The Economics of Class Size

D W Schanzenbach, University of Chicago, Chicago, IL, USA

Introduction

Class-size reduction is a politically popular but relatively expensive education reform. Understanding the causal relationship between class size and student achievement is critical for determining whether class-size reduction can be recommended as a policy to improve student outcomes. We begin with a review of the theory of why class size might matter, followed by a discussion of the empirical strategies for identifying the causal impact of class size on student achievement. Next, the empirical literature on class-size reduction is reviewed, focusing on studies using experimental and quasi-experimental techniques because these rely on the most credible strategies for identifying the true causal relationship between class size and student achievement.

Why Class Size Might Matter

For most readers, class size and academic outcomes are probably intuitively linked. Nonetheless, it may be helpful to formalize the idea somewhat. Lazear (2001) puts forth a useful theory of educational production. In it, reducing class size decreases the amount of time that the classroom is disrupted, increasing time devoted to productive tasks. A simple summary of the model is as follows: a child is behaving in class at a given moment with probability p, and misbehaving with probability $(1 - p)$. In the model, misbehavior is broadly defined, ranging from talking or fighting, to behaviors such as asking questions that slow down the class or monopolizing the teacher's time. When there are n children in the classroom, p^n is the probability that the entire class is behaving and learning is taking place (assuming that p is independent across children). Assuming a constant disruption rate, having fewer students in the class means that learning is taking place in a larger fraction of time.

In the model, the impact of reducing class size depends not only on the size of the class, but also on the behavior of the students in it. As a result, the Lazear theory predicts that class-size effects should be larger for classes with more poorly behaved students. For example, on average p may decrease with age, so the impact of class-size reduction might be smaller for high school students than elementary school students. The impact of class-size reduction – all else equal – is predicted to be larger for groups with lower propensities to behave.

Empirical Approaches to Studying the Impact of Class Size

Economists typically model the relationship between student achievement and class size for student i in school j as

$$Y_{ij} = aS_{ij} + bF_{ij} + \varepsilon_{ij} \qquad [1]$$

where Y represents a measure of student achievement. S contains information on school-level inputs that impact achievement, such as class size, F contains family inputs, such as parental education, and ε is an error term. Both S and F measure inputs over the child's entire lifetime, and may contain inputs that are not observable to the econometrician. A negative coefficient on class size would suggest that student achievement declines as class size increases.

The problem with estimating eqn [1] (and similar versions of it) is that class size may be endogenous such that $E(\varepsilon \mid S,F) \neq 0$. For example, if students are assigned to small classes or better teachers in a compensatory manner – perhaps, because of low baseline test scores, or low levels of family inputs – but that information is not available to the researcher, the estimated impact of school resources will be biased. The most obvious such example is remedial or special-education courses, which tend to be small in size. Similarly, bias will result if parents who are more involved in their children's education are more likely to push for a smaller class or better teachers, and parental involvement is not measured in the dataset.

Due to these confounding factors, researchers have relied on strategies that use (plausibly) exogenous variation in class size in order to identify the causal impact of class size on student achievement. In other words, to identify the effect of class size, the variation in class size must come from factors that are more or less out of the control of decision makers such as parents and educators. The easiest example of this is where students are randomly assigned to classes of different size. The benefit of using a randomized experiment is that the treatment assignment is unrelated to any omitted characteristics. Such a design allows researchers to isolate the impact of the policy they are trying to test, without confounding factors such as parental pressure or compensating assignments. Thus, an experimental study typically offers more compelling evidence than a nonexperimental study, which simply observes the relationship between Y and S in the real world.

With a well-designed experimental assignment, a straightforward comparison of means by class type will provide an unbiased estimate of the impact of class size on achievement. In the case of (an idealized version of) a class-size experiment in which students are randomized within schools, the equation to be estimated might be as follows:

$$Y_{ics} = \beta_0 + \beta_1 SMALL_{cs} + X_{ics}\gamma + \alpha_s + \upsilon_{ics} \qquad [2]$$

where *SMALL* is an indicator variable for randomly assigned small-class treatment, and c indexes class c in school s. X is a vector of student-level characteristics. When treatments are randomized, student-level covariates are not related to class assignment and their inclusion should not change the estimated effect on class size, but should just contribute to the overall explanatory power of the model. A school-level fixed effect, α, is included, so that identification of small-class effects are identified off of within-school comparisons. Finally, the error term υ contains class-level and individual-level components, reflecting random differences in teacher and student quality.

Nonexperimental Research

There have been volumes of research looking at the relationship between class size and student performance in nonexperimental settings. These are well summarized in a pair of influential meta-analyses by Hanushek (1986, 1997). In them, he argues that the lion's share of the economics of education literature finds no consistent relationship between class size and student performance. In the 277 estimates from 59 published studies included in the 1997 paper, only 15% showed a positive, statistically significant impact of class size on student outcomes. A re-analysis by Krueger (2003) raised questions about the validity of the meta-analysis, and argued that the literature largely supports a positive impact of reduced class size. Interested readers are encouraged to refer to Krueger (2001) and Hanushek (2003) for an overview of the debate.

The usefulness of the Hanushek meta-analyses is limited by the underlying quality of the studies included in them. Most – but certainly not all – of the underlying studies relied only on observational variation and did not have a research design that would allow the estimation of the causal impact of class size on student achievement. This is problematic for several reasons. For one, within-school variation in class size (in the absence of a true experiment) is seldom random. If, for example, there is compensatory assignment to smaller classes, then the coefficient on class size will not only pick up the true effect of being in a smaller class but also any correlated omitted variables such as special-education status or poor prior achievement as discussed above (see also Boozer and Rouse, 2001). In addition, if there are only small differences in class size within school – for example, one classroom with 22 students and the other with 23 students – then one would need a large amount of data to precisely estimate the effect of such a small difference in class size. In general, it is inappropriate to base public policy on research that does not have a compelling identification strategy. As a result, much more weight should be placed on the experimental and quasi-experimental evidence outlined below.

Experimental Research

The most influential studies of class-size reduction are based on data from Project STAR, a large-scale randomized trial in the US state of Tennessee to test the impact of reducing class sizes in grades K-3. Mosteller (1995) described Project Steps to Achieving Resilience (STAR) as "one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools." In the experiment, students were randomly assigned within school to one of three treatment types: a small-size class (target of 13–17 students), a regular-size class (target of 22–25 students), or a regular-size class with a full-time teacher's aide. Importantly, teachers were also randomly assigned to class types. The experiment took place in 79 public schools across a variety of geographic locations (inner-city, suburban, and rural; predominantly low income, and middle class) for a single cohort of students in kindergarten through third grade in the years 1985–89. An eventual 11 600 students and 1330 teachers took part in the experiment. It is worth noting that the experiment made some students better off than they would have been, but of course did not otherwise increase class sizes beyond their normal range in the state at the time. Thus, no students were made worse off by being assigned to abnormally large classes.

In the ideal implementation of this experiment, students were to remain with the same randomly assigned class type from kindergarten through the end of third grade. In practice, though, there were several major sources of deviation from this model. Students who entered a participating school while the cohort was in first, second, or third grades were added to the experiment and randomly assigned to a class type. There were a substantial number of new entrants – 45% of eventual participants entered after kindergarten. A relatively large fraction of students also exited Project STAR schools (45% of overall participants), due to school moves, grade retention, or grade skipping, which also caused deviations from the original plan (see Krueger (1999) and Hanushek (1999) for further discussion). Fortunately, many of these exiting students are recaptured in the follow-up analysis described below, which includes all students in Tennessee in grades 4–8, and a nationwide match of college-entrance-exam takers when the cohort is around twelfth grade. In addition, in response to parental concerns about fairness to students, all students in regular and regular-aide classes were re-randomized across the two treatment groups in first grade. This deviation is less problematic for evaluation of the program, because it involved new randomization. In general, studies have found no difference in performance between regular and regular-aide classes, so these two groups are sometimes combined as the control group.

Finally, a smaller number of students (about 10% of participants) were moved from one class type to another in a nonrandom manner. It has been reported that most of these moves were due to student misbehavior, and were not typically the result of parental requests for moves to small classes (Krueger, 1999). Of all transitions, 25% were into small (more desirable) classes. This weakness of the experiment can be addressed through use of an intent-to-treat setup – that is, to use the variation caused by initial randomly assigned class type instead of the actual (possibly nonrandom) class type attended.

In practice, the nonrandom transitions and new entrants described above complicate the approach somewhat. Due to nonrandom transitions after initial assignment, it would be inappropriate to use current-year class type; instead, initial class-type assignment (the intent-to-treat measure) is typically used in studies using Project STAR data. That is, all impacts are measured with regard to the class that students were assigned to, and not the class that they actually attended. The intent-to-treat measure used in this case likely understates the impact of small classes by up to 15% (Krueger, 1999). Nonetheless, the conservative intent-to-treat measure based on random assignment is typically considered preferable to models which measure the impact of actual class type attended in cases in which there is nonrandom movement between classes. A simple example may help illustrate this: if a child were moved from a regular class to a small class because his parents insisted on the move, it is also reasonable to assume that the parents are especially active in other aspects of the student's education. For example, they may monitor homework closely or provide other education-enhancing opportunities. Researchers cannot control for these characteristics because they do not have perfect measures of the home environment. In the ideal case in which class type is randomly assigned, these home-environment measures are uncorrelated with class type and their impacts are absorbed in the error term in eqn [2]. When the effect of actual (nonrandom) class attended is measured instead, some of the impacts of the active home environment also may be picked up because actual attendance may be correlated with this home-environment component of the error term. Using the experimentally induced variation in this case means that not all students actually attend their assigned class type, and some students' test scores will count toward the regular-size class they were assigned to, even though they actually attended small classes. This approach circumvents the causation problem, but provides an understatement of the true impact. Krueger (1999) provides a more detailed discussion of this matter.

As described above, new entrants into the program were randomly assigned to class types. So, even though new entrants in first, second, and third grades on average are more disadvantaged than the kindergarten entrants, randomization allows us to compare new entrants in each grade to other new entrants in the same school across class types. In practice, then, the school-level fixed effect in eqn [2] is replaced with a fixed effect that combines school with a student's grade of entry (K, 1, 2, or 3) to the experiment, as this is the pool within which random assignment was determined (K representing kindergarten). In general, work on Project STAR has employed the following approach:

$$Y_{igs} = \beta_{0g} + \beta_{1g} SMALL_{is} + \beta_{2g} AIDE_{is} + \beta_{3g} X_{is} + \alpha_{sw} + \varepsilon_{igs} \quad [3]$$

Here g indexes the grade of the outcome measure. Both the *SMALL* and *AIDE* variables are measured as initial assignment, and not actual class attendance. The fixed effect varies by the randomization pool – school interacted with entry wave w. The coefficient on the control for classes with a teacher aide is sometimes omitted, as there appears to be no difference in outcomes between regular and regular-aide classes. As a result, the coefficient of interest measuring the small-class effect is similar whether or not aide classes are separately controlled. For precision, other student-level covariates such as gender, race, and free-lunch status are included in the vector X of control variables, but because of random assignment, including these controls does not change the magnitude of the small-class effect. The dependent variable is an outcome such as the mean math and reading score on the Stanford Achievement Test (SAT) for each grade, or whether a student took a college entrance exam.

Checks for Randomization

Due to the experimental design, impacts of reduced class size are straightforward to measure as the within-school (and entry wave) difference between class types, provided the randomization was done correctly. A compelling check of randomization is to examine a pretest to ensure that there are no measurable differences in the dependent variable between class types before the program begins. Unfortunately, no baseline test measure was collected in Project STAR. Another way to investigate whether randomization was done properly is to compare student characteristics that are related to student achievement but cannot be manipulated in response to treatment, such as student race, gender, and age. If there are no systematic differences in observable characteristics across class types, this provides support that the randomization was done properly. A similar check should be done on observable teacher characteristics.

Table 1 presents estimates of differences in nonmutable characteristics across initial treatment assignment. This is similar to tables presented in Krueger (1999) and Krueger and Whitmore (2001). The estimating equation is similar to eqn [3] above, with student or teacher characteristic on the left-hand side and indicators for small-class assignment and school-by-entry-wave fixed effects. Standard errors are clustered at the school level. Each table entry represents a separate regression, and only the coefficient and standard error on small-class assignment are reported. Since none of the coefficients are large or statistically significant, this is evidence that the randomization was done correctly, at least with regard to observable characteristics. The single exception is that the teacher having a master's degree or higher is marginally significant (p = 0.06). This means that teachers with more education were slightly less likely to be assigned to small classes. The results below are virtually unchanged if direct controls for teacher characteristics are included. Now that the randomization is validated, it is straightforward to turn to results of the experiment.

Achievement Results

Table 2 reports the impact of initial assignment to a small class on student test scores in grades K–3. Equation [3] is estimated, and each table entry reflects a separate regression. Test scores are normalized into z-scores based on the regular and regular-aide population. Average math and reading scores are reported in most cases, though if a student was missing a test score for one test but not both, the score for the non-missing test is used. The coefficient on the indicator variable for small class can be interpreted as the standard-deviation impact of the treatment. As many researchers have found (Word *et al.*, 1990; Krueger, 1999; Krueger and Whitmore, 2001), the table indicates that overall, students benefit about 0.15 standard deviations from assignment to a small class. When the results are disaggregated by race, it appears that black students benefited more from being assigned to a small class than the overall population, suggesting that reducing class size might be an effective strategy to reduce the black–white achievement gap. Krueger and Whitmore (2002) find that this result is largely driven by a larger treatment effect for all students regardless of race

Table 1 Testing whether covariates appear randomly assigned[a]

	(1)	(2)	(3)	(4)
Panel A: Student characteristics	Female = 1 0.000 (0.012)	White = 1 −0.002 (0.006)	Free Lunch = 1 −0.014 (0.011)	Age in 1985 (in years) −0.012 (0.011)
Panel B: Teacher characteristics	Female = 1 −0.001 (0.006)	White = 1 −0.001 (0.018)	Master's degree or higher = 1 −0.051 (0.027)	Total experience (in years) −0.155 (0.470)

[a]Each entry represents a separate regression. Only coefficients on initial assignment to small class are reported. Standard errors, in parentheses, are clustered by randomization pool. Other covariates include randomization-pool fixed effects.

Table 2 Small-class effects on test scores during the experiment[a]

	(1)	(2)	(3)	(4)
Panel A: Overall	Kindergarten 0.187 (0.039)	Grade 1 0.189 (0.035)	Grade 2 0.141 (0.034)	Grade 3 0.152 (0.030)
Panel B: Black students only	Kindergarten 0.214 (0.074)	Grade 1 0.249 (0.063)	Grade 2 0.207 (0.054)	Grade 3 0.242 (0.060)
Panel C: Free-lunch students only	Kindergarten 0.188 (0.046)	Grade 1 0.195 (0.042)	Grade 2 0.174 (0.041)	Grade 3 0.174 (0.039)

[a]Each entry represents a separate regression. Only coefficients on initial assignment to small class are reported. Standard errors are in parentheses, clustered by randomization pool. Other covariates include randomization-pool fixed effects and student demographic characteristics.

in predominantly black schools, suggesting that benefits from additional resources are higher in such schools. Benefits are also larger for students from low socioeconomic status families, measured by whether they receive free or reduced-price lunch.

In fourth grade, the class-size reduction experiment concluded and all students were returned to regular-sized classes. At the same time, the assessment test was changed from the SAT to the Comprehensive Test of Basic Skills (CTBS). Both tests are multiple-choice standardized tests that measure reading and math achievement, and are taken by students at the end of the school year. The CTBS results are scaled in the same manner as the SAT, in terms of standard deviation units. One important difference in the data is that all students in public schools statewide who had ever participated in Project STAR are included in the follow-up study, even if they had been retained a grade. It is estimated that 20% of students had been retained a grade by eighth grade, but this did not vary with initial class assignment. As a result, some students took the fourth-grade test in 1990, while others took it in later years or even took it more than once. In the analysis reported here, all scores from grade g – no matter what year a student was in that grade – are compared. In the event of multiple attempts at grade g's test, the first available score is used. As in **Table 2**, all estimates are conditional on school-by-entry wave fixed effects and only the coefficient on small class is reported.

Results for grades 4–8 are reported in **Table 3**. Overall, there is a persistent positive impact of small-class assignment that is statistically significant (or borderline significant) through eighth grade, as has been found in previous studies (e.g., Krueger and Whitmore, 2001). The magnitude of the gain is one-third to one-half the size that was observed while the students were in the experimental classes. When the results are disaggregated, though, the impact appears to remain stronger with black and free-lunch students than with more advantaged students. There is also some evidence that nonacademic outcomes such as the rates of criminal behavior and teen pregnancy are improved (Krueger and Whitmore, 2002).

Another potential measure of student achievement is whether these students take the SAT or the American College Test (ACT) college-entrance exam, which can be used as an early proxy for college attendance. In order to measure this, Project STAR student data were matched to the national databases of college-entry test records, as described in Krueger and Whitmore (2001, 2002). To examine whether assignment to a small class influences the college-entrance exam test-taking rate, a binary variable indicating that a college-entrance exam was taken is the dependent variable in eqn [3]. The impact of small-class assignment on college test taking is included as the final column in **Table 3**. Overall, test-taking rates increase by about 2 percentage points. Black students were 5 percentage points more likely to take the SAT or ACT if they were assigned to a small rather than regular-size class. On average, 38% of black students assigned to small classes took at least one of the college-entrance exams, compared with 33% in regular classes. Such a striking difference in test-taking rates between the small and regular class students could occur by chance less than one in 10 000 tries. Krueger and Whitmore (2002) interpret the magnitude of these effects by reference to the resulting reduction in the black–white test-taking gap. In regular classes, the black–white gap in taking a college entrance exam was 12.9 percentage points, compared to 5.1 percentage points for students in small classes. Thus, assigning all students to a small class is estimated to reduce the black–white gap in the test-taking rate by an impressive 60%. After controlling for increased selection into the test among small-class students, the impact on test scores for blacks is 0.15 standard deviations – about the same as the test-score impact in third grade.

Additional Caveats

An important limitation to the experiment was nonrandom movement across class-type assignment, as well as sample attrition during the treatment phase. As discussed briefly above (and at much greater length in the referenced works, especially Krueger, 1999), nonrandom movement can be addressed through using initial class-type

Table 3 Small-class effects on long-term test scores[a]

	Grade 4 (z-score) (1)	*Grade 5 (z-score) (2)*	*Grade 6 (z-score) (3)*	*Grade 7 (z-score) (4)*	*Grade 8 (z-score) (5)*	*Took college entrance test (1 = yes) (6)*
Panel A: Overall	0.035	0.048	0.060	0.040	0.036	0.024
	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)	(0.010)
Panel B: Black	0.078	0.080	0.105	0.066	0.063	0.050
students only	(0.048)	(0.043)	(0.045)	(0.042)	(0.046)	(0.018)
Panel C: Free-lunch	0.029	0.058	0.080	0.067	0.064	0.031
students only	(0.036)	(0.031)	(0.034)	(0.031)	(0.034)	(0.014)

[a]Each entry represents a separate regression. Only coefficients on initial assignment to small class are reported. Standard errors are in parentheses, clustered by randomization pool. Other covariates include randomization pool-fixed effects and student demographic characteristics.

assignment and not the actual (nonrandom) class-type attended. The attrition problem is significantly addressed through the statewide and nationwide matches used for the follow-up analyses.

Another concern often raised about the results of randomized experiments generally is that the measured effect may be driven by Hawthorne effects and might not be generalized to nonexperimental settings. That is, people participating in the experiment might act differently than they normally do because they know they are being studied. Although one cannot directly test for Hawthorne effects, Krueger (1999) attempts to shed light on the issue by investigating differences in achievement using the variation across only regular-sized classes, as there is little reason to think that Hawthorne effects would cause some classes in the treatment group to behave differently relative to other treatment-group classes. Class size in regular-sized classes ranged from 16 to 30 students, but the bulk of the distribution was between 20 and 26 students. Whether or not school effects are controlled for, students in a regular class with slightly fewer members out-scored larger regular classes. The estimated magnitude of a one-student reduction in class size was consistent with the magnitude of the experimental results (which estimates the impact of a seven-student reduction).

Finally, another concern is whether the findings of this experiment may be generalized to other settings. Along many measures, Tennessee in the mid-1980s looks reasonably similar to other places that might be interested in implementing a class-size reduction policy, so it would be reasonable to expect similar effects as those in the experiment. On the other hand, the Tennessee sample has lower levels of education inputs than the United States overall at the time, as measured by spending per student and education level of teachers. If adding resources has a greater impact when the baseline levels are already low, this might mean that schools with higher levels of spending could experience a smaller impact of class-size reduction. In addition, in order to be eligible for the experiment, schools were required to have a large enough enrolment to support three classrooms per grade. As a result, Project STAR schools were about 30% larger than average schools in Tennessee or across the United States. If larger schools are somehow differently effective with additional resources, then the findings in Project STAR may not be generalizable to smaller school settings (see Schanzenbach (2007) for further discussion of these points).

Quasi-Experimental Research

As true randomized experiments are rare, researchers must also look for quasi-experimental approaches that allow isolation of the causal impact of class-size reduction. One of the strengths of quasi-experimental approaches is that the participants are unaware that they are being studied, so Hawthone effects are unlikely.

The most famous quasi-experimental approach to studying class-size reduction comes from Angrist and Lavy's (1999) use of a strict maximum class-size rule in Israel and a regression discontinuity (RD) approach. In Israel, maximum class size is dictated by Maimonides' rule, which specifies that no more than 40 students shall be in one class. As a result, if the school's total enrolment in a grade is 40 students or fewer, there will only be one classroom with a class size equal to the total enrolment. If the enrolment increases from 40 to 41 students, though, a second class must be added, and the average class size declines precipitously from 40 students to 20.5 students. Similarly, if enrolment increases from 80 to 81 students, a school must move from two to three classes and the average class size falls from 40 students to 27 students.

Using the local variation around the enrolment sizes that are multiples of 40 students, Angrist and Lavy isolate the causal impact of class-size reduction. They find strong improvements overall in both math and reading scores, of a magnitude that is consistent with Project STAR's experimental results. Like in Project STAR, they also find larger improvements among disadvantaged students.

Urquiola (2006) uses a similar RD approach in Bolivia and finds that a one standard-deviation reduction in class size (about eight students in his data) improves test-score performance by 0.2 to 0.3 standard deviations. Browning and Heinesen (2007) also find similar results on data from Denmark, even though average class size is much smaller in their study (20 pupils per classroom, compared to 31 students in Angrist and Lavy's Israeli data). Urquiola and Verhoogen (2008) provide a cautionary tale about possible endogenous responses of schools to class-size caps, and show that in Chilean data, these endogenous responses of schools lead to violations of the assumptions of the RD design. In addition, caveats about the external validity of these studies are required as was the case with Project STAR.

Another quasi-experimental approach comes from Hoxby's (2000) study of class size in the US state of Connecticut. In the study, Hoxby isolates the variation in enrolment that comes from random fluctuations in cohort sizes across adjacent years. That is, taking away any preexisting trends in enrolment that might signify that a school district is waning or booming, the effect of class size is identified by variation in cohort size that reflects a temporary random shock in population size that may have been caused by an unusually small (or large) birth rate in a given year. Using this approach, Hoxby finds no positive effect of reduced class size, but has the statistical precision to rule out an effect as large as about one-fifth the size found in Project STAR. The discrepancy between these results and those of other well-identified experimental and quasi-experimental studies remains a puzzle.

Policy-Induced Variation

Another potentially promising approach to studying the effects of class-size reduction comes from sharp changes in policies regarding class size. The most famous recent example comes from the US state of California, where in 1996 a law was passed to give strong monetary incentives to schools to reduce class size in grades K–3 to 20 or fewer students. Unfortunately, from a research-design perspective, the take-up of the policy was nearly universal within a short period of time, so there was very little variation to exploit and evaluate its impact. In addition, test scores are only available starting in grade 4, so any evaluation of the policy was forced to use later test scores (instead of scores during the year that the reduced class size was experienced) as the outcome measure. Not surprisingly, the best evaluation of the policy found inconclusive results (Bohrnstedt and Stecher, 2002). It is unfortunate that this policy intervention, costing more than $1 billion per year, did not yield useful information about the impact of class-size reduction on student outcomes.

Discussion

The bulk of the research using credible identification of the impacts of class-size reduction suggests that reducing class size will significantly improve test scores. In addition, the benefits appear to be larger for disadvantaged groups such as African American students and children from families with low socioeconomic status. The long-term follow-up of the Project STAR class-size experiment finds that the gains appear to persist even after students are returned to regular-sized classes. The prior research has been less able to credibly isolate potential nonlinear effects of class size, which is an important consideration for policymakers considering a potential class-size reduction.

An important question that policymakers must ask prior to embarking on class-size reduction is whether the projected benefits outweigh the costs. A cost-benefit study of Project STAR found that the overall benefits outweighed the costs (Krueger and Whitmore, 2001). The answer in other cases will depend on the school's situation. What is the current level of educational inputs? Are there many disadvantaged students? Do we want to put extra weight on questions of equity – for example, the potential for small classes to reduce the black–white achievement gap? Is there a ready supply of qualified individuals available to meet the increased demand for classroom teachers? And, of course, what is the next best use of the available funds?

See also: Empirical Research Methods in the Economics of Education; Education Production Functions: Concepts; Education Production Functions: Evidence from Developed Countries; Education Production Functions: Evidence from Developing Countries.

Bibliography

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**(2), 533–575.
Bohrnstedt, G. W. and Stecher, B. M. (2002). What we have learned about class size reduction in California. *CSR Research Consortium Capstone Report.*
Boozer, M. and Rouse, C. (2001). Intraschool variation in class size: Patterns and implications. *Journal of Urban Economics* **50**(1), 163–189.
Browning, M. and Heinesen, E. (2007). Class size, teacher hours and educational attainment. *Scandinavian Journal of Economics* **109**(2), 415–438.
Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* **24**, 1141–1177.
Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis* **19**(2), 141–164.
Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis* **21**, 154–164.
Hanushek, E. A. (2003). The failure of inputs-based schooling policies. *The Economic Journal* **113**(485), F64–F98.
Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics* **115**(4), 1239–1285.
Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* **114**(2), 497–532.
Kureger, A. B. (2003). Economic considerations and class size. *The Economic Journal* **113**(485), F34–F63.
Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *Economic Journal* **111**, 1–28.
Krueger, A. B. and Whitmore, D. (2002). Would smaller classes help close the black–white achievement gap? In Chubb, J. E. and Loveless, T. (eds.) *Bridging the Achievement Gap*, pp 11–46. Washington, DC: Brookings Institution Press.
Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics* **116**, 777–803.
Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future of Children* 5(2), 113–127.
Schanzenbach, D. W. (2007). What have researchers learned from project STAR? *Brookings Papers on Education Policy* **2007**, 205–228.

Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia. *Review of Economics and Statistics* **88**(1), 171–177.

Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression–discontinuity design. *American Economic Review* **99** (1), 179–215.

Word, E., Johnston, J., Bain, H. P., *et al.* (1990). Student/teacher achievement ratio (STAR): Tennessee's K-3 class size study. *Final Summary Report 1985–1990.* Nashville, TN: Tennessee State Department of Education.

Further Reading

Achilles, C. M., Nye, B. A., Zaharias, J. B., and DeWayne Fulton, B. (1993). The lasting benefits study (LBS) in grades 4 and 5 (1990–1991): A legacy from Tennessee's four-year (K-3) class-size study (1985–1989), project STAR. *Research Paper*, HEROS.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., and Willms, J. D. (2001). Does class size matter? *Scientific American* **285**(5), 78–86.