TeachME PROFESSIONAL DEVELOPMENT
CEU's for Teachers and Educators

How Class Size Impacts Student Outcomes

Executive Summary

Public education has undergone major reforms in the last 30 years with the rise in high-stakes testing, accountability, and charter schools, as well as the current shift toward Common Core Standards. In the midst of these reforms, some policymakers have argued that class size does not matter. This opinion has a popular proponent in Malcolm Gladwell, who uses small class size as an example of a "thing we are convinced is such a big advantage [but] might not be such an advantage at all."

These critics are mistaken. Class size matters. Research supports the common-sense notion that children learn more and teachers are more effective in smaller classes.

This policy brief summarizes the academic literature on the impact of class size and finds that class size is an important determinant of a variety of student outcomes, ranging from test scores to broader life outcomes. Smaller classes are particularly effective at raising achievement levels of low-income and minority children.

Considering the body of research as a whole, the following policy recommendations emerge:

- Class size is an important determinant of student outcomes, and one that can be directly determined by policy. All else being equal, increasing class sizes will harm student outcomes.
- The evidence suggests that increasing class size will harm not only children's test scores in the short run, but also their long-run human capital formation. Money saved today by increasing class sizes will result in more substantial social and educational costs in the future.
- The payoff from class-size reduction is greater for low-income and minority children, while any increases in class size will likely be most harmful to these populations.
- Policymakers should carefully weigh the efficacy of class-size policy against other potential uses of funds. While lower class size has a demonstrable cost, it may prove the more cost-effective policy overall.

Introduction

Public education has undergone major reforms in the last 30 years with the rise in high-stakes testing, accountability, and charter schools, as well as the current shift toward Common Core Standards. The availability of new datasets that follow large numbers of students into the workforce has allowed researchers to estimate the lifetime impact of being taught by teachers who increase students' standardized test scores.[1] In the midst of these new reforms and policy concerns, some have argued that class size does not matter. This opinion has a popular proponent in Malcolm Gladwell, who uses small class size as an example of a "thing we are convinced is such a big advantage [but] might not be such an advantage at all."

The critics are mistaken. Class size matters. Class size is one of the most-studied education policies, and an extremely rigorous body of research demonstrates the importance of class size in positively influencing student achievement. This policy brief first reviews the research on class size. Special attention is given to the literatures in economics and related fields that use designs aimed at disentangling causation from correlation. It then documents the recent rise in class size and considers how to compare the effects of class-size reduction with other commonly discussed policy alternatives.

Review of research

Research shows that students in the early grades perform better in small classes. This is especially the case for students who come from disadvantaged backgrounds, who experience even larger performance gains than average students when enrolled in smaller classes. Small class sizes enable teachers to be more effective, and research has shown that children who attend small classes in the early grades continue to benefit over their entire lifetime.[2]

The importance of research design

Isolating the causal impact of policies such as class-size reduction is critical, but challenging, for researchers. Sometimes people will argue based on less sophisticated analyses that class size does not matter. Simple correlational arguments may be misleading, though. Since variation in class size is driven by a host of influences, the simple correlation between class size and outcomes is confounded by other factors. Perhaps the most common misinterpretation is caused by low-achieving or special needs students being systematically assigned to smaller classes. In these cases, a simple correlation would find class size is *negatively* associated with achievement, but such a

finding could not be validly generalized to conclude that class size does not matter or that smaller classes are harmful. Instead, because class size itself is correlated with other variables that also have an impact on achievement, such as students' special needs status, the estimated relationship between class size and outcomes would be severely biased.

The academic research has many examples of poor-quality studies that fail to isolate the causal impact of class size, most of them written and published prior to the so-called "credibility revolution" in economics.[3] Eric Hanushek has surveyed much of the early research on class size, as well as other educational inputs such as per-pupil spending, in a

> *Importantly, small classes have been found to have positive impacts not only on test scores during the duration of the class-size reduction experiment, but also on life outcomes in the years after the experiment ended.*

pair of older but influential articles from 1986 and 1997, which have been revived in Gladwell's popularized book.[4] Based on these surveys, he concluded at the time that "there is not a strong or consistent relationship between student performance and school resources" such as class size or spending. In a thorough re-analysis of Hanushek's literature summary, Krueger demonstrates that this conclusion relies on a faulty summary of the data. In particular, Hanushek's summary is based on 277 estimates drawn from 59 studies, but while more estimates are drawn from some studies than others, each estimate is weighted equally. As a result, Hanushek's literature summary places a disproportionate weight on studies that analyzed smaller subsets of data. Krueger argues that since studies, not individual estimates, are what are accepted for publication, weighting by study is more appropriate than weighting by the number of estimates. When Krueger re-analyzed the data giving each study equal weight, he found that there is indeed a systematic positive relationship between school resources and student performance in the literature surveyed by Hanushek.

More troubling, many of the studies included in the survey employed research designs that would not allow researchers to isolate causal effects. For example, one-third of the studies ignored the relationship between different measures of school inputs, and held constant per-pupil spending while studying the "impact" of class size. Because smaller classes cannot be had without increased spending on teachers, it is inappropriate to include spending as a control variable and effectively hold spending constant when investigating class size. The resulting estimate does not provide insight about the impact of reducing class size, but instead estimates a convoluted value that is something like the impact of reducing class size while simultaneously paying teachers less, which is unrealistic.[5] Such evidence does not reflect the impact of class size and should not be used to inform policy.[6] Nonetheless, in his 2013 book *David and Goliath*, Malcolm Gladwell uncritically cites the Hanushek literature summary and its argument that the class size literature is inconclusive.[7] As demonstrated below, well-designed studies generally—with a few notable exceptions—find strong class-size impacts.

The modern research paradigm strongly prefers the use of research designs that can credibly isolate the cause-and-effect relationship between inputs and outcomes. Scholars generally agree that true randomized experiments, such as the Project STAR class-size experiment described below, are the "gold standard" for isolating causal impacts. When an experiment is not available, researchers are sometimes able to employ other techniques that mimic experiments—termed "quasi-experiments" in the literature—that can better infer causality.

In implementing a quasi-experimental study, there must be some sort of variation in class size that is random or nearly random. Such variation is hard to come by, and in many cases there is no way for researchers to isolate the impact of class size. Thus, some of the older and better-designed studies inform the policy debate more accurately than newer studies that employ less sophisticated and simpler correlational designs.

Evidence from Tennessee's STAR randomized experiment

The best evidence on the impact of reducing class sizes comes from Tennessee's Student Teacher Achievement Ratio (STAR) experiment.[8] A randomized experiment is generally considered to be the gold standard of social science research. In STAR, over 11,500 students and 1,300 teachers in 79 Tennessee elementary schools were randomly assigned to small or regular-sized classes from 1985-89. The students were in the experiment from kindergarten through third grades. Because the STAR experiment employed random assignment, any differences in outcomes can be attributed with great confidence to being assigned to a smaller class. In other words, students were not more or less likely to be assigned to small classes based on achievement levels, socio-economic background, or more difficult-to-measure characteristics such as parental involvement.[9]

The results from STAR are unequivocal. Students' achievement on math and reading standardized tests improved by about 0.15 to 0.20 standard deviations (or 5 percentile rank points) from being assigned to a small class of 13-17 students instead of a regular-sized class of 22-25 students.[10] When the results were disaggregated by race, black students showed greater gains from being assigned to a small class, suggesting that reducing class size might be an effective strategy to reduce the black-white achievement gap.[11] Small-class benefits in STAR were also larger for students from low socio-economic-status families, as measured by eligibility for the free- or reduced-priced lunch program.

A follow-up study of the most effective teachers in STAR found that teachers used a variety of strategies to promote learning and that small classes allowed them to be more effective in employing these strategies. For example, they closely monitored the progress of student learning in their classes, were able to re-teach using alternative strategies when children did not learn a concept, had excellent organizational skills, and maintained superior personal interactions with their students.[12]

Importantly, small classes have been found to have positive impacts not only on test scores during the duration of the class-size reduction experiment, but also on life outcomes in the years after the experiment ended. Students who were originally assigned to small classes

did better than their school-mates who were assigned to regular-sized classes across a variety of outcomes, including juvenile criminal behavior, teen pregnancy, high school graduation, college enrollment and completion, quality of college attended, savings behavior, marriage rates, residential location and homeownership.[13]

Most other quasi-experimental evidence is consistent with STAR

True randomized experiments such as Tennessee's random assignment of students across an entire state to experimental and control groups are quite rare. Therefore, researchers must also look for quasi-experimental approaches that allow isolation of the causal impact of class-size reduction. Other high-quality studies that isolate the effect of small class size in elementary school on student outcomes generally show results similar to those found in STAR.

For example, a quasi-experimental approach was used to evaluate Wisconsin's targeted class-size reduction program. In the Student Achievement Guarantee in Education (SAGE) program, high-poverty school districts could apply to implement a pupil-teacher ratio of 15-to-1 in grades K-3.[14] While most participating schools reduced class sizes, some schools chose to attain the target pupil-teacher ratio by using two-teacher teams in classes of 30 students. Test scores of first-grade students in SAGE schools were higher in math, reading, and language arts compared with the scores of those in selected comparison schools in the same districts with average pupil-teacher ratios of 22.4 to 24.5. Attending small classes improved student achievement by approximately 0.2 standard deviations.[15]

The most famous quasi-experimental approach to studying class-size reduction comes from Angrist and Lavy's use of a strict maximum-class-size rule in Israel and a regression discontinuity (RD) approach.[16] In Israel, there is a strict maximum class size of 40 students. As a result, class size drops dramatically when enrollment in a grade in a school approaches the point when the rule requires the school to add a new classroom—i.e., when enrollment tips above a multiple of 40. For example, if a grade has 80 students, then a school could offer as few as 2 classrooms, with the maximum allowable class size of 40 students in each. If a grade has 81 students, however, the school is required to offer at least 3 classrooms, and consequently the maximum average class size falls to 27 students. In practice, some schools add an additional classroom prior to hitting the 40-student cap. Nonetheless the maximum-class-size rule is a good predictor of actual class sizes and can be used in an instrumental-variables research design to isolate the causal impact of class size on student achievement. Using the variation in narrow bands around enrollment sizes that are multiples of 40 students, Angrist and Lavy find strong improvements overall in both math and reading scores, of a magnitude nearly identical to that of Project STAR's experimental results. Consistent with the STAR results, they also find larger improvements among disadvantaged students.

Several subsequent papers have identified the impact of smaller class sizes using maximum class-size rules in other international settings.[17] (Note that quasi-experimental approaches tend to require large datasets and data spanning a large number of years. Such datasets are more likely to derive from settings outside the United States.) Most recently,

Fredriksson *et al.* evaluated the long-term impact of class size using data from students in Sweden between ages 10 and 13 who were facing a maximum-class-size rule of 30 students.[18] At age 13, students in smaller classes had higher cognitive and non-cognitive skills, such as effort, motivation and self-confidence. In adulthood (between ages 27 and 42), those who had been in smaller classes had higher levels of completed education, wages, and earnings. Urquiola used a similar regression discontinuity approach in Bolivia and found that a one standard-deviation reduction in class size (about 8 students in his data) improves test score performance by 0.2 to 0.3 standard deviations.[19] Browning and Heinesen derive similar results from data from Denmark, even though the average class size is much smaller in their study (20 pupils per classroom, compared with 31 students in Angrist and Lavy's Israeli data).[20]

A different quasi-experimental approach is to use variation in enrollment driven by small variations in cohort sizes across different years. Hoxby takes this approach using data from the state of Connecticut, finding no statistically significant positive effect of smaller class size.[21] One drawback of the Connecticut study is that test scores are only measured in the fall, so the impact of the prior year's class size may be somewhat mitigated by the time spent away from school in the summer. The discrepancy between Hoxby's Connecticut results and those of other studies that also use research designs capable of uncovering causal relationships is an unresolved puzzle. Despite the overwhelming pattern in the literature of positive class-size impacts, Malcolm Gladwell, intent on supporting his point about what he calls the "theory of desirable difficulty," described only the Hoxby results in his description of research on class size in his recent book.[22]

Results from statewide class-size-reduction policies

Based in part on the research evidence on the impact of class-size reduction, several U.S. states, including California, Texas and Florida, have implemented class-size caps. The most widely studied of these policies is the 1996 California law that gave strong monetary incentives to schools to reduce class size in grades K-3 to 20 or fewer students. Sometimes when a new policy is introduced it is phased in slowly across locations, which gives researchers the opportunity to compare outcomes in schools that have adopted the policy with those that have not yet done so. In California, however, the policy was nearly universally adopted within a short period of time, so there was very little opportunity to compare early implementers with later implementers. Furthermore, test scores are only available starting in grade 4, so any evaluation of the policy is forced to use test scores from later than the year in which the reduced class size was experienced. Although there were positive impacts on achievement due to class-size reductions on the order of 0.05 to 0.10 standard deviations, these impacts may have been offset because many inexperienced teachers had to be hired to staff the new classrooms, reducing average teacher quality.[23]

Why are small classes more effective?

The mechanisms at work linking small classes to higher achievement include a mixture of higher levels of student engagement, increased time on task, and the opportunity small

classes provide for high-quality teachers to better tailor their instruction to the students in the class. For example, observations of STAR classrooms found that in small classes students spent more time on task, and teachers spent more time on instruction and less on classroom management.[24] Similar results have been found in other settings.[25] However, qualitative research from the pupil-teacher ratio reduction in Wisconsin's SAGE program indicates that such beneficial adaptations in teachers' practices will not necessarily occur. It is important to provide professional-development support to instruct teachers on how to adapt their teaching practices to smaller classes.[26]

In addition, small classes may have a positive impact on student "engagement behaviors," which include the amount of effort put forth, initiative taken, and participation by a student. Not surprisingly, these characteristics have been shown to be important to classroom learning. Finn finds that students who were in small classes in STAR continued to have higher engagement ratings in subsequent grades.[27]

It is sometimes argued that class size only matters for inexperienced or low-quality teachers because more effective teachers are better able to adapt their teaching styles to accommodate larger classrooms. The evidence suggests that the opposite is true. In STAR, the positive impacts of small classes were found to be larger for experienced teachers.[28] Experienced teachers are better able to take advantage of smaller class sizes to make pedagogical changes.

What does the evidence say about how small is small enough?

The best evidence on class-size reduction is from the STAR experiment, which estimated substantial positive impacts from class-size reduction from an average of 22 to an average of 15. In fact, the class sizes targeted in STAR were informed by influential work by Glass and Smith that found strong impacts from class sizes below 20.[29] Based on this, some researchers conclude that the evidence supports better outcomes only if classes are below some threshold number such as 15 or 20. Sometimes the argument is extended to suggest that reducing class size is not effective unless classes are reduced to within this range. The broader pattern in the literature finds positive impacts from class-size reductions using variation across a wider range of class sizes, including class-size reductions mandated by maximum class-size rules set at 30 (Sweden) or 40 (Israel). In fact, the per-pupil impact is reasonably stable across class-size reductions of different sizes and from different baseline class sizes. For example, when scaled by a 7-student class-size reduction as in the Tennessee experiment, the Israeli results imply a 0.18 standard deviation increase in math scores, which is nearly identical to the Tennessee results.[30] The weight of the evidence suggests that class-size impacts might be more or less linear across the range of class sizes observed in the literature—that is, from roughly 15 to 40 students per class. It would be inappropriate to extrapolate outside of this range (as is done in the Gladwell book).

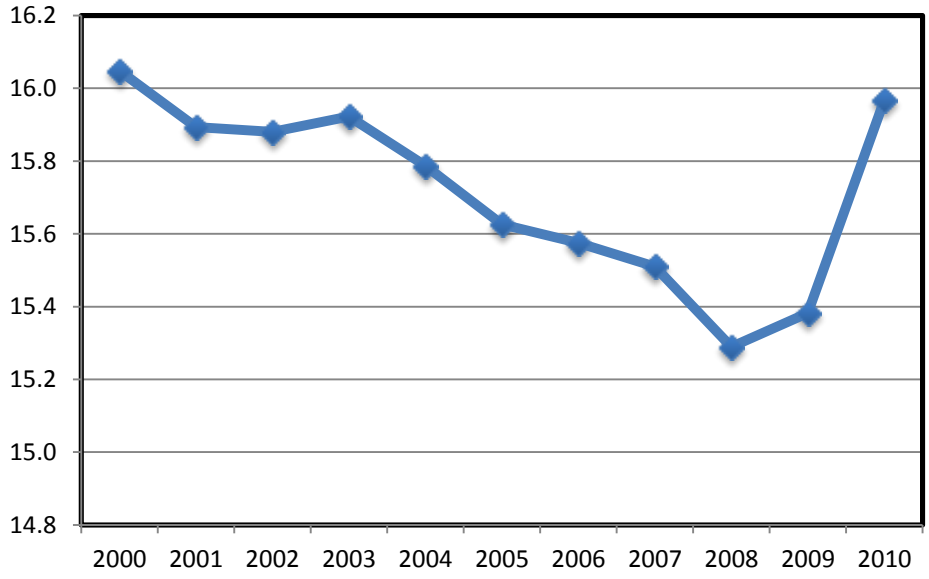
Do small classes matter in later grades?

Most of the high-quality evidence on class-size reduction is based on studies of the early grades. The available high-quality evidence on the impact of class size on outcomes in

older grades is more limited, and more research in this area is needed. A notable exception is Dee and West, who estimate class-size effects using variation in class sizes experienced by students across classes in different subjects, and by students taking classes from the same teachers in different class periods. The study finds that smaller class sizes in eighth grade have a positive impact on test scores and measures of student engagement, and finds some evidence that these impacts are larger in urban schools.[31]

Recent Developments

Student-teacher ratios in public schools fell steadily over the past 40 years until recently. Between 2008 and 2010, however, the student-teacher ratio increased by 5%, from 15.3 to 16.0 (see Figure 1). Note that actual class sizes are typically larger than student-teacher ratios, because these ratios include special teachers who are not included in class-



Source: Digest of Education Statistics (table 78, 2012; table 69, 2011)

Figure 1. Student/Teacher Ratios in Public Elementary and Secondary Schools

size counts, such as teachers for students with disabilities.[32] For example, imagine a grade level in a school that contains three "regular" classes with 24 students in each and one compensatory class with only 12. This school would have a pupil-teacher ratio of 21, even though most of the students in that grade (in fact, 85% of them) are in classes with 24 students. This is a reason why simple correlations between class size and student outcomes may be misleading. If some students are placed in smaller classes because they have low

performance levels, this biases the estimate of the positive effect of small classes downward.[33]

According to the Schools and Staffing Survey, in 2011-12 the average United States class size for public primary school teachers in self-contained classes was 21.6, up from 20.3 in 2007-08.[34] During this time frame, the recession forced California to abandon its class-size reduction policy, which had provided incentives for districts to adopt a 20-student cap in grades K through 3.[35] In response, the average K-3 class size increased from 23 students in 2008-09 to 26 students in 2012-13.

Table 1. Hypothetical Distribution of Students with Different Numbers of Teachers

		Allocation with 24 teachers			Allocation with 23 teachers	
Grade	Enrollment	Number of classes	Class size		Number of classes	Class size
K	100	4	25		4	25
1	100	4	25		4	25
2	100	4	25		4	25
3	100	4	25		4	25
4	100	4	25		4	25
5	100	4	25		3	33.3
Total	600	24			23	
Average class size			25			26.4
Average pupil-teacher ratio			25			26.1

Small increases in average class sizes can mask large class-size increases in some districts and schools. For example, sometimes policymakers will calculate the cost savings from increasing the average class size by a single student, arguing or implying that the impact on test scores from this “modest” one-student increase will be negligible.[36] This line of reasoning is misleading because actual classes and teachers are not easily divisible into fractions.[37] As illustrated in Table 1, imagine a K-5 school that has 100 students in each grade with four classrooms for each grade. Each of the 24 classes in the school has a class size of 25 students. If this school had to lay off one fifth-grade teacher, the aggregate numbers would not increase very much. The average pupil-teacher ratio would increase only slightly, from 25.0 to 26.1, while the average class size would increase from 25.0 to 26.4. These averages mask the sharp increase in class size experienced by the fifth-grade students, from 25 to 33.3. The negative impact of increasing class size by 8 students in

fifth grade would be expected to be sizeable, but it might not raise alarms to the average parent told that the pupil-teacher ratio increased by only 1 student.[38]

Discussion and Analysis

Recently some policymakers and education analysts have argued that manipulating other educational inputs would be more effective or more cost-effective than class-size reduction. By and large, though, these suggestions do not pit class-size reductions against some other policy alternative that has been implemented and evaluated. It is only appropriate to compare effectiveness across a variety of policy alternatives.

For example, recent studies have found that teachers with high value added on standardized test scores also have an impact on such subsequent outcomes for their students as wage earnings.[39] Based on these findings, some argue that giving students a high-test-score value-added teacher is more cost-effective than class-size policy. The problem with this suggestion is that there are few—if any—policies that have been designed, implemented and evaluated that increase the availability of teachers with high-test-score value added and result in higher student achievement. It's one thing to measure the impact of teachers on their students' standardized test scores, but it is a separate challenge to design a policy lever to bring more teachers into the classroom who can raise test scores. A recent report from the Institute of Education Sciences documents that disadvantaged students are taught by teachers with lower value added on tests.[40] At this point we know relatively little about how to increase teacher quality, much less how much it will cost to induce more high-quality teachers to work and stay in the schools that need them. Much more needs to be done in terms of pilot programs, policy design and evaluation before improving teacher quality can be considered a viable policy option.

Another proposal has been floated (e.g. by Bill Gates) to pay high-quality teachers bonus payments for taking on extra students.[41] It is certainly possible that such a reallocation of students could increase overall achievement, but it is also possible that it would backfire. For example, imagine a school with a grade containing two classes. One teacher is an excellent, experienced teacher, while the other is an untested, first-year "rookie" teacher. One option would be for both teachers to get classes with 25 students. Another option would be to pay the experienced teacher a bonus to take a class of 29 students, leaving the rookie teacher with a class of 21 students. All else equal, children in the experienced teacher's class would likely record lower test score gains if there were 29 students than if there were 25, but these gains would be enjoyed by more students. Perhaps the 21 students in the rookie teacher's classroom would be better off than if they would have been in a classroom of 25 students, though the research is less clear about whether the rookie teacher will be more effective in a small class. In this hypothetical case, it is possible that the aggregate test score gains could be larger when the classrooms have unequal sizes, especially if the experienced teacher is substantially more skilled at raising test scores than the rookie teacher. Whether it is an effective policy, however, hinges crucially on a variety of factors: how large the skill differential is between teachers, how large a bonus payment is required to induce the experienced teacher to accept a larger class, what the next best

use is for the funds used for the bonus payment, and whether the gains persist over time. While this is a potentially interesting area for policy development, much more pilot testing needs to be done before it could be considered a credibly policy alternative to class-size reduction.

Recommendations

The academic literature strongly supports the common-sense notion that class size is an important determinant of student outcomes. Class-size reduction has been shown to improve a variety of measures, ranging from contemporaneous test scores to later-life outcomes such as college completion.

Based on the research literature, I offer the following policy recommendations:

- Class size is an important determinant of student outcomes and one that can be directly influenced by policy. All else being equal, increasing class sizes will harm student outcomes.
- The evidence suggests that increasing class size will harm not only children's test scores in the short run but also their long-term human capital formation. Money saved today by increasing class sizes will be offset by more substantial social and educational costs in the future.
- The payoff from class-size reduction is larger for low-income and minority children, while any increases in class size will likely be most harmful to these populations.
- Policymakers should carefully weigh the efficacy of class-size-reduction policy against other potential uses of funds. While lower class size has a demonstrable cost, it may prove the more cost-effective policy overall.

"This document was developed from the public domain document: Does Class Size Matter? - Schanzenbach, D.W., National Education Policy Center (http://nepc.colorado.edu/publication/does-class-size-matter)."