Empir Econ (2015) 48:1267–1286
DOI 10.1007/s00181-014-0810-1

Disruption, learning, and the heterogeneous benefits of smaller classes

Graham McKee · Katharine R. E. Sims · Steven G. Rivkin

Received: 24 February 2012 / Accepted: 19 February 2014 / Published online: 1 July 2014

Abstract Prior research suggests that the benefits from smaller classes may vary along multiple dimensions. In this paper we develop a flexible model of education production that incorporates the classroom-level time lost to disruption and the rate of learning during productive time as a function of teacher quality and individual propensity to acquire knowledge. We then investigate heterogeneity in class size effects by school poverty share, family income, teacher experience, and achievement percentile using data from Project STAR. We find that the benefits of small classes are consistently higher in schools with a larger low-income enrollment share. Conditional on school poverty share, we find little or no evidence that lower-income or lower-achieving students tend to realize larger benefits of smaller classes. Instead, we find that the return to smaller classes tends to increase with achievement regardless of school poverty share. Given the generally higher levels of disruption reported in higher poverty schools, this set of findings is consistent with, though not direct evidence of, the notion that reduced time lost to disruption is a primary mechanism through which smaller classes raise achievement and a compelling explanation for the empirical finding that class-size effects tend to be larger for lower-income children.

G. McKee
Department of Economics, Princeton University, Princeton, NJ, USA

K. R. E. Sims
Department of Economics, Amherst College, Amherst, MA, USA

S. G. Rivkin (✉)
Department of Economics, University of Illinois at Chicago, Chicago, IL, USA
e-mail: sgrivkin@uic.edu

S. G. Rivkin
National Bureau of Economic Research, Cambridge, MA, USA

Keywords Economics of education · Education · Class size · Heterogeneous benefits

JEL Classification I20 · I21 · I28 · H41

1 Introduction

Since the release of the Coleman Report on equality of educational opportunity (Coleman et al. 1966), there has been active debate on the magnitude and distribution of the benefits of smaller classes. The purposeful sorting of families into schools and the differential allocation of resources among schools makes it difficult to identify the causal relationship between smaller classes and student achievement. Not surprisingly, studies of the Tennessee STAR experiment, which randomly assigned students to small or large classes, provide much of the clearest evidence on class size effects. Word et al. (1990), Finn and Achilles (1990), Krueger (1999), Nye et al. (1999, 2000), Finn et al. (2001), Krueger and Whitmore (2001, 2002), Schanzenbach (2007), Chetty et al. (2011), and Dynarski et al. (2011) all find that smaller class sizes have a significant and lasting impact on academic achievement and educational attainment.[1]

Prior research on class size suggests several possible dimensions of heterogeneity in returns to smaller class size. First, returns may vary by family background characteristics such as income and race. Some analyses that estimate separate effects by income suggest that lower income students realize larger benefits from class size reduction. Schanzenbach (2007) summarizes results based on the STAR experimental data, reporting that the treatment effect is larger for low-income students and black students, the latter result being driven by a larger effect for all students in predominantly black schools. However, two recent papers, Hyunkuk et al. (2012) and Leuven et al. (2008), find no evidence of heterogeneity by family background. Second, returns may vary across the achievement distribution. Quantile treatment effect (QTE) estimates suggest that higher-achieving students may realize larger benefits. Research using the STAR experimental data by Ding and Lehrer (2011) and Konstantopoulos (2008) finds that the benefit of smaller classes increases with achievement. A study by Bandiera et al. (2010) also indicates that high achieving university students gain the most from smaller classes. However, studies by Eide and Showalter (1998), Levin (2001), and Ma and Koenker (2006), based on non-experimental data, find little evidence of a significant benefit of smaller classes anywhere in the achievement distribution.

Prior research also investigates heterogeneity by teacher quality based on characteristics including experience and salary. Both Mueller (2013) and Schanzenbach (2007) report generally larger benefits for more experienced and presumably higher quality teachers from analyses of the STAR data, though there is some inconsistency across

[1] There are also a number of papers that rely on quasi-random variation in class size to identify effects. Angrist and Lavy (1999), Bressoux et al. (2009), Fredriksson et al. (2013), Hoxby (2000), Jepsen and Rivkin (2009), Leuven et al. (2008), and Rivkin et al. (2005) use plausibly exogenous variation in class size in Israel, France, Sweden, Connecticut, California, Norway, and Texas, respectively. With the exception of Hoxby (2000) and Leuven et al. (2008), these studies generally support the finding that smaller classes significantly increase achievement in the early grades (grade 5 and below).

grades. In contrast, Woessmann (2005), Woessmann and West (2006), and Altinok and Kingdon (2012) find that the benefit from smaller classes is larger in countries with lower relative teacher salaries and presumably lower teacher quality. Although the experimental nature of the STAR data would suggest placing a higher weight on those studies, the lack of a consistent pattern across grades, the contrasting findings from the non-experimental studies, and the imprecision of the proxies for teacher quality suggest that there remains much uncertainty about the relationship between teacher quality and the benefit from smaller classes.

The purpose of this paper is to integrate multiple dimensions of heterogeneity into a single framework based on a model of education production that incorporates both the time lost to disruption and the rate of learning during productive time. We then use that framework as a basis for an empirical investigation of the distribution of the returns to smaller classes by school poverty share, family income, teacher quality, and achievement quantile using the Tennessee STAR data.

Most prior research has not been grounded in theoretical foundations that would motivate or explain the dimensions of heterogeneity. The main exception is Lazear (2001). Our conceptual framework extends the model of education production developed in Lazear (2001) by adding the possibility of variation in the impact of class size on knowledge acquisition during non-disrupted class time. Lazear (2001) illustrates that schools enrolling students with a higher propensity to disrupt realize a larger benefit from small classes through a greater reduction in the time lost to disruptive behavior. If disruption and poverty are positively correlated, then disruption would provide a plausible explanation for the tendency for research to find that low-income students realize larger benefits from smaller classes. However, Lazear provides no direct evidence of a disruption effect and largely ignores potential differences in the benefits of smaller classes conditional on time available for learning.

Moreover, the benefits of additional time available for learning are likely to vary among students and classrooms on the basis of student and teacher skills. In the case of students, one possibility is that higher achievers realize a larger benefit from improvements in the learning environment resulting from smaller classes if they tend to grasp new material more readily. Alternatively, those with lower achievement or lower income might realize larger benefits if there are diminishing returns with respect to home educational inputs or higher achievers tend to already know more of what is being taught in the classroom. Moreover, if effort is an important determinant of achievement, smaller classes might increase effort much more for initially lower achievers.[2] Finally, smaller classes may create additional opportunities to differentiate instruction or use innovative curricula (e.g., Rice 1999); in this case the relationship between achievement and class size effects could work in either direction. In the case of teachers, the ambiguity between skill and class size effect holds as well. Higher skill may

[2] In order to identify the channels through which smaller classes raise achievement for lower- income children, Babcock and Betts (2009) use information from report cards to classify baseline effort and academic achievement prior to a change in class size. The authors find that students who were previously graded with low effort realize larger benefits from class size reduction, while there appears to be little heterogeneity on the basis of achievement. To the extent that effort is negatively related to disruptive behavior, this finding is also consistent with Lazear's emphasis on disruption as a mechanism that contributes to heterogeneity in class-size effects.

raise the return to smaller classes if that return depends primarily on decisions regarding the reallocation of class time or changes in pedagogy. Alternatively, less skilled teachers may realize a larger return to smaller classes if deficiencies in classroom management or presentation become less damaging as class size decreases.

This ambiguity in the predicted variation of the benefits of smaller classes introduced by consideration of student and teacher skills contrasts with the clear prediction in the model of education production developed by Lazear (2001). The richer but less definitive conceptual framework which we develop in this paper highlights a number of factors that could introduce variation in the benefits of smaller classes. Of particular interest is the possibility for complex relationships between the return to smaller classes and student demographic characteristics resulting from the fact that students are both the recipients of teaching and other education inputs and themselves inputs into education production through their influence on the classroom environment. For instance, evidence in Schanzenbach (2007) that the higher benefit for black students is driven by a larger return for both black and nonblack students in high proportion black schools is supportive of the Lazear (2001) emphasis on classroom environment, while the larger return for higher achievers suggests the importance of student-level variation.

Our empirical analysis investigates heterogeneity by school poverty share, family income, teacher experience, and achievement quantile in an effort to learn more about the pattern of class size effects and the contributions of the aforementioned causal pathways. We find that the benefits of small classes are consistently higher in schools with a larger low-income enrollment share. Conditional on school poverty share, we find little or no evidence that lower-income or lower-achieving students tend to realize larger benefits of smaller classes than their schoolmates, even in high-poverty schools. Instead, QTE estimates find that the return to smaller classes tends to increase with achievement. Given the generally higher levels of disruption reported in higher poverty schools in the Early Childhood Longitudinal Survey (ECLS), this set of findings is consistent with, though not direct evidence of, the notion that reduced time lost to disruption is a primary mechanism through which smaller classes raise achievement and a compelling explanation for the empirical finding that class-size effects tend to be larger for lower-income children.

The remainder of the paper is structured as follows. Section 2 develops the conceptual model through which disruption and the efficacy of class time determine the benefits of class size reduction. Section 3 describes the Tennessee STAR data and empirical specifications, and Section 4 reports a series of class size estimates. Section 5 summarizes the analysis and discusses policy implications.

2 Conceptual framework

Based on the patterns of class size effects found in existing research, we expand on the work of Lazear (2001) to develop a model of the relationship between class size and learning that incorporates additional sources of heterogeneity. In this framework, variation in the return to smaller classes can come from the classroom time lost to disruption and the rate of learning during productive time as a function of teacher quality and individual propensity to acquire knowledge. Student heterogeneity may be driven by differences in effort, ability, or other individual or family factors.

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

$$\text{learning}_{icj} = \left[\rho(d_{cj})^{n_{cj}}\right][f(n_{cj}, a_{icj}, q_{cj})] + X_{icj}. \quad (1)$$

In this equation, total learning is a function of the fraction of classroom time available (first term in brackets), the rate of knowledge acquisition during uninterrupted time (second term in brackets), and other student, family, community, and school factors (X).

The first term in brackets, $\rho(d_{cj})^{n_{cj}}$, represents the share of classroom time available for learning, following Lazear (2001). The proportion of time a student is not disrupting the class is given as ρ, which is modeled as a function of d, the classroom average propensity to disrupt the class.[3]

The second term in brackets, $f(n_{cj}, a_{icj}, q_{cj})$, represents the rate of learning during non-disrupted time as a function of class size (n), student propensity to acquire knowledge (a), and quality of instruction (q). Although all students in the classroom may experience both the same amount of instructional time free from disruption and the same teacher, the value of that instructional time also depends on the propensity to gain knowledge. Therefore the benefits of smaller classes may vary both between and within classrooms and schools.

In order to illustrate the relationship between these three factors and the benefits from class size reduction, we take the derivative of Eq. 1 with respect to n and then again with respect to d (Eq. 2), a (Eq. 3), and q (Eq. 4). For simplicity, we drop the subscripts:

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

Equation (2) illustrates how the benefit of class size reduction varies with respect to the propensity to disrupt class. The product of the two functions enclosed in squiggly brackets determines the sign. The term in the first squiggly bracket 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 relationship in the second squiggly bracket is ambiguous and depends on the magnitudes of the respective terms. The expression to the right of the plus sign is negative, because the quality of instruction and propensity to acquire knowledge are assumed to decrease as class size increases, (i.e., $\frac{\partial f(n,a,q)}{\partial n} < 0$), and class size, n, is positive. The expression

[3] The classroom average propensity to disrupt is determined by teacher skill at classroom management as well as underlying student behavior. Differences in classroom management skill introduce another pathway through which teacher performance can influence the benefits of smaller classes.

to the left of the plus sign is positive, because the product of $ln(\rho)$ and n lies between 0 and -1 for a reasonable range of ρ, and $f(n, a, q)$ is positive.[4]

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 $\rho(d)$, i.e., higher propensity for disruption, a reduction in the power n has a larger effect on the share of time available for learning and thus a larger effect on achievement. Second, at lower values of $\rho(d)$, any improvements in teacher quality or the rate at which a student acquires knowledge resulting from the lower class size has a smaller impact on learning. This holds because students in classrooms with lower values of $\rho(d)$ have less instruction time free from disruption and less time to benefit from an increase in the efficacy of instruction. This effect, not considered by Lazear, may partially, fully, or even more than fully offset the higher benefit in more disruptive classrooms resulting from a larger decrease in the amount of time lost to disruption; it depends upon the magnitude of the effect of class size on the efficacy of learning time and how that effect varies across students in classrooms with different levels of disruption.[5] Note again that there may be variation within classrooms in the effect of class size on the rate of knowledge acquisition and thus the benefit of smaller classes.

Therefore the empirical evidence indicating that lower-income children realize larger benefits from smaller classes may not be driven by the relationship between poverty and disruption. Evidence shows that higher poverty schools experience greater difficulty attracting and retaining teachers, and lower-income children are disadvantaged on average with regard to family educational resources. These disadvantages could also elevate the return to class size for lower-income children if smaller classes mitigate deficiencies in teacher quality or the return to smaller classes diminishes with the level of family inputs.

Equation (3) describes the cross-partial derivative of learning with respect to class size and teacher quality, while Eq. (4) describes the cross-partial derivative of learning with respect to class size and the rate of knowledge acquisition:

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

$$\frac{\partial^2 \text{learning}}{\partial n \partial a} = [\rho(d)]^n * \left\{[\ln(\rho(d)] * \frac{\partial f(n, a, q)}{\partial a} + \frac{\partial^2 f(n, q, a)}{\partial n \partial a}\right\}. \quad (4)$$

4 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 %.

5 If the underlying propensity to disrupt is positively correlated with the poverty level, the elements that primarily determine the relationship between the poverty rate and benefit of smaller classes are $\rho(d)^{n-1}$, $[n\ln(\rho(d)) + 1]$, and $\frac{\partial \rho(d)}{\partial d}$. Because $\rho(d)^{n-1}$ and $[n\ln(\rho(d)) + 1]$ decrease in magnitude as d rises and ρ declines, these terms would tend to make the benefit of class size reduction increase at a decreasing rate as d and presumably the rate of poverty rise. Therefore in order for the increase in the benefit of class size reduction to be larger at higher levels of d the term $\frac{\partial \rho(d)}{\partial d}$ would have to be larger in magnitude. This may well be the case, particularly if high-poverty schools find it very difficult to attract and retain teachers and administrators and if they experience a severe lack of resources for support services.

The two equations possess parallel structures, meaning that much of the discussion of (3) pertains to (4) as well.

As is the case with disruption, the relationship between the benefit of class size reduction and teacher quality depends on the sign of the sum of the two functions enclosed in squiggly brackets. The first is negative, as the product of $[\ln \rho(d)]$ (roughly the average disruption of a single student) and $\frac{\partial f(n,q,a)}{\partial a}$ is negative. This product reflects the fact that higher quality teachers raise the value of an increase in time available for learning. The second function might be positive or negative depending on how the relationship between the efficacy of class time and class size varies with teacher quality. If higher quality teachers take better advantage of class size reduction, $\frac{\partial^2 f(n,q,a)}{\partial n \partial q}$ would be negative and the overall benefit of smaller classes would be *ceteris paribus* higher for students with higher teacher quality.

However, if class size reduction has a larger effect for students with a lower quality teacher, then $\frac{\partial^2 f(n,q,a)}{\partial n \partial q}$ would be positive. In this case the two terms in the squiggly brackets would be offsetting, and the relationship between the benefit of class size reduction and teacher quality would be ambiguous. This could occur if a smaller class mitigated deficiencies of less skilled teachers and closed the gap in teacher quality.

The parallel structures of Eqs. (3) and (4) indicate that a student who acquires knowledge at a faster rate may realize either a larger or a smaller benefit from reduced class size than a student who acquires knowledge at a slower rate. Again, the direction and strength of the relationship between the rate of knowledge acquisition and the class size effect depends upon whether smaller classes increase the rate of knowledge acquisition more for faster learners. If so, these students will unambiguously realize a larger benefit to smaller classes. If not, the two terms in the squiggly brackets would be offsetting: initially faster learners realize a larger benefit from the reduction in time lost to disruption, but initially slower learners realize a larger increase in the rate of knowledge acquisition during non-disrupted time.

In summary, this model of the link between learning and class size highlights time available for learning and the rate of learning during non-disrupted time as primary dimensions through which class size affects achievement. The introduction of the second dimension into the Lazear framework introduces ambiguity into the relationship between class size and a range of factors including income. Therefore the relationship between student, teacher, and school characteristics and the benefits of smaller classes becomes an empirical question, which we investigate in the next section. In order to provide additional information about the potential importance of specific causal mechanisms we reference information from the ECLS and Project STAR on disruption and teacher practices.

3 Data and empirical specifications

This section describes the Tennessee STAR data and the specifications used to investigate the pattern of heterogeneity in the effects of smaller classes on achievement.

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	1755	47.1 %	0.499	4091	48.6 %	0.500	0.29
Non-white	1755	31.6 %	0.465	4091	33.2 %	0.471	0.24
Age	1755	5.47	0.344	4091	5.46	0.344	0.28
Class size	1755	15.1	1.50	4091	22.5	2.21	0.00
Math SAT	1755	−0.052	1.07	4091	−0.225	1.01	0.00
Reading SAT	1731	−0.024	1.04	4030	−0.198	0.993	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

3.1 Data

Project STAR was legislated by the Tennessee government and carried out by researchers from 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). Further details of the Project STAR data are described in Finn and Achilles (1990) and Finn et al. (2007).

Previous research finds significant benefits to smaller classes but no significant difference between regular classes and regular classes with teacher aides, so we group students in the two types of large classes together. Following Ding and Lehrer (2011), 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 by student mobility or teacher preferences.[6] We transform the mathematics and reading achievement test scores to have means of zero and standard deviations of one based on the test scores of all students.

Table 1 presents summary statistics for students in large and small classes. As there are no pre-experiment test scores, we cannot examine whether the kindergarten groups are similar with respect to baseline achievement, but the two 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.

[6] Hanushek (1999) 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. Ding and Lehrer (2010) estimate dynamic treatment effects for STAR.

3.2 Empirical model

Equation (5) describes the basic estimation model. End of year achievement (A) for student i in school s is a function of an indicator for a small class (SMALL), a vector of student, teacher, and classroom characteristics (X), a school fixed effect (δ), and a random error (ε).

$$A_{is} = \beta_0 + \beta_1 \text{SMALL}_{is} + \beta_k X_{is} + \delta_s + \varepsilon_{is}. \quad (5)$$

All specifications include controls 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, and subsidized lunch eligibility indicators. Since the randomization occurred within schools, we include school fixed effects in the OLS specifications.

This model is estimated for all students combined and by school poverty share and family income in order to investigate potential sources of heterogeneity in class size effects. In order to allow for a flexible relationship between school poverty share and family income, we divide students into four categories based on share of students eligible for subsidized lunch (0–25; 25–50; 50– 75; and 75–100 %).[7] Estimates from OLS regressions by school poverty share, family income or both come from a single specification where all variables are interacted in the specified dimension.

We then use within-school variation in teacher experience as a proxy for differences in teacher effectiveness to investigate another potential dimension of variation in class size effects. In order to learn more about the potential relationship, we use three different definitions of teacher experience: 0 years of prior experience, fewer than 2 years of prior experience, and fewer than 3 years of prior experience. Unfortunately, the small number of inexperienced teachers precludes the estimation of separate specifications by poverty level and teacher experience.

Finally, in order to investigate differences by achievement, we estimate QTE specifications for all schools combined and by poverty share.[8] The QTE specification is:

$$A_{is} = \alpha_0^{(\tau)} + \alpha_1^{(\tau)} \text{SMALL}_{is} + \alpha_k^{(\tau)} X_{is} + \delta_s + \eta_{is}^{(\tau)}, \quad (6)$$

where the τ's represent quantiles of the achievement distribution. QTE estimates provide estimates of the distance between the small class and large class post-treatment achievement distributions at a given quantile. To capture the effects across the distribution, we estimate the above specification for each τ in {.1, .25, .5, .75, .9}. In a final set of specifications, we also interact class size with subsidized lunch eligibility and estimate QTE specifications that include those interactions in order to test for the presence of differences by individual income. The QTE specifications control for

7 The average share eligible for a subsidized lunch in each category are 11.9, 35.6, 59.2, and 90.8 %.

8 See Firpo (2007), Hao and Naiman (2007) and Koenker and Hallock (2001) for further information on quantile regression.

school effects using the two-stage method of Canay (2011) which assumes the school effects are constant across the achievement distribution (see table notes).

4 Results

This section reports estimates and hypothesis tests for the specifications described above. The results, along with supplementary evidence, provide information on the likely contributions of specific factors from the conceptual model.

The top panel of Table 2 reports estimates of the average benefit of a small class on mathematics and reading achievement for all students and by school poverty share, while the bottom panel allows the effect to differ by family income and school poverty share. Estimates in the first column show the well-known Project STAR findings of a large and highly significant class size effect and a larger effect for lower-income students. Note that the estimated effects for low-income students across all types of schools exceed those for all others by roughly 13 % in math (.203 vs .179) and 34 % (.222 vs .165) in reading.

By comparison, the differences in class size effects by school poverty share tend to be much larger in magnitude than the differences by family income. In the case of mathematics, the estimated effect of a small class equals 0.134 in schools where less than 25 % are eligible for a subsidized lunch, 0.144 in schools between 25 and 50 % low income, 0.163 in schools between 50 and 75 % low income, and a much higher 0.326 in schools with a poverty rate that exceeds 75 %. Comparing the highest poverty share to the lowest across the top panel we find an effect that more than doubles in size. The *p* values from tests of the hypotheses of equal effects with the 75–100 % poverty category reported in brackets below the standard errors support the notion that students in the highest poverty schools realize significantly larger benefits from smaller classes: the hypothesis that the effect in the highest poverty category equals the effect for math in another category is rejected at the 5 % level when the comparison is with the 0–25 or 25–50 % categories and at the 10 % level when the comparison is with the 50–75 % category. In contrast, there is no evidence of significant differences by poverty rate for schools with poverty shares below 75 %.

The estimates for reading present a similar though not quite as pronounced pattern. Gains are still largest for the highest poverty share group and the estimated benefit is smallest for the lowest poverty share group, but the magnitude in the 50–75 % category is smaller than that in the 25–50 % group. The estimated benefit is not significant at the 5 % level for the lowest poverty share group, while the estimates for the other three categories are significant at the 5 % level. In addition, the coefficient magnitude in the highest poverty category is roughly twice as large as the estimate in the lowest category, although the *p* values reported below the standard errors show that hypotheses of equal effects with the top poverty category are not rejected at any conventional level.

A potential explanation for this pattern of findings is that class time lost to disruptive behavior increases with the poverty rate and that the loss in class time adversely affects achievement. Evidence on the variation in disruptive behavior by school poverty and the effects of disruptive behavior on time available for learning would be needed to directly test this hypothesis. Although the STAR data do not contain information on

Table 2 Estimated effects of small class on math and reading achievement by school poverty share and student income

	All	School poverty share			
		0–25 %	25–50 %	50–75 %	75–100 %
All students					
Math	0.187	0.134	0.144	0.163	0.326
	(0.026)	(0.066)	(0.039)	(0.062)	(0.061)
		[0.032]	[0.012]	[0.062]	
Reading	0.189	0.130	0.186	0.143	0.258
	(0.026)	(0.074)	(0.041)	(0.058)	(0.054)
		[0.165]	[0.293]	[0.148]	
By student income					
Math					
Not low income	0.179	0.116	0.155	0.212	0.379
	(0.037)	(0.071)	(0.051)	(0.098)	(0.179)
		[0.173]	[0.229]	[0.413]	
Low income	0.203	0.052	0.123	0.123	0.325
	(0.038)	(0.161)	(0.062)	(0.081)	(0.065)
		[0.115]	[0.024]	[0.051]	
Reading					
Not low income	0.165	0.123	0.172	0.071	0.191
	(0.040)	(0.084)	(0.056)	(0.010)	(0.158)
		[0.703]	[0.912]	[0.520]	
Low income	0.222	0.025	0.208	0.191	0.276
	(0.033)	(0.146)	(0.055)	(0.071)	(0.057)
		[0.110]	[0.387]	[0.351]	
N (math/reading)					
All	5,846/5,761	983/955	2,502/2,474	1,053/1,041	1,308/1,291
Not low income	3,029/2,980	864/838	1,617/1,601	432/426	116/115
Low income	2,817/2,781	119/117	885/873	621/615	1,192/1,176

OLS regressions with controls for school fixed effects, teachers with fewer than 3 years of experience and teachers with an advanced degree, and for the student's race-ethnicity, gender, age, special education status, whether or not the student is repeating kindergarten, and subsidized lunch eligibility. Student income is measured by eligibility for subsidized lunch. The class size coefficients for math and reading in each panel come from separate regressions. For each panel, the estimates in the first column come from one regression for each subject, and the estimates from the remaining four columns come from a second regression for each subject in which all variables are fully interacted with poverty share (top panel) or poverty share and an indicator for individual student subsidized lunch eligibility (bottom two panels). Standard errors are in parentheses, and p values of tests of equality with the effect in the 75–100 % poverty category are in brackets. Hypothesis tests (not reported) show that none of the effects in the 0–25, 25–50, and 50–75 % poverty share categories are significantly different from one another

disruption, evidence from the nationally representative ECLS shows that disruptive behavior tends to increase with poverty. Table 3 reports the distribution of kindergarten teacher responses to the statement: "The level of child misbehavior in this school inter-

Table 3 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"				
Share eligible for subsidized lunch	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
0–25 %					
Fall survey	37.4	45.5	9.7	5.7	1.7
Spring survey	37.9	45.3	9.5	5.7	1.7
25–50 %					
Fall survey	30.8	44.9	11.8	9.0	3.6
Spring survey	30.7	44.8	12.1	8.9	3.6
50–75 %					
Fall survey	20.6	49.0	13.4	13.4	3.7
Spring survey	19.8	49.1	13.7	13.9	3.5
75–100 %					
Fall survey	20.1	54.9	10.8	10.7	3.5
Spring survey	20.4	54.8	9.8	11.3	3.8

Columns indicate the distribution (%) of Kindergarten teacher responses to the following statement: "The level of child misbehavior in this school interferes with my teaching." *Source* Early Childhood Longitudinal Survey. Kindergarten Cohort: 1,741 respondents to fall survey and 1,709 respondents to spring survey

feres 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 higher share of subsidized lunch eligible students are more likely to agree or strongly agree with the statement, although somewhat surprisingly the highest reported levels of agreement on average are in schools with a 50–75 % share of low-income students. Note that these responses come from the nationally representative ECLS-Kindergarten cohort, but it is likely that similar relationships would hold for Tennessee schools.

Disruption thus provides a possible explanation for the observed pattern of estimates, but the pattern is also consistent with the hypotheses that a lower quality of instruction, fewer family resources, or a lower level of achievement raise the benefit of smaller classes. In the case of the quality of instruction, prior evidence suggests that teacher turnover, share of teachers lacking full certification, and share of teachers with little or no experience tend to increase with the school poverty share. Consequently, a positive relationship between poverty share and the benefits of smaller classes could result from less effective teaching. In fact, less experienced teachers may realize a larger benefit from class-size reduction precisely because it mitigates their deficiencies in classroom management. If class size effects rise with school poverty and if inexperienced teachers realize larger benefits from smaller classes, then lower quality instruction would provide a plausible explanation for the observed class size effect–poverty relationship. In the case of student income and achievement, if students with lower income or lower achievement realize larger benefits to class-size reduction than

Table 4 Estimated effects of small class by teacher inexperience

Definition of inexperienced teachers	0 years prior experience	0 or 1 year prior experience	0–2 years prior experience
Math			
Small class	0.16	0.18	0.21
	(0.03)	(0.03)	(0.03)
Small class * inexperienced teacher	0.48	0.08	−0.14
	(0.15)	(0.09)	(0.08)
Reading			
Small class	0.17	0.17	0.20
	(0.03)	(0.03)	(0.03)
Small class * inexperienced teacher	0.63	0.14	−0.05
	(0.17)	(0.10)	(0.08)
N (math/reading)	5,846/5,761	5,846/5,761	5,846/5,761

OLS regressions with controls for school fixed effects, teachers with an advanced degree, student's race-ethnicity, gender, age, special education status, whether or not the student is repeating kindergarten, and subsidized lunch eligibility. Student income is measured by eligibility for subsidized lunch. Standard errors in parentheses

their higher-income or higher-achieving schoolmates, then factors related to lower levels of academic preparation and family resources could provide plausible explanations for the positive relationship between the poverty rate and benefit from smaller classes. However, in the absence of such significant differences by student or teacher characteristics the empirical findings would provide support, albeit not direct evidence, in favor of the focus on disruptive behavior.

We focus now on within-school variation in student income and achievement and teacher experience in order to learn more about the sources of the relationship between the magnitude of class–size effects and school poverty. Consider first the possibility that low family income increases the benefit of smaller classes. The bottom panel of Table 2 reports estimates by subsidized lunch eligibility and school poverty share. In the case of mathematics, the return to class size is smaller for low-income (subsidized lunch eligible) students in all school poverty share categories compared to not low-income students. In the case of reading, estimated benefits are less for low-income students in low poverty share schools but higher for all other schools. One possible explanation is that schools play a relatively more important role in reading instruction for lower-income children than for middle- or upper- income children who may have more reading support outside of school, and this may offset factors that elevate the benefit of smaller classes for not low-income students in mathematics. Note, however, that the hypothesis of equal effects within school poverty share categories is not rejected in any category for either subject (not reported).

We turn now to the possibility that a lower quality of instruction, as measured by the share of teachers with little or no prior experience, raises the benefit of smaller classes. The interaction terms between the small class and inexperience indictors reported in Table 4 show that students in classes with first-year teachers but not teachers with one or two years of prior experience appear to realize a significant additional benefit

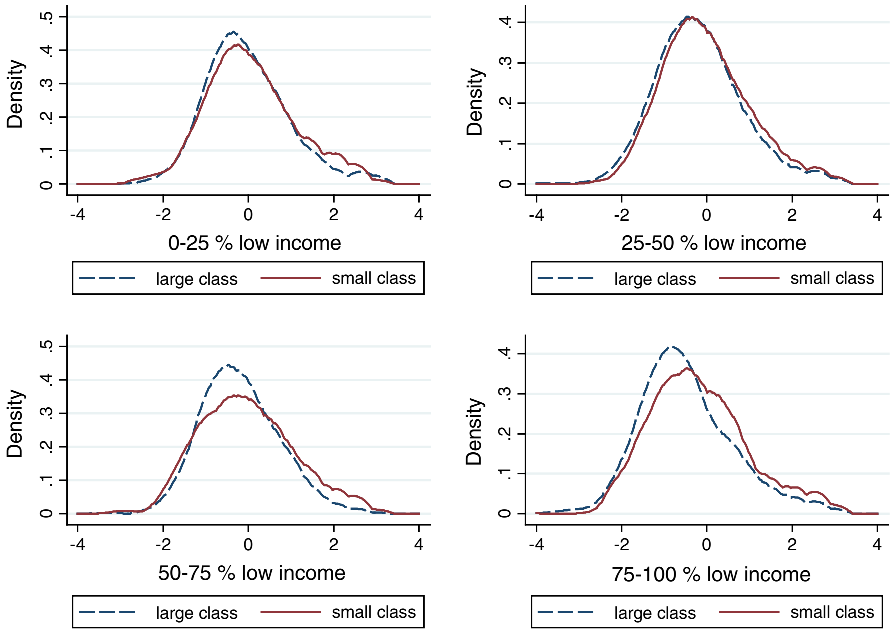

Fig. 1 Kernel density of kindergarten standardized achievement in math by class size and school poverty share

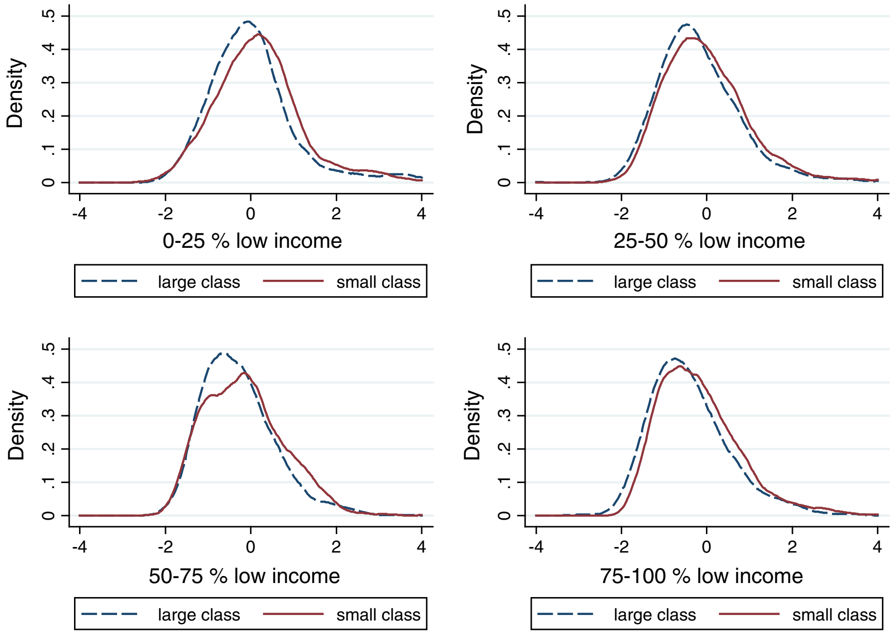

Fig. 2 Kernel density of kindergarten standardized achievement in reading by class size and school poverty share

Table 5 QTE estimated effects of small class on math and reading achievement by school poverty share

	All	School poverty share			
		0–25 %	25–50 %	50–75 %	75–100 %
Math					
Achievement quantile					
10	0.11	0.11	0.06	0.17	0.19
	(0.04)	(0.11)	(0.06)	(0.08)	(0.10)
25	0.14	0.16	0.11	0.13	0.33
	(0.03)	(0.08)	(0.05)	(0.07)	(0.08)
50	0.18	0.15	0.15	0.13	0.24
	(0.03)	(0.06)	(0.06)	(0.08)	(0.08)
75	0.23	0.19	0.18	0.25	0.30
	(0.04)	(0.10)	(0.07)	(0.11)	(0.09)
90	0.31	0.33	0.20	0.21	0.46
	(0.07)	(0.16)	(0.08)	(0.13)	(0.11)
Reading					
Achievement quantile					
10	0.11	0.11	0.11	0.15	0.15
	(0.03)	(0.11)	(0.05)	(0.07)	(0.06)
25	0.15	0.07	0.16	0.13	0.13
	(0.03)	(0.07)	(0.04)	(0.06)	(0.07)
50	0.18	0.22	0.21	0.07	0.22
	(0.03)	(0.07)	(0.04)	(0.07)	(0.06)
75	0.21	0.25	0.14	0.14	0.36
	(0.04)	(0.09)	(0.06)	(0.08)	(0.08)
90	0.29	0.22	0.32	0.19	0.34
	(0.06)	(0.16)	(0.09)	(0.11)	(0.12)
N (math/reading)	5,846/5,761	983/955	2,502/2,474	1,053/1,041	1,308/1,291

QTE regressions with controls for teachers with fewer than 3 years of experience and teachers with an advanced degree, and for the student's race-ethnicity, gender, age, special education status, whether or not the student is repeating kindergarten, and subsidized lunch eligibility. Student income is measured by eligibility for subsidized lunch. The five columns report estimates from separate regressions for each poverty share sample. Standard errors are in parentheses. School effects are controlled for using the two-stage method of Canay (2011), which assumes the school effects are constant across the achievement distribution. In the first stage, school effects are estimated via standard fixed effects regression. In the second stage, test scores are adjusted by subtracting the estimated school effects from the first stage, and the QTE regressions are run using the adjusted test scores

from class-size reduction. Given that the fraction of teachers with no prior experience tends to increase with the poverty share, this finding suggests that a lower quality of instruction contributes to the higher benefit to smaller classes in high-poverty schools.

Next we examine variation in class size effects by school poverty share and achievement. Given the crudeness of the subsidized lunch variable as a proxy for family background, this less parametric approach is likely to paint a richer picture of differences among children that are related to their academic preparation. Figures 1 and 2 display

Table 6 QTE estimated effects of free lunch status interacted with small class on math and reading achievement by school poverty share

	All	School poverty share			
		0–25 %	25–50 %	50–75 %	75–100 %
Math					
Achievement quantile					
10	0.07	0.10	0.00	0.04	0.27
	(0.10)	(0.37)	(0.13)	(0.19)	(0.31)
25	0.04	0.26	−0.04	0.02	0.02
	(0.06)	(0.24)	(0.09)	(0.14)	(0.31)
50	0.01	0.20	0.00	0.02	−0.30
	(0.06)	(0.21)	(0.11)	(0.17)	(0.31)
75	−0.02	−0.04	−0.03	−0.31	−0.04
	(0.09)	(0.24)	(0.14)	(0.21)	(0.22)
90	−0.10	−0.37	−0.26	−0.40	0.06
	(0.13)	(0.49)	(0.17)	(0.25)	(0.37)
Reading					
Achievement quantile					
10	0.00	−0.03	−0.13	0.09	0.18
	(0.07)	(0.28)	(0.11)	(0.17)	(0.22)
25	−0.03	−0.06	−0.04	0.13	−0.04
	(0.06)	(0.25)	(0.08)	(0.13)	(0.19)
50	0.01	−0.07	0.06	0.02	0.25
	(0.05)	(0.23)	(0.09)	(0.12)	(0.25)
75	0.10	0.10	0.14	−0.02	0.14
	(0.07)	(0.24)	(0.11)	(0.18)	(0.25)
90	0.00	0.04	−0.06	0.16	−0.04
	(0.11)	(0.45)	(0.16)	(0.24)	(0.39)
N (math/reading)	5,846/5,761	983/955	2,502/2,474	1,053/1,041	1,308/1,291

QTE regressions with controls for teachers with fewer than 3 years of experience and teachers with an advanced degree, and for the student's race-ethnicity, gender, age, special education status, whether or not the student is repeating kindergarten, and subsidized lunch eligibility. Student income is measured by eligibility for subsidized lunch. The five columns report estimates of subsidized lunch interaction terms from separate regressions for each poverty share sample. Standard errors are in parentheses. School effects are controlled for using the two-stage method of Canay (2011), which assumes the school effects are constant across the achievement distribution. In the first stage, school effects are estimated via standard fixed effects regression. In the second stage, test scores are adjusted by subtracting the estimated school effects from the first stage, and the QTE regressions are run using the adjusted test scores

mathematics and reading achievement kernel density estimates by class size for the four school poverty share categories, and Table 5 reports QTE estimates for the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution by category. The figures do not reveal a definitive picture, although the results for mathematics shown in Fig. 1 do suggest a pattern of larger class size effects at higher achievement levels in all four poverty categories.

The estimates for all schools combined reported in the first column of Table 5 are consistent with the pattern of effect size rising with achievement, from .11 to .31 in math and .11 to .29 in reading. When we divide schools into poverty categories, there is somewhat more variation. Nonetheless, the estimates for mathematics show a fairly consistent pattern of increasing class size effects as achievement rises, while the estimates for reading also present a general trend of increasing effects with achievement. For both reading and mathematics, the 90th percentile effect exceeds the 10th percentile effect by at least a factor of two for all but the 3rd poverty category. It is notable that students in the highest school poverty category tend to receive substantial benefit from smaller classes across the achievement distribution, although the effects on mathematics and reading appear to be substantially larger for students at the 75th and 90th percentiles of the respective achievement distributions.

Finally, we examine whether there are significant differences by income for students in the same part of the achievement distribution by including interactions between the small class and subsidized lunch indicators in the QTE specifications. Table 6 reports the coefficients on the interaction terms, and these reveal no consistent patterns or significant differences by income.

5 Conclusion

We construct a model of learning that expands upon the Lazear (2001) framework to permit class size to affect both the time available for learning and the rate of knowledge acquisition during that time and then investigate empirically the pattern of heterogeneity in the returns to smaller classes in an effort to learn more about the contributions of specific channels highlighted in the conceptual framework. The pattern of estimates may suggest that disruption affects the magnitude of the benefit of smaller classes and contributes to observed variation by family income.

The estimates reveal heterogeneity along three dimensions in the effect of smaller classes on both math and reading achievement for a sample of kindergarten students. First, the benefits of smaller classes are consistently higher in schools with a larger low-income enrollment share. Second, it appears that first-year teachers realize larger benefits to smaller classes. Finally, the return to smaller classes appears to increase with achievement. In contrast, there is little or no evidence showing that lower-income children realize larger benefits conditional on school poverty share or achievement.

Although we cannot provide direct evidence showing that higher levels of disruption increase the return to smaller classes, the set of empirical findings and supportive evidence is consistent with a role for disruption in the determination of the magnitude of the benefit of smaller classes. Several pieces of evidence support this explanation: (1) the generally positive relationship between the class-size effect and school poverty share for all children regardless of individual income; (2) the positive relationship between class-size effect and achievement given the negative relationship between achievement and poverty; (3) the ECLS evidence that disruption tends to increase with poverty; and (4) observational evidence from the STAR experiment that found little evidence that teacher behavior or pedagogy differed significantly by class size along any dimension including the extent of differentiated instruction

(Finn et al. 2010). Moreover, the finding that students in classes with first-year teachers realize a larger benefit is potentially consistent with disruption playing an important role given the absence of evidence that teachers modified practices in response to changes in class size. In fact, greater difficulty experienced by first-year teachers in managing student behavior could actually contribute to the higher levels of disruption in high-poverty schools.

Importantly, the findings provide mixed evidence on the likely impact of class size reduction on the achievement gap. Overall, smaller classes appear to increase achievement variance within schools by providing disproportionate benefit to the more successful students. Yet at the same time, the higher benefits to children in higher poverty schools contributes to the closing of the achievement gap across schools. The larger benefits for high achieving students in high-poverty schools suggests that class size reduction has a particularly strong effect on higher achievers whose family incomes may constrain schooling options. Perhaps more importantly, 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 might outweigh impacts on dispersion in the consideration of whether the benefits of class size reduction exceed the costs.

Finally, these experimental estimates almost certainly understate the benefit of class size reduction in higher poverty schools by ignoring any general equilibrium or externality effects. Class size related changes to the quality of teaching are potentially of particular importance. Decreases in the level of disruption not only expand the time available for learning but they also improve the quality of the work environment. Such improvements to working conditions would be expected to expand the supply of teachers and improve teacher quality over time, and the finding that inexperienced teachers realize larger benefits from smaller classes suggests that small classes are likely to be particularly appealing to new teachers. Future research should continue to explore the multiple dimensions of heterogeneous benefits, ideally with better classroom level data on teacher quality and time lost to disruption.

References

Altinok N, Kingdon G (2012) New evidence on class size effects: a pupil fixed effects approach. Oxf Bull Econ Stat 74(2):203–234

Angrist J, Lavy V (1999) Using Maimonides' rule to estimate the effect of class size on scholastic achievement. Q J Econ 114(2):533–576

Babcock P, Betts J (2009) Reduced-class distinctions: effort, ability, and the education production function. J Urban Econ 65(3):314–322

Bandiera O, Valentino L, Rasul I (2010) Heterogeneous class size effects: new evidence from a panel of university students. Econ J 120(549):1365–1398

Bressoux Pascal, Kramarz Francis, Prost Corrine (2009) Teachers' training, class size and students' outcomes: learning from administrative forecasting mistakes. Econ J 119(March):540–561

Canay I (2011) A simple approach to quantile regression for panel data. Econom J 14:368–386

Chetty R, Friedman JN, Hilger N, Saez E, Schanzenbach DW, Yagan D (2011) How does your kindergarten classroom affect your earnings? Evidence from Project STAR. Q J Econ 126(4):1593–1660

Cho H, Glewwe P, Whitler M (2012) Do reductions in class size raise students' test scores? Evidence from population variation in Minnesota's elementary schools. Econ Educ Rev 31(3):77–95

Coleman JS, Campbell EQ, Hobson CJ, McPartland J, Mood AM, Weinfeld FD, York RL (1966) Equality of educational opportunity. U.S. Government Printing Office, Washington, DC

Ding W, Lehrer S (2010) Estimating treatment effects from contaminated multiperiod education experiments: the dynamic impacts of class size reductions. Rev Econ Stat 92(1):31–42
Ding W, Lehrer SF (2011) Experimental estimates of the impacts of class size on test scores: robustness and heterogeneity. Educ Econ 19(3):229–252
Dynarski S, Hyman JM, Schanzenbach DW (2011) Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. NBER Working Paper 17533
Eide E, Showalter MH (1998) The effect of school quality on student performance: a quantile regression approach. Econ Lett 58:345–350
Finn JD, Achilles CA (1990) Answers and questions about class size: a statewide experiment. Am Educ Res J 27:557–577
Finn JD, Gerber SB, Achilles CM, Boyd-Zaharias J (2001) The enduring effects of small classes. Teach Coll Rec 103:145–183
Finn JD, Boyd-Zaharias J, Fish RM, Gerber SB (2007) Project STAR and beyond: database users guide. HEROS, Lebanon. http://www.laboratoriorevelli.it/corsi/emetrix/starUsersGuide.pdf. Accessed June 2014
Finn JD, Suriani AE, Achilles CM (2010) Small classes in the early grades: one policy, multiple outcomes In: Reynolds AJ, Rolnick AJ, Englund MM, Temple JA (eds) Childhood programs and practices in the first decade of life. Cambridge University Press, New York, pp 287–308
Firpo S (2007) Efficient semiparametric estimation of quantile treatment effects. Econometrica 75(1 (January)):259–276
Fredriksson P, Ockert B, Oosterbeek H (2013) Long-term effects of class size. Q J Econ 128(1):249–285
Hanushek E (1999) Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. Educ Eval Policy Anal 21(2 (Summer)):143–163
Hao L, Naiman D (2007) Quantile regression. SAGE, Thousand Oaks
Hoxby CM (2000) The effects of class size on student achievement: new evidence from population variation. Q J Econ 115(4):1239–1286
Jepsen C, Rivkin S (2009) Class size reduction and student achievement the potential tradeoff between teacher quality and class size. J Hum Resour 44(1):223–250
Koenker R, Hallock K (2001) Quantile regression. J Econ Perspect 15(4 (Fall)):143–156
Konstantopoulos S (2008) Do small classes reduce the achievement gap between low and high achievers, evidence from Project STAR? Elem Sch J 108:275–291
Krueger A (1999) Experimental estimates of education production functions. Q J Econ 114(2 (May)):497–532
Krueger A, Whitmore D (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. Econ J 111(January):1–28
Krueger A, Whitmore D (2002) Would smaller classes help close the black-white achievement gap? In: Chubb J, Loveless T (eds) Bridging the achievement gap. Brookings Institute Press, Washington, DC
Lazear E (2001) Educational production. Q J Econ 116(3 (August)):777–803
Leuven E, Oosterbeek H, Ronning M (2008) Quasi-experimental estimates of the effect of class size on achievement in Norway. Scand J Econ 110(4):663–693
Levin J (2001) For whom the reductions count: a quantile regression analysis of class size and peer effects on scholastic achievement. Empir Econ 26:221–246
Ma L, Koenker R (2006) Quantile regression methods for recursive structural equation models. J Econom 134(September):471–506
Mueller S (2013) Teacher experience and the class size effect—experimental evidence. J Public Econ 98(February):44–52
Nye B, Hedges LV, Konstantopoulos S (1999) The long-term effects of small classes: a five-year follow-up of the Tennessee class size experiment. Educ Eval Policy Anal 21:127–142
Nye B, Hedges LV, Konstantopoulos S (2000) Effects of small classes on academic achievement: the results of the Tennessee class size experiment. Am Educ Res J 37:123–151
Rice J (1999) The impact of class size on instructional strategies and the use of time in high school mathematics and science courses. Educ Eval Policy Anal 21(2 (Summer):):215–229
Rivkin SG, Hanushek EA, Kain JF (2005) Teachers, schools, and academic achievement. Econometrica 73(2):417–458
Schanzenbach DW (2007) What have researchers learned from Project STAR? Brook Papers Educ Policy 9:205–228
Woessmann L (2005) Educational production in Europe. Econ Policy 20(43):445–504

Woessmann L, West M (2006) Class-size effects in school systems around the world: evidence from between-grade variation in TIMSS. Eur Econ Rev 50:695–736

Word E, Johnston J, Bain HP, Fulton DB, Boyd-Zaharias J, Lintz MN, Achilles CM, Folger J, Breda C (1990) Final report: student/teacher achievement ratio (STAR): Tennessee's K-3 class-size study. Tennessee State Department of Education, Nashville. http://www.heros-inc.org