

Should Class Size Reduction Substitute for Innovation?

Eric A. Hanushek
University of Rochester

Testimony before the
Committee on Health, Education, Labor and Pensions
United States Senate

June 23, 1999

Improvement of our schools should receive a very high priority, and I believe that the seriousness with which you are approaching the re-authorization of the Elementary and Secondary Education Act is laudatory. Education is known to have huge impacts on the welfare of the Nation – entering into the growth of the U.S. economy, the incomes of its citizens, and the distribution of success across the population. But recent analyses demonstrate clearly that we are not getting what we should out of the expenditures and resources devoted to schools, and, worse, that our schools are falling behind.

I want to make a few simple points today.

- The move to mandate smaller classes, backed by proposed federal spending of \$1.4 billion this fiscal year, is misguided and could even hurt student achievement
- The accumulated evidence of the impact of reduced class size on student performance gives no reason to expect that the wave of class size reductions that have been proposed will have *any* effect on achievement.
- Class size reductions are very expensive and take resources and attention away from potentially productive reform efforts.
- Teacher quality is much, much more important than class size, and attention should focus on ways to improve teacher quality.
- The Federal government should concentrate on developing evidence about incentives and other reform approaches and should avoid large expenditures on unproven approaches.

Supporters of broad class size reductions generally point to a few studies or a few experiences that suggest improved performance with smaller classes and then rely on the “obviousness” of the proposed policies to carry the day. To be sure, there are U.S. classrooms that are overcrowded. But blanket reductions are wasteful and ineffective, unless the future looks nothing like the past. A thorough review of the scientific evidence shows a startling finding: class size reduction may be one of the *least* effective educational investments.

Proposals to Reduce Class Sizes

The widespread belief that lowering class sizes immediately improves education has been echoed by politicians in both parties during this election year. A majority of Governors are either proposing or actively considering class size reduction initiatives. These states are following on the heels of California, which reduced K-3 class sizes under Republican governor Pete Wilson after the state generated a revenue windfall in 1996. GOP proposals both in Congress and in many states to shift education dollars from “administration” to “classrooms” are also often promoted as enabling school districts to reduce class sizes.

The President proposed to spend \$1.4 billion in federal funds in the coming year to reduce class sizes in grades 1-3. These initiatives are designed to help bring classes in the early grades down to 18 students per class by 2005, an undertaking estimated to require 100,000 additional teachers.

Class size reduction is undoubtedly politically popular. The concern here is whether it is educationally desirable.

The Evidence on Class Size¹

A wide range of perspectives can be taken in attempting to pinpoint the effectiveness of reduced class sizes. No matter what the source of evidence, the answer about effectiveness is the same: broad policies of class size reduction are very expensive and have little effect on student achievement.

1. The United States has extensive experience with class size reduction and it has not worked.

Between 1965 and 1995 -- most of the period of the Elementary and Secondary Education Act -- pupil-teacher ratios fell by 30 percent, from about 25:1 to about 17:1 overall.² These reductions have been an important component of the dramatic increases in school spending that have occurred over this period. Table 1 shows the pattern of pupil-teacher ratios, teacher attributes, and real spending per pupil since 1965-66. The one-third fall in pupil-teacher ratios is a significant contributor to the 140 percent increase in real spending per student in average daily attendance (ADA). (The table further shows that other teacher attributes — advanced degrees and experience — also grew significantly.)

While we lack information about student achievement for this entire period, the information that we have from 1970 for the National Assessment of Educational Progress

¹A more detailed discussion of the evidence along with citations for the relevant work can be found in Eric A. Hanushek, "The Evidence on Class Size," Occasional Paper No. 98-1, W. Allen Wallis Institute of Political Economy, University of Rochester, February 1998 and Eric A. Hanushek, "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(2), forthcoming Summer 1999. The complete text of these is also available at: <http://petty.econ.rochester.edu>.

²Pupil-teacher ratios differ from class size for a variety of reasons including the provision of specialized instruction (as with special education), the use of teachers in supervisory and administrative roles, and the contractual classroom obligations of teachers. Nonetheless, even though we have little longitudinal data for class sizes, average class size will tend to move with pupil-teacher ratios.

(NAEP) indicates that our 17-year-olds were performing roughly the same in 1996 as in 1970. There are some differences by subject area. For science, the average scale score of 17-year-olds *falls* 9 points between 1969 and 1996. For math, 17-year-olds improve 3 points between 1973 and 1996. For reading, they improve 2 points between 1971 and 1996. Writing performance, which is only available since 1984, shows a fall of 7 points, by 1996. Only the fall in science (and in writing since 1984) is a statistically significant difference. There have been improvements at earlier ages, but they are not maintained and are not reflected in the skills that students take to college and to the job market. The overall picture is one of stagnant performance.

Table 1. Public School Resources in the United States, 1966-1996

Resource	65-66	70-71	75-76	80-81	85-86	90-91	95-96
Pupil-teacher ratio	24.7	22.3	20.4	18.7	17.9	17.2	17.3
% teachers with master's degree or more	23.3	27.5	37.5	49.6	51.4	53.1	56.2
median years teacher experience	8	8	8	12	15	15	15
current expenditure/ADA (1996-97 \$'s)	\$2,678	\$3,645	\$4,308	\$4,589	\$5,484	\$6,239	\$6,434

One common explanation for why the lower pupil-teacher ratio hasn't resulted in increased overall performance is that more students are now designated as special education students, whose classes are much smaller than regular ones. About 12.5 percent of students are now identified as having disabilities covered under special education legislation (up from 8 percent at the introduction of programs in the late 1970s). Indeed, the federal and state mandates for the education of handicapped students have placed significant requirements on hiring staff and providing extensive services. On average these students cost somewhat more than twice that of regular instruction. While these programs could account for as much as a third of the increased intensity of teachers over the 1980s, substantial reductions in class size have been directed at regular class room instruction as well.

In sum, the proposals to reduce class sizes are nothing new. We have been pursuing these policies for decades. The aggregate evidence shows no improvements in student performance that can be related to the overall pupil-teacher ratio reductions.

2. International comparisons suggest no relationship between pupil-teacher ratios and student performance.

The recent results of the performance of U.S. students on international math and science examinations have sobered many. Our high school seniors performed near the bottom of the rankings of the 21 nations participating in the Third International Mathematics and Science Study (TIMSS). This showing has nothing to do with more selective students taking the tests in other countries — our best students performed badly.

At the same time, the dramatic differences in pupil-teacher ratios and in class sizes across countries are unrelated to measures of mathematics and science achievement. Of course there are many differences across countries that are difficult to adjust for in any analysis, but if smaller classes were strongly related to high student achievement, then one would expect U.S. class sizes to be much larger than those in other countries. In fact, just the opposite is true. Asian countries that routinely outperform the U.S. generally have class sizes of 30-40 students. Ironically, the international differences suggest that there is a slight *positive* relationship between pupil-teacher ratios and student achievement.

3. Extensive econometric investigations show no systematic relationship between class size and student performance.

Over the past three decades, there has been significant research in deciphering what factors affect student achievement. This work, employing sophisticated econometric techniques, provides considerable evidence about the effects of class size on performance.

These extensive statistical investigations show almost as many positive as negative estimates of the effects of reducing class size. Table 2 summarizes the 277 separate published estimates of the effect of teacher-pupil ratios on student achievement. Only 15 percent give much confidence (i.e., are statistically significant) that there is the expected improvement from reducing class sizes. The bulk (85 percent) either suggest that achievement worsens (13 percent) or gives little confidence that there is any effect at all.

Because of the controversial nature of these conclusions, they have been carefully scrutinized – and the policy conclusions remain unaffected. The subsequent discussions have clarified one important aspect of these analyses. The existing studies do show that sometimes variations in class size have significant influences on performance. The difficulty, when thought of in terms of making policy from Washington or from state capitals, is that nobody has been able to identify the overall circumstances that lead to beneficial effects. This finding has important policy implications that are discussed below.

These studies are important because they provide detailed views of differences across classrooms – views that separate the influence of schools from that of family, peers, and other factors. As a group, they cover the influence of class size on a variety of student outcomes, on performance at different grades, and on achievement in different kinds of schools and different areas of the country. In sum, they provide broad and solid evidence.

Table 2. Percentage Distribution of Estimated Influence of Teacher-pupil on Student Performance, By Level of Schooling

School level	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
All schools	277	15%	13%	27%	25%	20%
Elementary schools	136	13	20	25	20	23
Secondary schools	141	17	7	28	31	17

4. Project STAR in Tennessee does not support overall reductions in class size except perhaps at kindergarten.

Much of the current enthusiasm for reductions in class size is based on the results of a random-assignment experimental program in the State of Tennessee in the mid 1980s. The common reference to this program, Project STAR, is an assertion that the positive results there justify a variety of overall reductions in class size. This study is the primary reference in the Clinton proposal as well as Governor Pete Wilson’s dramatic class size reductions in California in 1996.

The study is conceptually simple, even if some questions about its actual implementation remain. Students and teachers in the STAR experiment were randomly assigned to small classes (13-17 students) or large classes (22-25 students) with or without aides. Each participating school had one of each type of class. Students were kept in these small or large classes from kindergarten through third grade, and their achievement was measured at the end of each year.

Considerable uncertainty exists about the correct interpretation of the STAR results, because a number of problems with the design and interpretation of the experiment. The approach of Project STAR – the use of random assignment experimentation – is a real milestone in education research, and it should be employed much more often. The superiority of the research approach should not, however, be confused with the results or the policy interpretation of them.

The STAR evidence showed that the gains made were mainly in kindergarten. The STAR data are summarized by Figures 1 and 2. At the end of kindergarten, children in small classes

score better than those in large classes. They then maintain this differential for the next three years.

If smaller classes were valuable in each grade, the achievement gap would widen. It does not. In fact, the gap remains essentially unchanged through the sixth grade, even though the experimental students from the small classes return to larger classes for the fourth through sixth grades. The inescapable conclusion is that the smaller classes at best matter in kindergarten and perhaps first grade. The data do not suggest that improvements will result from class size reductions at later grades.

The STAR data suggest that perhaps achievement would improve if kindergarten classes were moved to sizes considerably below today's average. In addition, the effects were greater for minority students during the first two years. The President's plan gives greater assistance to Title 1 schools and targets the early grades, but it excludes kindergarten.

Nonetheless, the STAR evidence pertains to a one-third reduction in class sizes, a reduction approximately equal to the overall decline in pupil-teacher between 1950 and today. As we have seen, that reduction has not led to overall improvement in student achievement.

Interpreting the Evidence on Class Size

None of this says that smaller classes never matter. The class size evidence refers to the normal ranges observed in schools—roughly between 15 and 40 students per class. A class of 100 would likely produce different effects than a class of 5, but such a comparison is irrelevant for purposes of the broad policies currently being considered.

Existing teachers may well not adjust their classroom behavior with fewer children in the classroom, and new teachers hired to staff the additional smaller classes may not be as good as existing teachers. There may be situations – of specific teachers, specific groups of students, and specific subject matters – where the huge expense of smaller classes may be very beneficial for student achievement. At the same time, there are other situations where a large scale class size reduction policy could take away from other education priorities and result in stagnant or worse student achievement.

The complexity of the situation is that we do not know how to describe *a priori* situations where reduced class size will be beneficial. It makes little sense to dictate an across-the-board class size reduction policy from Washington. A national policy can only expect average gains, which appear to be very small, at a great expense.

It is also important to remember that bad implementation can actually worsen achievement. When California implemented its large-scale class reduction last year, the state scrambled to hire thousands of new teachers; 31 percent of California's new teachers are working with only emergency credentials, with a disproportionate number working in urban districts. Due

to lack of space, some schools have resorted to placing two teachers in a single classroom with forty students.³

Much of the case for reduced class size rests on “common-sense” arguments. With fewer students, teachers can devote more attention to each child and can tailor the material to the individual child’s needs. But, consider, for example, a movement from class sizes of 26 to class sizes of 23. This represents an increase in teacher costs alone of over ten percent. It is relevant to ask whether teachers would in fact notice such a change and alter their approach. The observational information from Project STAR suggested no noticeable changes in typical teacher behavior from the much larger changes in the experiment.

The small classes in California have 20 students in them - about the size of the large classes in STAR. No evidence from STAR relates to the likely effects of such a policy change. Indeed, the STAR study was based on previous research which suggested that a class size of 15 or fewer would be needed to make a significant improvement in classroom performance. The Clinton administration proposals point to class sizes of 18, instead of the 20 in California, but they still do not get down to the STAR levels.

The policy issue is not defined exclusively by whether we should expect positive effects from reducing class sizes. Even if we were confident of positive effects, the case for general policies to reduce class size would not yet be made. Class size reduction is one of the most expensive propositions that can be considered. The policy experiment of Project STAR involved increasing the number of classroom teachers by one-third, a policy with massive expenditure implications if implemented on a wide- scale basis. In recognition of fiscal realities, the expense of such policies puts natural limits on what is feasible, leading many reductions to be in the end rather marginal. Marginal changes, however, are even less likely to lead to underlying changes in the behavior of teachers.

Teacher Quality, Not Quantity

Considerable evidence shows that teacher quality is one of the most important factors in student achievement. Whether or not large-scale reductions in class sizes help or hurt will depend mostly on whether the new teachers are better or worse than the existing teachers. Unfortunately, class size reduction proposals usually are not accompanied by plans to recruit qualified teachers, and the current organization of schools and incentives to hire and retain teachers do little to ensure that the teacher force will improve. Reducing class sizes may likely have a negative effect by increasing the quantity of teachers at a time when what we need most is to increase teacher quality.

³ Wexler, Edward, et. al., **California’s Class Size Reduction: Implications for Equity, Practice & Implementation**, WestEd and PACE, March 1998.

Many current proposals call for states to adopt training and certification procedures that have not been adequately evaluated and tested. Simply trying to raise certification standards in the current system is unlikely to raise teacher quality. Indeed, certification as practiced today already deters too many talented individuals from teaching, and teachers are rarely held accountable for student performance. Moreover, some states may actually have to lower certification standards just to attract enough teachers for each classroom. If we are to have a real impact on teaching, we must evaluate actual teaching performance and use such evaluations in school decisions. We cannot rely on requirements for entry, but must switch to using actual performance in the classroom.

Superior Approaches

The states and federal government are in a unique position to initiate programs that promise true improvement in our schools. They are not programs that mandate or push local schools to adopt one-size-fits all approaches – such as lowering overall class sizes or altering the certification of teachers. Instead they are programs that develop information about improved incentives in schools.

The largest impediment to any constructive change in schools is that nobody in today's schools has much of an incentive to improve student performance.⁴ Careers simply are not made on the basis of student outcomes. The flow of resources is not related positively to performance – indeed it is more likely to be perversely related to performance. Let us return to class size proposals for a moment. Given that school incentives do not push toward better student performance or toward conserving on expenditure, it is little wonder that decisions about class size are made on the basis of “fairness” and not productivity. After all, would it be fair to some teachers to have to teach large classes or to some students to have to have less attention in a larger classroom? If schools were more motivated toward performance, the discussion might shift to identifying those situations where changing class sizes would have their largest impact. For example, reducing kindergarten class sizes might be important in communities that lack preschools; communities that face teacher shortages might instead raise teacher salaries in order to improve its applicant pool and to recruit more qualified teachers.

The unfortunate fact is, however, that we have little experience with alternative incentive structures. A very productive use of state and federal funds would be to conduct a series of planned interventions that could be used to evaluate improvements. Minimally, instead of funding lowered class sizes everywhere, the states and federal government could team together to mandate more extensive random-assignment trials and evaluation of the benefits of lowered class sizes, à la Tennessee.

⁴A full discussion of the issues of incentives and of experimentation is found in **Making Schools Work: Improving Performance and Controlling Costs** (Eric A. Hanushek with others, Washington, DC: Brookings Institution, 1994).

More usefully, they could work to develop a series of experiments that investigates alternative incentive schemes – from merit pay to private contracting to wider choice of schools. A new program of trials with altered performance incentives could place an indelible positive stamp on the nation’s future by committing to learning about how schools can be improved. Today we do not know enough to develop an effective program of improvement. Nor will continuation of past research programs help, because they must rely upon the existing structure of schools with the existing incentives (or lack of incentives).

The issues of incentives and of devising ways to obtain appropriate information is set out in more detail in *Making Schools Work*. These are clearly complicated issues. They would require considerable change in focus by the federal and state governments— turning from trying to dictate how schools do their jobs to setting up incentives for good performance. They also openly admit that there are many gaps in our knowledge and that improving education is more likely if we attack the knowledge problems directly instead of continuing policies that we know do not work.

Investing in Schools

There are powerful reasons to expand and improve investment in human capital. Educational investments are in fact very important for the U.S. economy, which has been built on a skilled labor force and has capitalized on the presence of skills, making human capital investments very important to the economy. Moreover, many authors show that the labor market value of the increased skills, as measured by schooling level, has increased dramatically in recent years. This valuation demonstrates that the economy continues to need an evermore skilled labor force. Economists have recently spent considerable time and effort trying to understand why some countries grow faster than others, and the majority opinion is that a nation’s stock of human capital is an important component of differential growth rates. In addition, Americans have long thought of education as a primary ingredient in providing equality of opportunity to society—as a way of cutting down or breaking intergenerational correlations of income and of trying to provide opportunity to all of society. Taken together, these provide important and relatively uncontroversial reasons for us to continue our attention to education.

Acknowledging the need for investment does not, however, lead to unqualified support for any policies labeled “investment in our youth” or “school improvement.” Recent policy discussions have been laced with programs that fundamentally involve haphazard and ineffective spending on schools and that offer little hope for gains in achievement. The current set of class size proposals falls into this category. President Clinton should leave class size policy to schools and districts, and remain faithful to an education policy defined in terms of results, not merely spending.

Figure 1. Project STAR results

Stanford Achievement Test -- reading

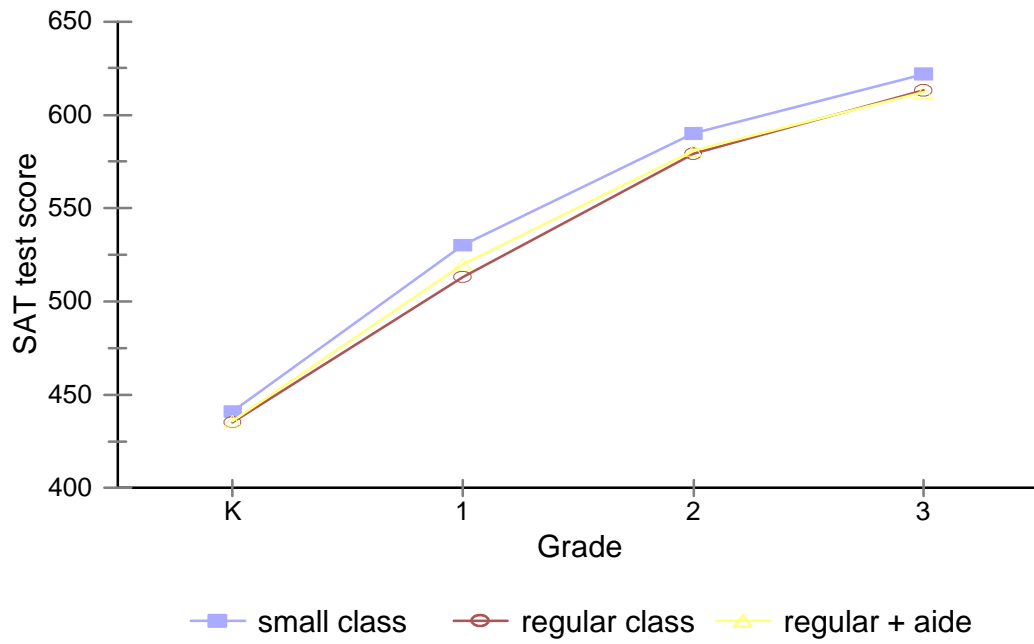


Figure 2. Project STAR results

Stanford Achievement Test -- math

