NEPC Review: Small Class Sizes for Improving Student Achievement in Primary and Secondary Schools: A Systematic Review (Campbell Collaboration, October 2018)

Reviewed by:

Clive Belfield
Queens College, City University of New York

November 2023

National Education Policy Center

School of Education
University of Colorado Boulder
nepc.colorado.edu

Acknowledgements

NEPC Staff

Faith Boninger
Publications Manager

Patricia Hinchey
Academic Editor

Elaine Duggan
Production Design

Alex Molnar
NEPC Director

Kevin Welner
NEPC Director

Suggested Citation: Belfield, C. (2023). *NEPC review: Small class sizes for improving student achievement in primary and secondary schools: A systematic review*. Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/review/class-size-reduction

Funding: This review was made possible in part by funding from the Great Lakes Center for Educational Research and Practice.

The National Education Policy Center (NEPC), a university research center housed at the University of Colorado Boulder School of Education, sponsors research, produces policy briefs, and publishes expert third-party reviews of think tank reports. NEPC publications are written in accessible language and are intended for a broad audience that includes academic experts, policymakers, the media, and the general public. Our mission is to provide high-quality information in support of democratic deliberation about education policy. We are guided by the belief that the democratic governance of public education is strengthened when policies are based on sound evidence and support a multiracial society that is inclusive, kind, and just. Visit us at: http://nepc.colorado.edu

NEPC Review: Small Class Sizes for Improving Student Achievement in Primary and Secondary Schools: A Systematic Review (Campbell Collaboration, October 2018)

Reviewed by:

Clive Belfield

Queens College, City University of New York

November 2023

Summary

This systematic review of research literature estimates the impact of class size reduction on reading and math achievement, two of the many possible outcomes of reducing class size. The review surveyed all international research evidence up to 2017 and identified 10 research studies meeting its methodological standards. A weighted analysis of the 10 studies—seven of which are from the U.S. and four of which are studies of Tennessee's STAR—produced estimates of how smaller class size affects achievement. The review found a small positive, statistically significant impact on reading achievement and a small negative, but statistically insignificant, impact on math achievement. Beyond these 10 studies, the review provides a comprehensive catalog of research on class size reduction, and it provides a thorough assessment of each study's methods. However, the study takes an extremely narrow view of what research is valid for determining the impact of reducing class size. Notably, more than 90% of studies relevant to the research question were not included. In addition, while related research has appeared in many countries, only three countries are represented in the body of evidence. The result is a very limited analysis of an already narrowed question. Finally, repeated assertions that reducing class size is "costly" are unexplained and unsupported, potentially misleading policymakers about the reform's potential. These limitations mean the study offers policymakers extremely limited useful guidance on policy or practice.

NEPC Review: Small Class Sizes for Improving Student Achievement in Primary and Secondary Schools: A Systematic Review (Campbell Collaboration, October 2018)

Reviewed by:

Clive Belfield

Queens College, City University of New York

November 2023

I. Introduction

Class size reduction (CSR) is an easily understood, substantive reform that may improve education outcomes. CSR's multiple effects—on learning, classroom and school climate, and teacher attitudes among other factors—are potentially significant over multiple years. For more than four decades, researchers have worked to identify CSR's effects in a range of school settings and for various student groups, making the reform one of the most intensively studied in education research.[1]

A 2018 study from the Campbell Collaboration, *Small Class Sizes for Improving Student Achievement in Primary and Secondary Schools* by Trine Filges and co-authors, reviews available research in an attempt to definitively answer one question: Does smaller class size boost student achievement in reading and math?[2] From a body of international evidence published from 1980 to 2017, several studies on reading and math achievement were culled and analyzed. The review's goal was to determine precise achievement estimates, based only on studies meeting high methodological standards.

Leaders making decisions about education reform and funding are wise to review research exploring CSR's impact on outcomes—and also to investigate to what extent studies like this one can reliably inform their choices.

II. Findings and Conclusions of the Review

The review offers straightforward findings on achievement: It finds that CSR has only a modest positive impact on reading and no significant impact on math.

Although conclusions include an assertion that more research is needed to determine more broadly how CSR affects outcomes and classroom dynamics, overall conclusions appear negative:

> Class size reduction is costly and the available evidence points to no or only very small effect sizes of small classes in comparison to larger classes. Taking the individual variation in effects into consideration, we cannot rule out the possibility that small classes may be counterproductive for some students.

The report terms further research into achievement and class dynamics "crucial" in order to determine "where money is best allocated." Nevertheless, in linking findings on effect sizes to cost and suggesting that small class sizes may be "counterproductive," the report implies that investment in CSR is not currently the "best" use of money.

III. The Review's Rationale for its Findings and Conclusions

Findings are based on analysis and synthesis of research studies meeting the report's standards for quality. Results from selected studies were combined to produce weighted averages for achievement in math and reading. The conclusion that CSR has minimal impact on learning are based on calculations involving those statistical averages.

IV. The Review's Use of Research Literature

An initial rigorous survey of the literature identified 8,000 possible sources. Extremely strict inclusion criteria were applied in order to cull studies to serve as a base for calculations. Those criteria embed a major flaw in the study, as discussed both here and below.

First, the inclusion criteria yielded a stark result. Of 8,000 possible sources, 374 articles were assessed for eligibility and 127 were considered for inclusion. Ultimately, only 10 of the potential 127 studies were included to determine average achievement. Thus, 92% of research—the vast majority of which had been published in peer-reviewed journals—was rejected. Moreover, of 10 studies included, four refer to the same experiment—in Tennessee, in the 1980s.

In addition, the insistence that studies include test scores means that calculations reflect only one year's achievement, ignoring both other measures of learning and the possibility of lagging or latent effects. In addition, only studies with almost zero "risk of bias" were

included. While biased studies are indeed less valid, they can still be helpful if the goal is to see whether CSR improves achievement rather than by how much. And, only studies that reported sufficient information for the report's intended calculations are included.

By so severely narrowing its body of evidence when so much evidence is available, the review significantly limits its usefulness.

V. Discussion of the Review's Methods

While the methods applied in the review are straightforward and valid as well as clearly explained, the issues with inclusion criteria make it difficult to generalize findings and draw any meaningful policy conclusions about CSR.

First, in completely excluding some broadly accepted econometric methods—such as instrumental variables techniques or quasi-experiments—the methodology assumes that only high-quality randomized controlled trials provide useful information. These trials are expensive to conduct, take years to run and even longer to analyze. Adopting the review's approach would lead CSR research to effectively grind to a standstill.

In addition, the limited number of studies makes it difficult to generalize findings. For example, it is difficult to generalize findings from a single state—Tennessee, for example—to education systems in other states as diverse as Alaska and Massachusetts. Because only three countries are represented in the evidence base (the U.S., France, and the Netherlands), still less can findings be generalized internationally to countries with different education systems, testing regimes, and baseline class sizes. Moreover, the latest study included was from 2003, while most others came from the 1980s, although teaching pedagogies, instructional productivity, and/or resources have changed within the last 20 years.

There is also an issue related to how each of the 10 studies were weighted to determine averages. CSR research has been dominated by one large-scale randomized trial: Project STAR, which was conducted in Tennessee in the 1980s. STAR has been the subject of over 45 research papers, an amount of research scrutiny massively different from that for any other research project. Thus, the salience and significance of STAR evidence far exceeds that of any other study on CSR. At issue, therefore, is how much weight to give to STAR studies relative to others. Unfortunately, there is no obvious answer to this question. The weighting approach used in this analysis is plausible, but many other weights are equally plausible.

Finally, the limits imposed by the inclusion criteria make it impossible to conduct a meaningful meta-analysis as the review intends.[3] The purpose of a meta-analysis is to apply statistical analysis to a large body of evidence in order to identify common characteristics of subsets. For example, if CSR was found to have no effect in only 10 of 40 studies, meta-analysis might reveal that a common characteristic of those 10 was that they all focused on first grade; researchers might reasonably conclude the analysis showed CSR to be ineffective in first grade. However, only five of the 10 studies were used in the review's meta-analysis, all focusing on K-3. While the review notes that as a limitation, it fails to acknowledge that the

extremely small sample includes too little variation to allow for an informative meta-analysis.

VI. Review of the Validity of the Findings and Conclusions

It is important to note first that while achievement is an important consideration, it is far from the only one worth considering when evaluating CSR's potential, because implementation varies widely and because the strategy affects so many other outcomes.[4] That said, the overall finding—that achievement is very modestly increased as class size is reduced—is valid, given the selected evidence and research method applied. However, this finding falls far short as a guide to class size reduction policies because it ignores not only CSR's other effects but also the vast majority of research evidence. As a result, it would be invalid to use this study to claim that class size reduction is ineffective.

Further, the review emphasizes that CSR policies are "costly," which should be a significant factor as policymakers consider implementation. However, the review's definition of "costly" is unclear—every action involves some costs. Presumably, the review means "relatively high cost," but there is no indication of what other strategies it is being compared to in order to judge its cost substantively higher, and there is no indication or evidence of known costs provided. It is somewhat perplexing to see the review undertake extensive, methodologically sophisticated research on achievement impacts and yet provide no evidence or justification for its claims about cost. Hence, the economic claim cannot be validated and it is unclear why the review emphasizes it.

CSR is a complex, intensive and multifaceted reform. It can affect students, teachers, parents, and the school environment; and it can do so both immediately and over time.[5] There are economic consequences for each of these groups and it is the aggregate of these consequences that matters. One weak result on achievement does not necessarily mean that the overall consequences of CSR are zero. If the consequences are positive in total, they can be monetized as "benefits" and compared to the costs of implementing CSR. A cost-benefit analysis would be the most valid way to determine if CSR policies are worth implementing. Such an analysis is challenging to perform and the results are likely to depend on the specific CSR policy. Nevertheless, analyses by the Washington State Institute for Public Policy have found that the overall benefits of CSR do exceed the costs.[6]

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

As the review itself clearly states, there are important questions as to which students benefit from CSR, in what ways they benefit, and for how long. Such knowledge will greatly help in identifying the optimal conditions for implementation of CSR policies. But this review offers

only a narrow review of a narrow question, and so it offers no insight into the broader questions known to be important in policy design.

Other unanswered questions specific to CSR policy details are also important: Given the baseline class size, how much should it be reduced for desired benefits? Reducing a class of 40 to 39 might well have a different effect from reducing a class of 20 to 19, even though the reduction of one student is the same in both cases. The smaller baseline of 20 might allow the teacher to give substantively more attention to individual students, while not providing the same benefit in the larger class.[7] Or, reducing a class of 40 to 30 might prompt a teacher to make changes in pedagogy that a reduction of 40 to 39 would not. New evidence finds that the effects of CSR are not linear per child.[8] Finally, because the studies forming the base of the review primarily reflect large reductions in class size (approximately from 22 to 12), it offers no insight into more typical marginal changes (say, from 22 to 21, or from 32 to 31) that policymakers allow.

Overall, the review's value is minimal in that it provides a severely restricted answer to a very narrow question. Policymakers and practitioners, by contrast, need to take a much broader view and must make immediate decisions based on contingent or imperfect evidence.

Finally, perhaps of even greater concern than its narrowness is that in emphasizing CSR as "costly"—without offering analysis or evidence—the review has the potential to mislead policymakers. Although some recent media commentaries recognize the complexity of designing CSR policy based on existing research,[9] there are also examples of overreliance on this review's findings to make the unwarranted claim that CSR is an unproductive reform strategy.[10]

Policymakers should not make the same mistake: This study provides extremely limited guidance for policy or practice.

Notes and References

1 For a recent analysis with extensive references, see:

Shen, T. & Konstantopoulos, S. (2022). Are class size and teacher characteristics associated with cognitive outcomes in early grades? *School Effectiveness and School Improvement, 33*(3), 333-359. Retrieved October 22, 2023, from https://doi.org/10.1080/09243453.2021.2023585

2 Filges, T., Sonne-Schmidt, C.S., & Nielsen, B.C.V. (2018, October 11). *Small class sizes for improving student achievement in primary and secondary schools: A systematic review*. Campbell Collaboration, *Campbell Systematic Reviews, 14*(1). Retrieved October 17, 2023, from https://doi.org/10.4073/csr.2018.10

3 The meta-analysis is deficient in two respects. One deficiency is that the four STAR studies are excluded from the meta-analysis. See page 30 of Filges, T., Sonne-Schmidt, C.S., & Nielsen, B.C.V. (2018, October 11). *Small class sizes for improving student achievement in primary and secondary schools: A systematic review*. Campbell Collaboration, *Campbell Systematic Reviews, 14*(1). Retrieved October 17, 2023, from https://doi.org/10.4073/csr.2018.10. The second deficiency is that the meta-analysis only provides a point estimate from combining the six remaining studies; a meta-analytical examination of study variability and heterogeneity is not performed. Thus, it is unknown why some studies yield strong results and others yield weaker results.

4 These other factors include: behavior and non-cognitive skills, at the student level; satisfaction and pay, at the teacher level; and climate, at the school level.

5 There is plenty of evidence on these broader consequences. Some of it was published before this Review; and more has emerged since. Evidence from before 2017 includes:

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D. & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics, 126*(4), 1593-1660. Retrieved October 20, 2023, from https://doi.org/10.1093/qje/qjr041

More recent evidence is from:

Leuven, E. & Lokken, S.A. (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources,* 55(1), 309-348. Retrieved October 19, 2023, from https://doi.org/10.3368/jhr.55.2.0217.8574R2

6 Evidence on the benefit—cost analysis of CSR:

Washington State Institute for Public Policy (n.d.). *Class size: reducing average class size by one student in grade 1* [webpage]. Retrieved October 19, 2023, from https://www.wsipp.wa.gov/BenefitCost/Program/203

7 We might assume teachers divide their attention equally across students. CSR40—39 will increase attention per student by 2.4%. But CSR20—19 will increase attention per student by 5.2%.

8 Connolly, M. & Haeck, C. (2022). Nonlinear class size effects on cognitive and noncognitive development of young children. *Journal of Labor Economics, 40*(1), S341-S382. Retrieved October 21, 2023, from https://www.journals.uchicago.edu/doi/10.1086/718328

9 For a rounded interpretation, see:

Barnum, M. (2022, June 11). Does class size really matter? A Chalkbeat look at the research. *Chalkbeat*. Retrieved October 19, 2023, from https://www.chalkbeat.org/2022/6/10/23162544/class-size-research

For a recent discussion of some of the tensions in CSR see:

Closson, T. (2023, October 18). Smaller classes? At elite schools, some parents say 'no thanks.' *New York Times*. Retrieved October 20, 2023, from https://www.nytimes.com/2023/10/18/nyregion/nyc-schools-class-size.html

10 For an opinion that relies heavily on the Review, see:

Bloomberg, M.R. (2023, August 30). *US students need better schools, not smaller classes*. Bloomberg. Retrieved October 20, 2023, from https://www.bloomberg.com/opinion/articles/2023-08-30/michael-r-bloomberg-us-students-need-better-schools-not-smaller-classes

For an overinterpretation of the relevance of the findings, see:

Barshay, J. (2021, June 7). *Proof points: Pandemic relief money is flowing to class-size reduction but research evidence for it isn't strong*. Hechinger Report. Retrieved November 3, 2023, from https://hechingerreport.org/proof-points-pandemic-relief-money-is-flowing-to-class-size-reduction-but-research-evidence-for-it-isnt-strong/

For a similar interpretation, see:

Wing Institute at Morningside Academy (2019, October 14). *Are small class sizes a panacea for improving student performance?* Retrieved November 3, 2023, from https://www.winginstitute.org/news/are-small-class-sizes-a-panacea-for-improving-student-performance/