> (.042)	<b>.163</b> (.044)	<b>.192</b> (.039)
.75	<b>.228</b> (.066)	<b>.222</b> (.064)	<b>.237</b> (.060)	<b>.191</b> (.059)
.90	<b>.392</b> (.085)	<b>.204</b> (.075)	<b>.230</b> (.103)	<b>.212</b> (.070)
OLS	0.169 (0.040)	0.162 (0.041)	0.149 (0.042)	0.186 (0.036)
Observations	3028	2815	2980	2781

Table 5. Distribution (%) of Kindergarten Teacher Responses to the Following Statement: "The Level of Child Misbehavior In this School Interferes with my teaching", by Share of Students Eligible for a Free or Subsidized Lunch and the Date of the Survey (1,741 respondents to fall survey and 1,709 respondents to spring survey)

	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
share of students eligible for subsidized lunch					
<b>Below Median</b>					
fall survey	40.3	47.5	7.1	4.2	0.9
spring survey	40.6	47.2	7.1	4.2	1.0
<b>Above Median</b>					
fall survey	26.6	47.7	12.1	10.2	3.4
spring survey	26.5	47.4	12.2	10.4	3.4

Source: Early Childhood Longitudinal Survey-Kindergarten Cohort