



REVIEW OF *CLASS SIZE: WHAT RESEARCH SAYS AND WHAT IT MEANS FOR STATE POLICY*

Reviewed By

Diane Whitmore Schanzenbach

Northwestern University

June 2011

Summary of Review

Class Size: What Research Says and What It Means for State Policy argues that increasing average class size by one student will save about 2% of total education spending with negligible impact on academic achievement. It justifies this conclusion on the basis that Class-Size Reduction (CSR) is not particularly effective and is not as cost-effective as other reforms. However, this conclusion is based on a misleading review of the CSR research literature. The report puts too much emphasis on studies that are of poor quality or that do not focus on settings that are particularly relevant to the debate on class-size policy in the United States. It argues that class-size reduction is less cost-effective than other reform policies, but it bases this contention on an incomplete accounting of the benefits of smaller classes and an uncritical, unexamined list of alternative policies. The report's estimates of the potential cost savings are flawed as, in reality, schools cannot structurally reduce class size by only one student. Well-documented and long-term non-academic gains from CSR are not addressed. Likewise, the recommendation for releasing the "least effective" teachers assumes a valid way of making such determinations is available.

Kevin Welner

Editor

William Mathis

Managing Director

Erik Gunn

Managing Editor

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: 303-735-5290
Fax: 303-492-7090

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

Publishing Director: Alex Molnar



This is one of a series of Think Twice think tank reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at <http://greatlakescenter.org>.

REVIEW OF *CLASS SIZE: WHAT RESEARCH SAYS AND WHAT IT MEANS FOR STATE POLICY*

Diane Whitmore Schanzenbach, Northwestern University

I. Introduction

Class Size: What Research Says and What It Means for State Policy, authored by Grover J. Whitehurst and Matthew M. Chingos and published by the Brown Center on Education Policy at the Brookings Institution, estimates the potential cost savings of increasing average pupil-teacher ratio by one student to be \$12 billion per year.¹ To assess the likely learning impacts on students of such an increase, the study reviews a subset of research evidence on class size.

II. Findings and Conclusions of the Report

The report argues that during the current budget crisis, increasing class size can save substantial amounts of money with relatively few detrimental effects on students. Furthermore, it contends that if class-size increases were coupled with the laying off of the least-effective instructors, any potential negative impacts from larger classes could be offset by the resulting better instruction. The report concludes by adding that class-size reduction policies are most effective when they are targeted toward groups that benefit most, and when they are carefully weighed against alternate policy options.

III. The Report's Rationale for Its Findings and Conclusions

To assess the likely impacts on students of an increase in class size, the study reviews some research evidence on class-size reduction. In its literature review, the report concentrates on only those studies the authors deem to be “high enough quality,” and they summarize the range of impact estimates across these studies. The report characterizes the positive results from the well-known and highly regarded Tennessee STAR Project—which used an experimental design—as unusually large compared to the rest of the literature. The report also points to some prior studies that find no impact of class size on student achievement, suggesting to the authors that there may be little or no negative impact on students of moderate increases in class size.

IV. The Report's Use of Research Literature

The report's primary contribution is in summarizing what it deems to be high-quality studies on the impact of class-size reduction. The report (correctly, in my opinion) emphasizes that only

studies that have employed careful research methodologies are appropriate to consider. Because public policy is primarily interested in whether a policy can change outcomes—and not simply whether there is a non-causal, correlational relationship—it is vital to place primary emphasis on studies that can isolate the causal impact of class size. The report argues that the highest-quality studies are based on randomized experiments, natural experiments (also known as “quasi-experiments”), or studies based on sophisticated mathematical modeling. Unfortunately, as seen below, it does not uniformly apply these standards to its literature review, and as a result the report presents misleading results.

V. Review of the Report’s Methods

The study is primarily based on a review of the literature. As noted, the report reviews several studies that the authors consider “high quality,” and this approach is sensible. However, the report’s execution of this approach paints a picture of the class-size literature that is less conclusive than it in fact is. A recent review coauthored by MIT professor Joshua Angrist characterizes the class-size literature as being in substantial agreement about the magnitude of the benefits of smaller class sizes.² He argues that the high-quality studies “have produced estimates within a remarkably narrow band” and that “the weight of the evidence suggests that class-size reductions generate modest achievement gains” on the order of a 0.2 to 0.3 standard deviation increase for a 10-student reduction in class size. The Brookings report comes to a different conclusion in part because it places too much emphasis on some mixed or negative studies that do not actually measure up to the authors’ expressed high standards. In addition, the STAR results are mischaracterized as too large and as out of line with the literature. As described below, the STAR study and its results deserve substantially stronger weight than that given in the Brookings report.

Study reviews

The study notes that “the most influential and credible study of CSR is the Student Teacher Achievement Ratio” (STAR). In STAR, students and teachers in 79 Tennessee elementary schools were randomly assigned to small or regular-sized classes from 1985 to 1989. Such randomized experiments are generally characterized as the gold standard of social science research. This is because any positive gains in outcomes can be attributed with great confidence to being assigned to a smaller class. There have been many secondary studies using STAR data, and those studies have consistently found positive impacts not only on test scores but also on later life outcomes such as criminal behavior and college enrollment.³ The findings indicate that reducing class size from an average of 22 to an average of 15 improves average math and reading scores by about 0.20 standard deviations. Low-income students and African American students experience somewhat larger improvements from being assigned to a smaller class.

The Brookings report mischaracterizes the STAR effect sizes as unusually large relative to the literature. As the only estimates that come from a large-scale, well-executed true random experiment, though, these are the most trustworthy estimates in the literature. Thus, the STAR

results are the benchmark at this point in time. Research studies with less ideal research designs compare the magnitude of their findings to those from this randomized experiment.

Another highly regarded class-size study described in the Brookings report used quasi-experimental methods and comes from schools in Israel.⁴ Israel has a maximum class-size rule of 40 students. As a result, if there are only 40 students in a given grade at a given school, they can be in one single large class. If a 41st student enrolls, however, a second class must be added, pushing average class size across the two classes down to 20.5. Leveraging this sharp change in class size, the authors of the study find strong positive impacts of smaller class sizes, on the order of a 0.026 standard deviation improvement in test scores for each student reduced in a class. This implies that a 7-student reduction in class-size increases test scores by 0.18 standard deviations. Interestingly, the estimated impact in this study is very close to the one found in the STAR experiment.

The report omits results from Wisconsin's Student Achievement Guarantee in Education (SAGE) program, which reduced pupil-teacher ratios in high-poverty schools from between 21:1 and 25:1 to between 12:1 and 15:1. Molnar and coauthors evaluated the impact of the program by

The rest of the papers described in the study fail to meet the quality threshold set out by the authors.

comparing test score growth rates in SAGE schools to those in comparison schools that are closely matched in terms of students' demographic characteristics and prior test scores.⁵ The results are highly consistent with the results from STAR and the Israeli study. Overall, small class attendance improved student achievement by approximately 0.2 standard deviations. Also consistent with STAR, African American students received larger benefits from attending smaller classes.

In addition, the report cites a well-regarded Connecticut study that found no positive effect of reduced class size.⁶ This study took two approaches. The first relied upon the relatively modest variation in enrollment and class sizes that are driven by random fluctuations in cohort sizes across adjacent years and found no impact of smaller classes. The second took the same type of approach as the Israeli study described above and found positive but statistically insignificant impacts of class-size reduction.

The rest of the papers described in the study fail to meet the quality threshold set out by the authors. That is, in order to qualify as a high-quality study, the study must be based on variation in class size that is random or nearly random. An example of a problematic study would be one that simply correlates class size to achievement outcomes without regard to why some students are assigned to larger vs. smaller classes. For instance, sometimes low-achieving or special needs students are systematically assigned to smaller classes where they can receive more individualized attention. Likewise, advanced placement classes are often quite small. More affluent districts frequently have smaller class sizes than impoverished districts. In natural settings, there are any number of other variables that are not included in these analyses but which could have a strong impact on achievement.

For example, one of the studies showing positive impacts of class-size reduction uses a large dataset from Texas that allows the researchers to follow individual students over time.⁷ The paper overall is a highly regarded study published in a top economics journal. However, the paper's primary objective is to measure the impact of teachers on their students, not its subsection on the impact of class size, where the research approach was not well suited to address this question. (In the study, the class-size effect is measured by the difference in class size that a student experiences from one year to the next after accounting for stable and unchanging characteristics of each individual student.)

The problem with the Texas study's approach is that it only addresses one possible confounding variable. There may be important determinants of year-to-year variation in class size that are unobserved by the researchers but have a direct impact on student achievement. For instance, if a principal observes that a student under-performed relative to her potential in a given year, the principal may take that into account in the following year's class assignments. If—as we would naturally expect—any of the underlying factors that cause the assignment to smaller classes are also correlated with academic achievement, then the class-size effect will be mismeasured.

Another confounded study, an unpublished working paper by one of the Brookings report's authors, examines a recent policy change in Florida and is cited as evidence against CSR as a reform strategy.⁸ The study purports to investigate the impact of class-size reduction, but in fact

In its choice of studies to cite, the Brookings report puts too much faith in weaker studies and, thus, improperly represents what we do know about the effects of CSR.

the policy change is much more complicated. The policy awarded additional resources to *all* schools. Schools with large classes were required to use the resources to reduce class size, while schools that already had small classes were also given a sizeable financial bonus that they could use in any manner. In addition, as documented in a recent Think Tank Review, there were other policy changes in Florida that were implemented at the same time, such as mandatory grade retention.⁹ This type of study—even if executed perfectly—would not allow one to isolate or draw conclusions about the impact of class-size reduction.¹⁰

The Brookings report cites a Jepsen and Rivkin study, of California class-size reduction, that finds positive impacts of class-size reduction, but smaller in magnitude than those found in Tennessee and Israel.¹¹ Unfortunately, the California policy was introduced in a manner that made it impossible to credibly measure the impact of the policy. Test scores were not measured prior to the introduction of the policy, nor were they measured in the primary grades (where the reduction occurred). It should also be noted that an additional, thoughtful study of the California policy by Bohrnstedt and Stecher was not cited in the Brookings report, but argues that the data limitations are so severe that drawing any conclusions about the impact on student achievement is not justified.¹² Although Jepsen and Rivkin make a valiant effort to isolate the impact of the policy, in the end the results are confounded by class-size reductions in earlier, untested grades.

The Brookings study also includes a pair of papers that investigate class size in eighth grade and find mixed results.¹³ Although each of these middle school papers has strengths and weaknesses, ultimately they are less relevant in the current policy context that is focused on early grades. It is worth noting, though, that the Brookings study omits a Danish paper that finds class-size effects for ninth-grade students that are of the same magnitude as those found in STAR, Israel and SAGE.¹⁴

To be sure, the number of strong studies on class-size reduction is small. However, in its choice of studies to cite, the Brookings report puts too much faith in weaker studies and, thus, improperly represents what we do know about the effects of CSR.

Cost-benefit analysis

The Brookings report estimates that approximately 2% of total K-12 public education spending—\$12 billion per year—can be saved by increasing average class size across the United States by a single student. While the Brookings report does extrapolate costs down to the effect of one student, it does not directly address instructional losses. It only opines, “But if schools choose the least effective teachers to let go, then the effect of increased teacher quality could make up for some or all of the possible negative impact of increasing class size.”

A key practical problem with the cost-savings calculation, though, is that classes and teachers are not so easily divisible. Indeed, this is why research designs such as those based on the maximum class-size rule in Israel are so strong: when a school sees an enrollment increase from 40 to 41, it has to hire an entire additional teacher and cannot instead hire a fraction of one. As a result, average class-size declines sharply when a second teacher is hired (from 40 to $41/2$, or 20.5).

The same concept applies in reverse when one considers increasing class size. Imagine a K-5 school that has 100 students per grade spread across four classrooms in each grade. Currently, each of the 24 classrooms has a class size of 25 students. If the district were to reduce the teaching force in this school by one, the new average pupil-teacher ratio would be 26—characterized in the report as a relatively small increase in class size for potentially large budgetary savings. To obtain this, though, one cannot increase each class size by a single student. Each grade still has 100 students, and unless the school/district engages in some creative multi-grade classrooms or redistricting, most grades will still have 4 teachers and a class size of 25. To save one teacher, though, one grade would have to be reduced from 4 classrooms to 3, raising the average class size in that grade from 25 to 33.3. Among the children in this grade, the negative impacts would be striking. The school’s decision about which grade-level should be disadvantaged in this manner would certainly be complicated.

Unaddressed in the Brookings cost analysis is a growing body of research on early childhood interventions that shows that CSR may have a significant and positive impact on long-term outcomes, even when short-term test scores show more modest effects or rapid fadeout. Research on the Perry Preschool Project in particular suggests lower adult criminal behavior of participants, and a recent study of Head Start shows important gains in a variety of outcomes measured in young adulthood, such as high school graduation rates and health status.¹⁵

Evidence from follow-up studies of the STAR participants suggests that there are long-term benefits on a variety of outcomes that would not have been predicted from test score increases. If these potential benefits are ignored, then a simple cost-benefit analysis based on short-term and narrowly measured gains is inadequate.

In addition, the claims of comparative cost-benefit analyses fail to acknowledge the uncertainty surrounding such estimates.

VI. Review of the Findings and Conclusions

The report gives a misleading summary of the literature. It includes studies that show mixed or negative results that are of insufficient quality, and studies that investigate older students who are less likely to be subject to low-class-size policy initiatives.

The report also mischaracterizes the STAR findings as unusually large relative to other studies. For example, comparing STAR to the Israeli study, the report says that the Israeli results are “on the lower end of the range of those found in the STAR study.” In contrast Joshua Angrist, one of the authors of the Israeli study, states in his econometrics textbook that the impacts he finds are nearly identical to those in STAR.¹⁶

Further, the report presents misleading statistics on the potential cost savings of a one-student increase in class size. Because one cannot dismiss a fraction of a teacher but instead must dismiss a whole one, a one-student increase in average pupil-teacher ratio would likely result in very large class-size increases for some students.

Finally, after saying, “there is no research from the U.S. that directly compares CSR to specific alternative investments,” the report states that CSR is the “least cost-effective” policy based on a study that only accounts for a subset of the benefits of the policy. In making this claim, the report provides no evidence in support of potential alternative policies or whether any or all of those alternate policies could feasibly be implemented on a large scale.

VII. Usefulness of the Report for Guidance of Policy and Practice

The Brookings report is of limited use in policy debates about the role of class size on student achievement or as an effective guideline for financial savings. It provides a misleading characterization of the prior research literature, and it implies that class-size increases will have little impact on students. It also bases much of its argument on the impractical and questionable assumption that any reduction in the teacher workforce can be made on the basis of instructional quality instead of according to the terms of teachers’ current contracts. It does, however, make the important point that class-size reduction may be more effective for disadvantaged students and young students—and consequently that potential increases in class size would be particularly detrimental to these groups. Overall, the authors fail to make their case that increasing class size is either relatively harmless to school quality or a cost-effective way of saving money.

Notes and References

1 Whitehurst, G.J. & Chingos, M.M. (May 2011). *Class Size: What Research Says and What It Means for State Policy*. Brown Center on Education Policy at Brookings.

2 Angrist, J.D. & Pischke, J. (2010). “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics.” *Journal of Economic Perspectives*, 24(2): 3–30.

3 Krueger, A.B. (1999). “Experimental Estimates of Education Production Functions,” *Quarterly Journal of Economics* 115(2): 497-532.

See also

Krueger, A.B. & 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.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W. & Yagan, D (forthcoming). “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR.” *Quarterly Journal of Economics*.

4 Angrist, J.D. & Lavy, V. (1999). “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement,” *Quarterly Journal of Economics* 114(2): 533-575.

5 Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A. & Ehrle, K. (1999). “Evaluating the SAGE Program: A Pilot Program in Targeted Pupil-Teacher Reduction in Wisconsin,” *Educational Evaluation and Policy Analysis* 21(2): 165-77.

6 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.

7 Rivkin, S.G., Hanushek, E.A. & Kain, J.F. (2005). “Teachers, Schools, and Academic Achievement,” *Econometrica* 73(2): 417-458.

8 Chingos, M. (2010). *The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida’s Statewide Mandate*. Cambridge: Harvard University’s Program on Education Policy and Governance Working Paper 10-03.

See also

Finn, J. (2010, July). *Review of “The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida’s Statewide Mandate.”* Boulder, CO: National Education Policy Center. Retrieved June 14, 2011, from <http://nepc.colorado.edu/thinktank/review-class-size-florida>.

9 See Chatterji, M. (2010, Nov.). *Review of "Closing the Racial Achievement Gap."* Boulder, CO: National Education Policy Center. Retrieved June 14, 2011, from <http://nepc.colorado.edu/files/TTR-HeritageFla-Chatterji.pdf>

10 Other problems with this unpublished study include that the policy seems to "impact" background characteristics such as the racial composition of a school and its enrollment. There are also troubling divergent trends in test scores across the treatment and control groups prior to the introduction of the program, making the validity of the design additionally suspect.

11 Jepsen, C. & Rivkin, S. (2009). "Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size," *Journal of Human Resources* 44(1): 223-250.

12 Bohrnstedt, G.W. & Stecher, B.M. (Eds.) (2002). *What We Have Learned About Class Size Reduction*. Sacramento, CA: California Department of Education.

13 See Woessmann, L. & West, M. "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS," *European Economic Review* 50(3):695-736.

Dee, T.S. & West, M. (forthcoming). "The Non-Cognitive Returns to Class Size," *Education Evaluation and Policy Analysis*.

14 Heinesen, E. (2009). "Estimating Class-Size Effects Using Within-School Variation in Subject-Specific Classes," *The Economic Journal* 120: 737-760. The results of this study indicate that an 8-student reduction in language class sizes improve test scores by 0.18 to 0.24 standard deviations.

15 Belfield, C.R., Nores, M., Barnett, S. & Schweinhart, L. (2006). "The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup." *Journal of Human Resources* XLI:162-190.

Deming, D. (2009). "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1(3): 111-34.

16 Angrist, J.D. & Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press, 267.

DOCUMENT REVIEWED:

**Class Size: What Research Says and
What it Means for State Policy**

AUTHORS:

Grover J. Whitehurst and Matthew M. Chingos

PUBLISHER/THINK TANK:

Brookings Institution, Brown Center on
Education Policy

DOCUMENT RELEASE DATE:

May 2011

REVIEW DATE:

June 16, 2011

REVIEWER:

Diane Whitmore Schanzenbach, Northwestern
University

E-MAIL ADDRESS:

dws@northwestern.edu

PHONE NUMBER:

(847) 491-3884

SUGGESTED CITATION:

Schanzenbach, D.W. (2011). *Review of "Class Size: What Research Says and What It Means for State Policy."* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-class-size-brookings>.