Disruption, Achievement and the Heterogeneous Benefits of Smaller Classes

Graham McKee
Steven Rivkin
Katharine R. E. Sims

Department of Economics
Amherst College
Amherst, MA 01002-5000

February 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 and student achievement and illustrates how these underlying sources of variation may drive heterogeneity in the benefits of class size reductions. We test for results consistent with this model using the Tennessee STAR data. The estimates show that students in higher poverty schools and with greater learning aptitude realize larger benefits from smaller classes.

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 learning. Lazear (2001) provides one explanation for the observed pattern of class size estimates, attributing the benefits of smaller classes to 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 of education production with two potential dimensions of underlying heterogeneity and use data from the Tennessee STAR random assignment experiment to investigate the pattern of benefits from smaller classes. In the model, the returns to education are a function of both the time available for learning and the value of a unit of instructional time. Following Lazear (2001), we assume that the time available for learning depends on how frequently students disrupt class and that the probability of disrupting the class is on average higher for lower income children. Where there is a higher initial propensity to disrupt, class size reductions are expected to have a larger effect on the share of time available for learning, thereby leading to larger benefits of smaller classes in high poverty schools.

We expand on Lazear's model by assuming that the time available for learning also interacts with the value of a unit of instructional time. This value in turn depends on both class size and the academic skill of the student. Whether students with high or low baseline achievement can be expected to gain more from class size reduction therefore depends on the shape of the function linking academic skill and quality of instruction.

One possibility is that when class size is reduced, there is 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. This would bestow particular benefit to higher achieving students who might otherwise gain little from class time.

A second possibility is that the increased quality of instruction due to smaller class size may differentially benefit initially low or high achieving students through their own 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 because they 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.

Our model therefore illustrates that the heterogeneity in benefits from reductions in class size cannot be signed *a priori* when both the time available for instruction and the quality of instruction play a role in the education production function. Rather the pattern of heterogeneity is an empirical question, and we attempt to learn more about variation in the benefits of smaller classes by both student income and achievement using quantile treatment effect (QTE) methods and the Tennessee STAR data.

The results support the existence of multiple dimensions of heterogeneity in the return to smaller classes. First, consistent with the hypothesized model, both QTE and OLS estimates

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. Additional evidence from the Early Childhood Longitudinal Study (ECLS) that shows higher rates of disruptive behavior in higher poverty schools supports this interpretation. The fact that heterogeneity in benefits of class size reduction 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.

Second, the QTE estimates show a consistent pattern of larger class size effects for students at the upper end of the achievement distribution. This pattern holds regardless of whether the sample includes all students or is split by individual income or the share of low income students in a school. This heterogeneity in benefits 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 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 Tenessee STAR data, and the estimates of heterogeneity in benefits of class size based on this data. Section V brings additional evidence to bear on the causal channels from the ECLS, and Section VI summarizes the analysis and discusses policy implications.

II. Selected Research on Class Size and Heterogeneous Benefits

Since the release of the Coleman Report on equality of educational opportunity in 1966, a great deal of research on the effects of smaller classes has been completed. Much of this literature focuses on the average treatment effect but it has investigated heterogeneous effects along various dimensions.

The purposeful sorting of families into schools and allocation of resources among schools clearly complicates identification, and a number of recent papers use quasi-random variation in class size to identify the causal effects of smaller classes. Focusing largely on average treatment effects and potential differences by identifiable sub-groups, Angrist and Lavy (1999), Hoxby (2000), Jepsen and Rivkin (2008), and Rivkin, Hanushek, and Kain (2005) identify plausibly exogenous variation in class size (in Israel, Connecticut, California, and Texas, respectively). These studies have generally found that smaller classes significantly increase achievement in the early grades (grade 5 and below), and that effects tend to be larger for lower-income students.[1]

Babcock and Betts (2009) delves deeper in an effort to uncover the causal channels through which smaller classes raise achievement for lower income children. Using information from first grade report cards to classify students on the basis of both effort and academic achievement, the authors find that those graded poorly on effort realize larger benefits from class size reduction, while there appears to be little heterogeneity on the basis of achievement.[2] This finding is consistent with Lazear's emphasis on behavior rather than skill.

---

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

[2] Some concerns have been raised that the difference-in-differences framework used by Babcock and Betts 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.

There is also a growing body of non-experimental research that investigates the distribution of class size effects across the achievement distribution using quantile treatment effect (QTE) methods, and the findings of three recent papers reveal little evidence of benefits to smaller classes at any points in the distribution. Eide and Showalter (1998) find no significant benefit of smaller classes at any quantile using US data. In two papers using Dutch data, Levin (2001), and Ma and Koenker (2006) find little evidence of a significant relationship between class size and achievement across most of the distribution. Though these two papers both use two-stage quantile treatment effect methods to account for unobserved factors, they differ in terms of the parameterization of class size. Levin groups classes into six size categories, while Ma and Koenker analyze the benefits of marginal class size reductions.

The divergence between the QTE estimates and estimates of the average class size effects raise the concern that at least some of the non-experimental approaches do not account adequately for confounding factors. Because findings based on ideal random assignment experiments are not contaminated by confounding factors, a number of researchers have used the Tennesse STAR class size experiment to investigate the benefits of smaller classes. Word et al., (1990), Finn and Achilles (1990), Krueger (1999), Finn, Gerber, Achilles, and Boyd-Zaharias (2001), and Krueger and Whitmore (2001) all find that smaller class sizes have a significant and lasting impact on academic achievement and educational attainment. These studies also tend to find larger benefits for low income students and students of color.

Several studies have also used the STAR data to investigate whether the benefits of class size reduction vary along the achievement distribution. In contrast to the non-experimental QTE estimates, both Ding and Lehrer (2005) and Konstantopoulos (2008) find that the benefits of smaller classes appear to increase with achievement. Although differences in grade level and

other factors or problems with the STAR experiment may account for the divergent findings, the pattern of results is certainly consistent with the possibility that the non-experimental methods failed to account fully for confounding factors.

III. Conceptual Framework

The existing empirical research on class size effects is conducted largely in the absence of a theoretical framework, and we believe that such a framework provides valuable structure for an empirical analysis of the distribution of the returns to smaller classes. This section builds on Lazear (2001) and develops a model of the relationship between class size and the quality of instruction that incorporates two potential sources of heterogeneity: differences in the propensity to disrupt the class and academic achievement[3]. Following the description of the model we discuss potential channels that could cause heterogeneity along the achievement distribution.

Equation (1) models learning for student i in classroom c in school j where class size is given as n:

(1) $learning_{icj} = [\rho(d_{cj})^n][q(n,a)_{icj}] + X_{icj}$

In this equation, total learning is a function of the amount of classroom time available, the quality of class time in terms of the amount learned, and other student, community, and school factors (X).

The first term in brackets on the right hand side, $\rho(d_{cj})^n$, captures the amount of classroom time available for learning; this term comes almost directly from the model in Lazear (2001). Here the proportion of time a student is not disrupting the class is given as ρ, which is

---

[3] We refer to achievement rather than ability because it captures more broadly differences in aptitude, effort, and other factors that could lead to variation in the rate of knowledge acquisition, due to either student heterogeneity in the rate of learning or teacher decisions regarding the distribution of effort or the focus of the curriculum.

modeled as a function of d, the classroom average propensity to disrupt the class. With n students in the class, the share of class time not lost to disruption is thus $\rho(d_{cj})^n$.

The second term in brackets, $q(n,a)$, represents the quality of a unit of instructional time as a function of both class size (n) and the academic preparation (a) of the student. Although all students in the classroom experience the same amount of instructional time free from disruption, the value of that instructional time is likely to depend on the baseline academic preparation of the student. Variation among students could result from the level at which the curriculum is taught, distribution of teacher effort, and differences in student aptitude or effort.

In order to illustrate the ways in which classroom propensity for disruption *(d)* and academic preparation *(a)* 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 each of the two potential dimensions of heterogeneity, d (equation 2) and a (equation 3).

Equation (2) illustrates how the learning benefit of class size reduction varies with respect to the propensity to disrupt class. For simplicity, we drop the subscripts:[4]

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

The product of the two functions enclosed in squiggly brackets determines the sign of the cross-partial derivative of learning with respect to class size *(n)* and propensity for disruption *(d)*. The first is negative, as the derivative of ρ 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 respective terms. The expression to

---

[4] The derivative of learning with respect to class size (n) equals $[\rho(d)]^n \ln(\rho(d))q(n,a)+[\rho(d)]^n \dfrac{\partial q(n,a)}{\partial n}$.
Taking the derivative of this relationship with respect to d produces Equation (2) and with respect to a produces equation 3.

the right of the plus sign is assumed to be negative, because the quality of instructional time is

assumed to decrease as class size increases, (i.e. $\frac{\partial q(n,a)}{\partial n} > 0$), and class size, *n*, is positive. The

expression to the right of the plus sign is assumed to be positive, because the product of *ln(ρ)* and

n lies between 0 and -1, and q(*n,a*) is positive.[5]

The relationship between the benefits of smaller classes and the degree of disruption (*d*)

thus depends upon the magnitudes of two counteracting effects. First, as was emphasized by

Lazear (2001), at lower values of *ρ(d)*, i.e. higher propensity for disruption, 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 *ρ(d)*, any improvement in the quality of instruction time

due to smaller classes has a lower overall impact, because classrooms with lower values of *ρ(d)*

also have less time available 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. Therefore we may not necessarily expect to see greater

benefits of smaller class size where the propensity for disruption is higher.

To explore how the second dimension of heterogeneity may affect achievement gains

from smaller classes, equation (3) illustrates the cross-partial derivative of learning with respect

to class size (*n*) and achievement (*a*):

(3) $\qquad \frac{\partial^2 learning}{\partial n \partial a} = [\rho(d)]^n * \{[\ln(\rho(d)] + \frac{\partial^2 q(n,a)}{\partial n \partial a} \bigg/ \frac{\partial q(n,a)}{\partial a}\}$

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. Rather it depends on the sign of

_____

[5] At a value of ρ below 0.95, $n\ln(\rho)$ +1 becomes negative, but at such a low value of ρ the share of class time available for instruction is well below 50 percent.

the sum of the two functions enclosed in squiggly brackets. The first term $[\ln\rho(d)]$, which is roughly the average disruption of a single student, is negative. The second term, given by the ratio, $\dfrac{\partial^2 q(n,a)}{\partial n \partial a} \Big/ \dfrac{\partial q(n,a)}{\partial a}$ might be positive or negative depending on how the value of a unit of instructional time (q) varies with class size (n) and achievement (a). If the second term is negative the benefit of smaller classes would rise with initial level of achievement; we discuss this case below. First, however, we offer two possible explanations for why this ratio is likely to be positive, thereby leading to a theoretically ambiguous relationship between the benefit of smaller classes and initial student achievement that would depend on the relative magnitudes of offsetting terms.

Consider first how teacher behavior might lead to a negative ratio. The denominator $\dfrac{\partial q(n,a)}{\partial a}$ will be negative if it is the case that teachers tend to target the curriculum and their efforts towards less well-prepared students, so that the value of a unit of instructional time is decreasing in student achievement. This would follow if schools, particularly in an accountability era with a focus on the pass rate, target the curriculum toward those at or near the bottom of the distribution. The numerator, $\dfrac{\partial^2 q(n,a)}{\partial n \partial a}$, will also be negative if the emphasis on the lower portion of the achievement distribution increases with class size. Such an 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.

An alternative and not mutually exclusive explanation for a positive ratio can arise directly from student heterogeneity. Consider the possibility that higher achievers learn more per

11

unit of class time ($\frac{\partial q(n,a)}{\partial a}$>0) but that lower achievers realize a larger benefit to smaller classes

($\frac{\partial^2 q(n,a)}{\partial n \partial a} > 0$). A larger benefit to smaller classes for lower achievers might come from a

disproportionate increase in effort due to less day dreaming.

Finally, consider the case in which the ratio is negative and the cross partial derivative of

learning with respect to *n* and *a* would be unambiguously negative. This will occur if students

higher up in the distribution of achievement tend to learn more per unit of instructional time,

($\frac{\partial q(n,a)}{\partial a}$>0), and if 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). Systematic differences in

effort or skill that produce initial differences in achievement could lead to larger benefits for

higher achieving children.

In summary, 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

and illustrates that the pattern of any heterogeneous effects cannot be predicted *a priori*.

Therefore the relationship between student characteristics and the benefits of class size becomes

an empirical question. Importantly, differences in district, state, or federal policies which affect

how teachers target instructional time may lead to differences in the pattern of heterogeneity

along each of these dimensions. In addition, a number of causal channels including class size

induced changes in the behavior of students and of teachers could produce a given pattern of

heterogeneity. Following the discussion of the estimates we provide additional information from

the ECLS and Project STAR in order to learn more about the underlying causal mechanisms.

IV. Empirical Analysis

This section describes the Tennessee STAR data, the QTE specifications used to investigate the pattern of heterogeneity in the effects of smaller classes on achievement and then presents the results. 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 description 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 3rd grade. Seventy-nine schools participated in the experiment. These schools 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.

Previous research finds no significant difference between regular classes and regular classes with teacher assistants and we group students in the two types of large classes together.[6] 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

---

[6] 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.

later grades necessitated by student mobility or teacher preferences. 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.

Table 1 presents summary statistics for students in large and small classes. Unfortunately, there are no pre-experiment test scores, so we cannot conclusively determine that the kindergarten groups are similar with respect to baseline achievement. 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%) are non-white. Age is also very similar. 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 based on the test scores of all students, some of whom are excluded from the regressions due to missing data.

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 end of year achievement for student i is a function of an intercept, an indicator for class size status and an error term that captures student heterogeneity, differences in teacher quality among classrooms, and random noise.

(3) $A_i = \beta_0 + \beta_1 SMALL_i + \varepsilon_i$,

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 βhat from the OLS regression would provide a consistent estimate of the average treatment effect of class size reduction, and QTE estimates would provide consistent estimates of the distance between the small class and large class post-treatment achievement distributions at a given quantile. The corresponding QTE specification is:

(4) $A_i = \alpha_0^{(\tau)} + \alpha_1^{(\tau)} SMALL_i + \eta_i^{(\tau)}$

where the τ's represent quantiles of the achievement distribution. To capture the effects across the distribution, we estimate the above specification for each τ in $\{.1, .25, .5, .75, .9\}$. 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. However, we do include controls for student and teacher characteristics in all specifications order to assure that the results are not driven by differences across the samples in these characteristics.

The theoretical model presented above focused on the potential role of disruption in the learning function. To explore this idea empirically we use the share of students eligible for

subsidized lunch as a proxy for the level of disruptive behavior. 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.

However, 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. In addition, we recognize as Babcock 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 individual free lunch eligibility status in all specifications and also estimate separate class size effects by student eligibility for a subsidized lunch rather than the school low income share for comparison.

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 τth quantiles of the small class and large class distributions provides a valid estimate of the treatment effect at the τth quantile only if a student would be in the τth quantile regardless of treatment status. Satisfaction of this rank preservation assumption cannot be demonstrated

16

empirically, and therefore the results are most cautiously interpreted as reflecting differences in the marginal distributions.

IV.c. 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.[7] 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 indicators for teachers with fewer than three years of experience and teachers with an advanced degree, student demographic information including race-ethnicity, gender, age, special education status, whether or not the student is repeating kindergarten, attendance record, 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. Although the plots do not adjust for differences in demographic characteristics, the random assignment experimental design and observed balance between small and large classes suggests that the

---

[7] See Hao and Naiman (2007) and Koenker (2001) for further information on quantile regression.

inclusion of such controls should not lead to substantial changes in the achievement differences by class size. This is confirmed by the quantile regression estimates reported in Table 2.

Table 2 gives estimates of the average benefit of a small class and the magnitudes of the shifts in the distributions at the 10th, 25th, 50th, 75th, and 90th percentiles for specifications with and without school fixed effects. The estimates indicate 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 one moves up the 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 90th-10th and 75th-25th percentile differences in effect magnitudes differ only slightly whether or not school fixed effects are included. Focusing on the first and third column (without school fixed effects), the 90th-10th percentile difference in mathematics exceeds 0.21 standard deviations and the 75th-25th 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 share of low income students in a school 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. The results provide strong support for the hypothesis that the benefits of class size reduction are larger in high poverty schools. First, differences in the OLS estimates (lower part of Table 3) 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.

Second, 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 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. Therefore, the pattern of increasing magnitudes as achievement rises appears more pronounced for the sample of schools with a larger faction of students eligible for a subsidized lunch. This pattern also appears in Figure 2, which compares the kernel density plots of kindergarten achievement by class size and by school share of students eligible for a subsidized lunch. The shift of the achievement distribution to the right is generally larger for students in schools with a high share of subsidized lunch eligible students, particularly for mathematics.

Although the results in Table 3 and Figure 2 provide compelling 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 therefore reports class size estimates by individual subsidized lunch eligibility rather than by the share of eligible students in a school and includes specifications both with and without school fixed effects. The difference in OLS estimates by individual income is far smaller than the difference by the school share of low income students. This is true even if we include school fixed effects, as shown in Table 4. This provides additional support for the disruption interpretation, as it suggests that it is something about the learning environment 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 low and high income students, providing additional evidence that this finding holds across schools and student populations.

V. Additional Evidence on Causal Mechanisms

The findings reveal heterogeneous class size effects by both the school share of low income students and achievement, although to interpret these results as supporting specific causal channels would require additional evidence. In this section we provide evidence on the likely causal mechanisms that contribute to observed heterogeneity in class size effects. This evidence includes classroom observations on teacher and student behavior in a sub-sample of Project STAR classrooms and teacher reports on student misbehavior from the Early Childhood Longitudinal Survey-Kindergarten Cohort. The pattern of evidence generally supports 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. 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. Specifically, 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 of teaching 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.

VI. 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 shares of low income students 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 across schools, 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 UsersGuide*. 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/teacherachievement 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 of Kindergarten standardized achievement by class size and subject
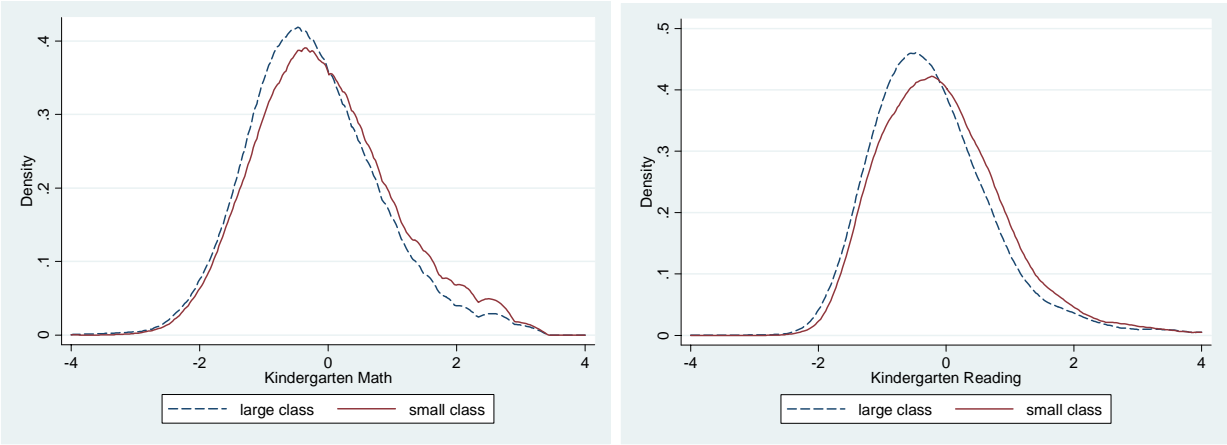
Figure 2. Kernel density of Kindergarten standardized achievement by share of students in a school eligible for a subsidized lunch, class size, and subject

High Share of Subsidized Lunch Eligible Students
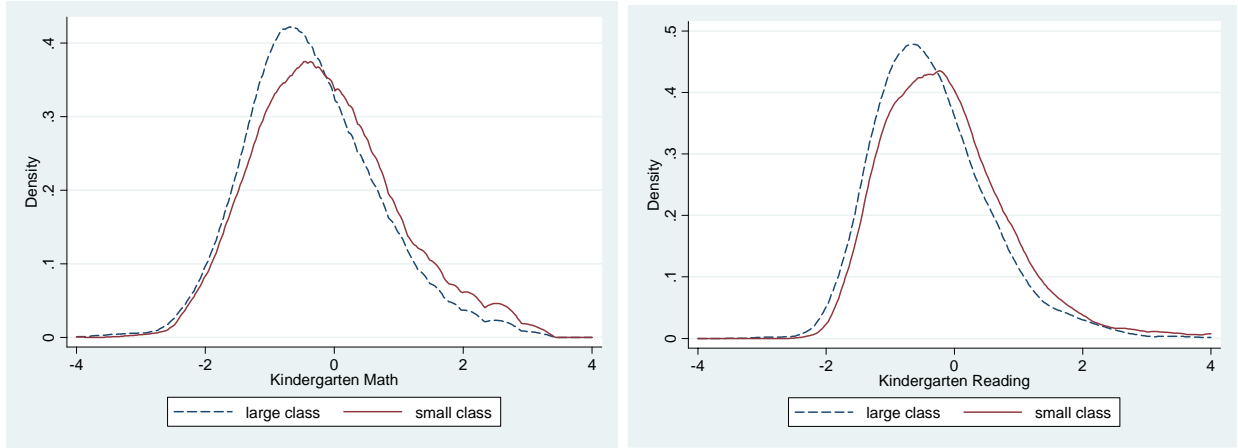
Low Share of Subsidized Lunch Eligible Students
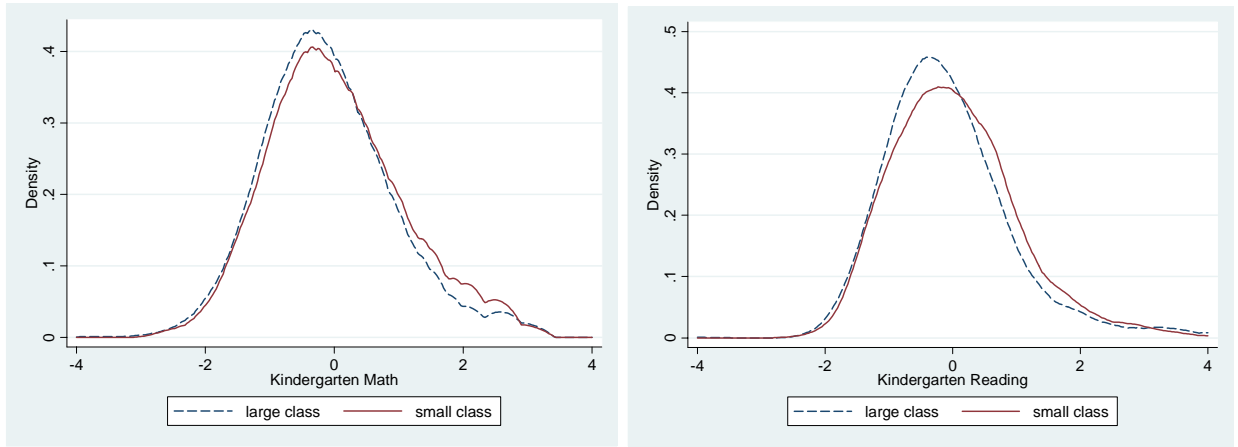
Table 1: Summary statistics by kindergarten class size

Demographic Indicators	Small classes			Regular classes			pr(|T|>t)
	Obs	Mean	SD	Obs	Mean	SD	
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 effects of small class on Kindergarten achievement

	Math		Reading	
QR estimates	**(1)**	**(2)**	**(3)**	**(4)**
.10	**.080**	**.100**	**.087**	**.114**
	(.039)	*(.032)*	*(.030)*	*(.029)*
.25	**.093**	**.115**	**.127**	**.139**
	(.028)	*(.032)*	*(.026)*	*(.020)*
.50	**.151**	**.170**	**.162**	**.182**
	(.029)	*(.027)*	*(.030)*	*(.027)*
.75	**.205**	**.208**	**.211**	**.212**
	(.044)	*(.034)*	*(.043)*	*(.029)*
.90	**.293**	**.248**	**.247**	**.294**
	(.061)	*(.055)*	*(.055)*	*(.057)*
OLS estimates	**.166**	**.189**	**.165**	**.191**
	(.029)	*(.026)*	*(.028)*	*(.026)*
Observations	5843	5843	5761	5761
School F.E.	no	yes	no	yes

Standard errors are given in parentheses. Quantile regression estimates are listed, followed by standard OLS estimates. All specifications include indicators for teachers with fewer than three years of experience and teachers with an advanced degree, and for the student's race-ethnicity, gender, age, special education status, whether or not they are repeating kindergarten, attendance record, and subsidized lunch eligibility. Specifications that do not include school fixed effects also include indicators for community type (suburban, rural, urban, and inner-city).

Table 3. Estimated effects of small class on Kindergarten achievement, by share of students in school eligible for subsidized lunch

	Math				Reading			
Share eligible:	low	high	low	high	low	high	low	high
QR estimates	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
.10	**.089**	**.081**	**.059**	**.144**	**.039**	**.151**	**.098**	**.138**
	(.045)	*(.062)*	*(.046)*	*(.054)*	*(.038)*	*(.038)*	*(.041)*	*(.035)*
.25	**.073**	**.107**	**.074**	**.126**	**.108**	**.143**	**.143**	**.128**
	(.040)	*(.038)*	*(.033)*	*(.035)*	*(.037)*	*(.036)*	*(.033)*	*(.032)*
.50	**.120**	**.178**	**.174**	**.147**	**.136**	**.214**	**.190**	**.184**
	(.045)	*(.038)*	*(.049)*	*(.043)*	*(.036)*	*(.035)*	*(.034)*	*(.043)*
.75	**.183**	**.254**	**.164**	**.295**	**.159**	**.231**	**.138**	**.255**
	(.063)	*(.062)*	*(.055)*	*(.048)*	*(.035)*	*(.053)*	*(.042)*	*(.034)*
.90	**.259**	**.274**	**.169**	**.352**	**.216**	**.294**	**.245**	**.312**
	(.084)	*(.083)*	*(.067)*	*(.069)*	*(.083)*	*(.073)*	*(.083)*	*(.062)*
OLS estimates	**0.124**	**0.210**	**0.147**	**0.233**	**0.103**	**0.224**	**0.139**	**0.245**
	(0.039)	(0.042)	(0.036)	(0.038)	(0.040)	(0.039)	(0.038)	(0.037)
Observations	2960	2883	2960	2883	2909	2852	2909	2852
School F.E.	no	no	yes	yes	no	no	yes	yes

Standard errors are given in parentheses. Quantile regression estimates are listed, followed by standard OLS estimates. All specifications include indicators for teachers with fewer than three years of experience and teachers with an advanced degree, and for the student's race-ethnicity, gender, age, special education status, whether or not they are repeating kindergarten, attendance record, and subsidized lunch eligibility. Specifications that do not include school fixed effects also include indicators for community type (suburban, rural, urban, and inner-city).

Table 4. Estimated effects of small class on Kindergarten achievement, by individual subsidized lunch eligibility

	Math				Reading			
Eligible for subsidized lunch	no	yes	no	yes	no	yes	no	yes
QR estimates	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
.10	**.038**	**.116**	**.006**	**.185**	**.052**	**.110**	**.132**	**.110**
	(.059)	(.052)	(.032)	(.052)	(.041)	(.045)	(.054)	(.028)
.25	**.088**	**.083**	**.082**	**.131**	**.145**	**.129**	**.144**	**.120**
	(.043)	(.037)	(.042)	(.046)	(.040)	(.030)	(.035)	(.034)
.50	**.171**	**.158**	**.167**	**.145**	**.163**	**.192**	**.153**	**.174**
	(.040)	(.042)	(.053)	(.052)	(.044)	(.039)	(.028)	(.048)
.75	**.228**	**.222**	**.270**	**.266**	**.237**	**.191**	**.151**	**.269**
	(.066)	(.064)	(.060)	(.062)	(.060)	(.059)	(.054)	(.041)
.90	**.392**	**.204**	**.299**	**.293**	**.230**	**.212**	**.237**	**.334**
	(.085)	(.075)	(.121)	(.063)	(.103)	(.070)	(.107)	(.057)
OLS estimates	**0.169**	**0.162**	**.173**	**.208**	**0.149**	**0.186**	**.161**	**.221**
	(0.040)	(0.041)	(.037)	(.038)	(0.042)	(0.036)	(.041)	(.033)
Observations	3028	2815	3028	2815	2980	2781	2980	2781
School F.E.	no	no	yes	yes	no	no	yes	yes

Standard errors are given in parentheses. Quantile regression estimates are listed, followed by standard OLS estimates. All specifications include indicators for teachers with fewer than three years of experience and teachers with an advanced degree, and for the student's race-ethnicity, gender, age, special education status, whether or not they are repeating kindergarten, attendance record, and subsidized lunch eligibility. Specifications that do not include school fixed effects also include indicators for community type (suburban, rural, urban, and inner-city).

Table 5. Teacher perceptions of disruption by share of students eligible for a subsidized lunch

	Low disruption	<--->			High Disruption
	"The level of child misbehavior in this school interferes with my teaching"				
	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
Low share eligible for subsidized lunch					
Fall survey	40.3	47.5	7.1	4.2	0.9
Spring survey	40.6	47.2	7.1	4.2	1.0
High share eligible for subsidized lunch					
Fall survey	26.6	47.7	12.1	10.2	3.4
Spring survey	26.5	47.4	12.2	10.4	3.4

Columns indicate the distribution (%) of Kindergarten teacher responses to the following statement: "The level of child misbehavior in this school interferes with my teaching." Classes with a low share eligible for subsidized lunch are those below the median share, classes with a high share are those above the median share.

Source: Early Childhood Longitudinal Survey. Kindergarten Cohort: 1,741 respondents to fall survey and 1,709 respondents to spring survey.