TESTIMONY PREPARED FOR:

IMPROVING STUDENT ACHIEVEMENT:

EXAMINING THE IMPACT OF TEACHER QUALITY AND CLASS SIZE

THE SUBCOMMITTEE ON POSTSECONDARY EDUCATION, TRAINING,

AND LIFE-LONG LEARNING

THURSDAY, APRIL 29, 1999

CAROLINE M. HOXBY

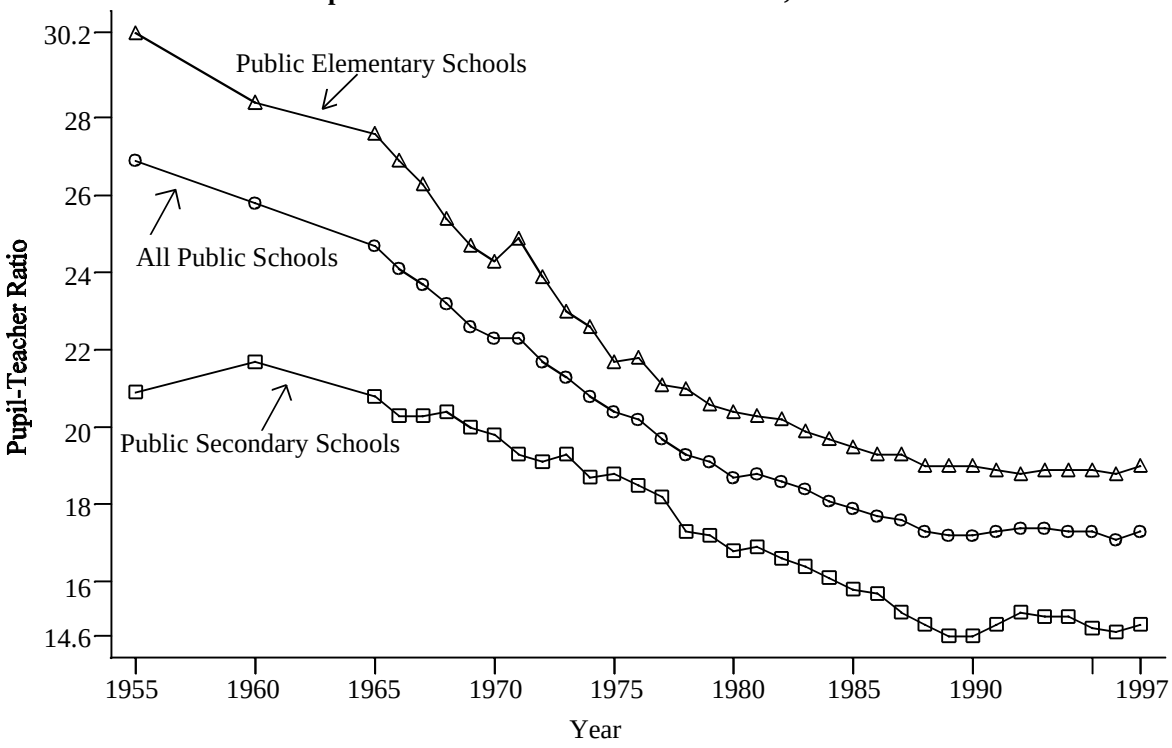
ASSOCIATE PROFESSOR OF ECONOMICS, HARVARD UNIVERSITY

CONTENTS

I. Introduction

One of the most commonly held beliefs about education is that students in smaller classes learn significantly more. As a result, reducing class size has been, for many years, the policy most widely used by American school districts that are attempting to improve student achievement. As a result of the widespread use of class size reduction policies, the pupil-teacher ratio has fallen in public schools from 27 in 1955 to 17 in 1997.[1] In other words, the pupil-teacher ratio in American public schools was 55 percent higher in 1955 than it is today. Figure 1 shows the fall in pupil-teacher ratios in American elementary and secondary schools since 1955.

Figure 1
Pupil-Teacher Ratios in American Schools, 1955-1997



The common belief in the efficacy of class size policies is not shared generally among researchers. In fact, it would be accurate to describe class size policies as highly controversial

among researchers, who disagree about whether reducing class size actually improves student achievement. In addition, it is unclear how to define class size. Parents usually think of the number of students in their child's regular classroom, but policies that attempt to reduce class size generally offer inducements for schools to hire more teachers. As a result, the policies usually lower the pupil-teacher ratio but do not always lower the size of regular classes. The additional teachers are often absorbed by special or administrative duties. In this testimony, I will focus primarily on evidence regarding the question of whether class size reductions improve student achievement. I will also comment, however, on the implementation of class size policies and on the question of whether teachers "matter" to student achievement.

Class size reduction is controversial among researchers because there are literally hundreds of studies, based on data from the federal and from state departments of education, that suggest that common beliefs about class size are not true. These studies, which generally show little or no effect of reducing class size, have been surveyed by Professor Eric Hanushek of the University of Rochester and, more recently, by Professor Julian Betts of the University of California-San Diego.[2] Some researchers are skeptical of older studies because students do not arbitrarily or randomly end up in smaller classes, as they would in a scientific experiment. Instead, we usually find small classes in schools that serve students from affluent families, sometimes find small classes in rural schools, and sometimes find small classes serving students who are difficult to educate. It is difficult to disentangle the effects of class size from the effects of growing up in an affluent family or the effects of being a difficult student. As a result, the research community is increasingly focused on studies that are based on *experiments*, either explicit policy experiments (such as Project Star) or natural experiments, in which students end up in smaller classes as the

result of arbitrary assignment.

II. Evidence Based on a Natural Experiment (1035 Schools in 165 Districts over 20 Years)

I recently completed the most comprehensive study that is based on a natural experiment. The data for the experiment cover every elementary school in the state of Connecticut and include all of the last 20 years.[3] The natural experiment was very simple. In every school attendance area, there is some random variation from year to year in the number of children whose birthdays make them eligible to start kindergarten. A family might find its older daughter in a first grade cohort of 50 children but find her sister, who is just one or two years younger, in a first grade cohort of only 36 children. Such random variation produces random variation in class size. If the elementary school ordinarily has two classrooms per grade, the older daughter is likely to end up in classes that average 25 students, while her sister is likely to end up in classes that average only 18 students. My study uses exactly this sort of random variation to analyze the effects of class size on student achievement.

The study also uses another sort of random variation which occurs when a cohort hits a maximum class size threshold. Suppose that an older brother is in a cohort of 44 students, his younger brother is in a cohort of 45 students, and the district has a maximum class size of 22. Then the older brother will end up in a class of 22 students (44 students divided between two classrooms), but the younger brother will end up in a class of 15 students (45 students divided between three classrooms, to avoid violating the maximum class size rule). I analyze the effects of the large differences in class size experienced by such students, who are otherwise in similar circumstances. In the study, student achievement is measured by Connecticut's state-wide

mastery tests, which are administered to fourth, sixth, and eighth graders.[4]

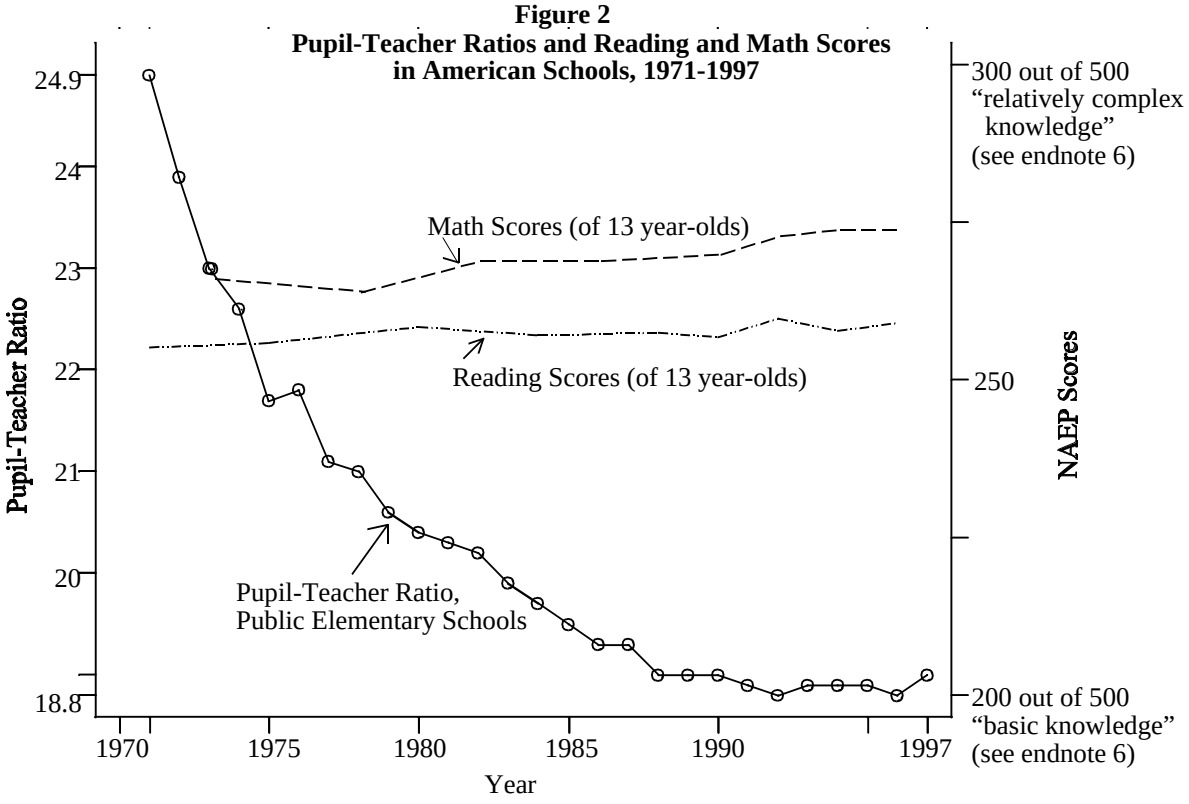
The study has several advantages. First, it is comprehensive: every elementary school in the state is included, regardless of whether it is urban, suburban, or rural. Second, the results are not confounded by factors like the students' background. The results depend on students who attend the *same* school just one or two years apart, so that the families who surround the school are the same. Third, I observe variation in class size that covers the entire policy range. I estimate the effects of reducing classes to sizes as small as 10 students and as large as 30 students. Fourth, because the data are very comprehensive, the results are precise so that I can look for tiny changes in student achievement on Connecticut's statewide tests. I can also look for effects just in schools that serve disadvantaged students.

The results of the study are easy to summarize. There appear to be *no* effects of class size reductions. This is true even though I would be able to identify statistically very small improvements in student achievement. For instance, if reducing class size by 2 students from a base of 20 students (a 10 percent reduction) had improved test scores by just 2/100^{ths} of a standard deviation, I would have found the improvement. The "no effect" result is the same even if I focus on just the early grades (grades one through three), just schools that serve disadvantaged students, just reductions in the size of classes that would otherwise be large (above 25 students), or just students who are in small classes throughout their entire elementary school experience. The "no effect" result remains even when I use students' birthdates to eliminate the problems that might be caused by parents who send their child to private school or search for a public school that has a small cohort in their child's grade.[5]

III. How to Reconcile the Conflicting Evidence on Class Size

How does one reconcile evidence from the natural experiment with evidence from Project Star, which suggest that class size reductions generate small, but statistically identifiable, improvements in student achievement? Moreover, how does one reconcile the evidence from Project Star with American history of the last 30 years--a period in which class size reductions have been enacted again and again but student achievement has been flat? Figure 2 juxtaposes pupil-teacher ratios and reading and math performance in American schools since 1971. (Reading and math are measured by the National Assessment of Educational Progress for 13 year-olds, whose scores should reflect their elementary school experience).[6] The figure shows that the pupil-teacher ratio in public elementary schools fell from 25 in 1971 to 19 in 1997 (a 24 percent

Figure 2
Pupil-Teacher Ratios and Reading and Math Scores in American Schools, 1971-1997



decrease), but that reading and math scores of 13 year-olds rose by less than 10 points on a 500-point scale (less than 1/10th of a standard deviation) over the same period. In other words, the historical evidence suggests that lowering class size has little effect on student achievement.

I believe that we can not only reconcile the different evidence, but can learn how to implement better policies in the process of reconciling the evidence. What makes Project Star different from typical policies of class size reduction? It was, one, an explicit experiment in which the policy was overtly evaluated and, two, an experiment upon which future funding for class size reductions depended. We know from scientific studies that people tend to be more productive when they know they are in an experiment that is being evaluated, regardless of whether the policy actually works. This is a placebo effect, often called the "Hawthorne effect," and it may account for the some of the differences between Project Star evidence and evidence based on natural experiments or recent history.[7] Far more importantly, Project Star contained *public evaluation, accountability, and incentives*. In other words, Project Star was not just money for class size reduction. It was money plus incentives that depended on *students'* performance. It is hard to overemphasize how rarely class size policy are combined with evaluation, especially on the basis of student performance. It is even more rare to find a policy that includes incentives, even just a threat that future classes will be larger if there is no evidence that smaller classes improve achievement. While Project Star did not give teachers special training for working with smaller classes, teachers and their administrators were certainly aware of the fact that they were expected to make good use of the opportunities presented by smaller classes.

In contrast, the natural experiment is more like typical class size policies. Teachers have smaller classes, but they are not aware of being evaluated and there is no link between their

students' achievement and their future enjoyment of smaller class sizes. In other words, they have no incentives. It is the norm in American educational policy to increase schools' resources without any accountability provisions. This may explain the pattern of recent history--increasing spending on schools with little effect on achievement.

There are a number of ways to give incentives to schools, including centralized evaluation (as in Project Star), explicit competition among schools for funding (this is used in Israel), local accountability systems, and incentives that come from parents' being able to choose among schools.[8] Developing countries now often enact policies that link school resources to evaluation and accountability, and it is reasonable to think that what is practicable there should be more than practicable here. In summary, the evidence would make me expect very little from any class size policy that did not include *some* incentives based on student performance. The ideal class size policy would not only include incentives, its continuance would be contingent on success, it would generate data that outside researchers could use for evaluation, and it would encourage administrators to train or advise their teachers to make good use of smaller classes.

IV. Important considerations related to class size reduction.

There are three other important facts that one should take into account when implementing a policy to reduce class size.

First, class size reductions are expensive. It is widely understood among researchers that the cost of a class size reduction is nearly proportional to the percentage reduction in class size. It is fairly accurate to say that a 10 percent reduction in class size requires almost a 10 percent increase in per-pupil spending in the long run. This is because teachers, school buildings, and

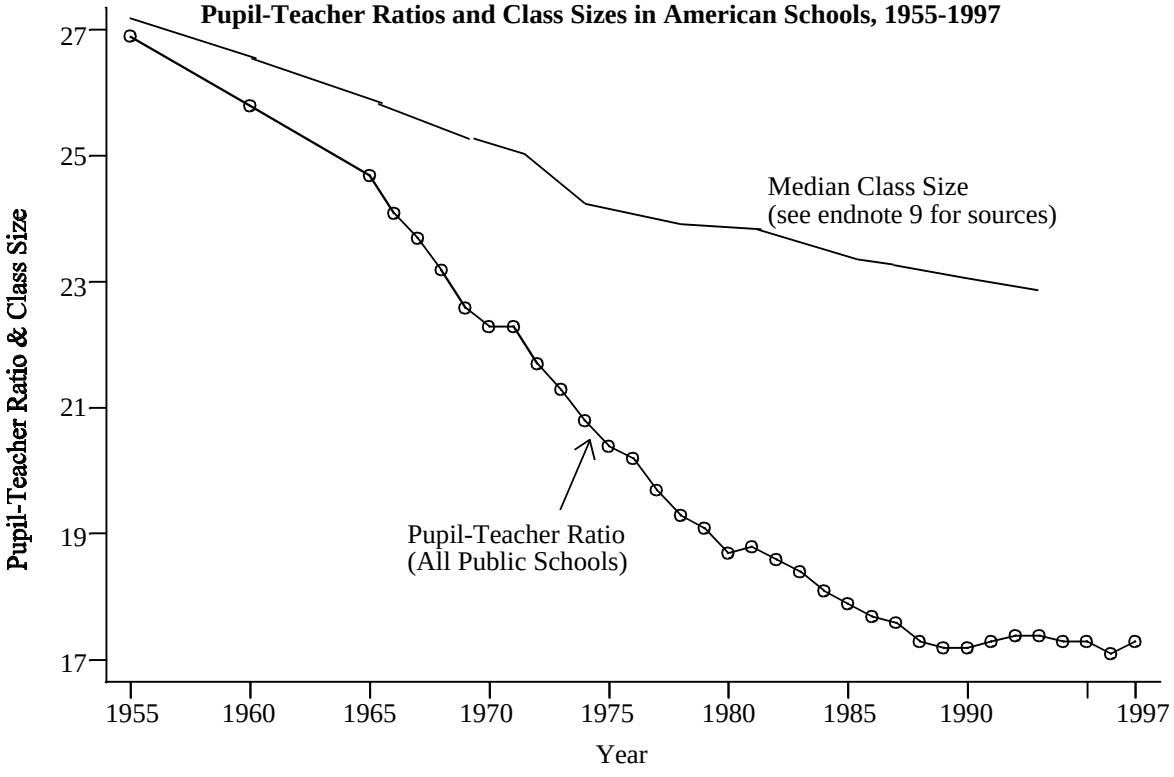
administrators increase in proportion to the number of classrooms. As a result, reducing class size from 20 students to 18 students requires about a 10 percent increase in spending, reducing class size from 22 students to 18 students requires about a 18 percent increase in spending, and reducing class from 24 students to 18 students requires about a 25 percent increase in spending. (It is worth noting that, because costs are related to the *percentage* reduction in class size, a 2-student reduction is more expensive when the initial class is smaller.) These spending increases are sufficiently large that every school should very carefully consider alternatives to class size reduction. Many proposed reforms that are dismissed as being too expensive are not more expensive than class size reduction.

Second, if a class size reduction is implemented everywhere at the same time, it will not have the same effects as Project Star, even if it includes the same incentives. The reason is that the demand for new teachers and classrooms cannot be met immediately without some deterioration in teacher quality and classroom quality. California's 1996 class size reduction produced evidence of the problems. There, students were often moved to portable classrooms and affluent districts "raided" good teachers from districts that serve disadvantaged students, resulting in a deterioration in teacher quality in the very schools where good teachers and smaller classes are thought to be most useful. (California also reduced certification requirements for new teachers, but such a change is not contemplated under the "100,000 teachers" plan.)

Third, as an historical matter, class size policies have been more successful at changing the pupil-teacher ratio than at lowering class size. The pupil-teacher ratio in a school is easy for the federal government to monitor, but class size in regular classes is difficult for the federal

government to monitor. If the intention is actually to lower class size, rather than just increase the number of teachers (who may be absorbed by duties other than regular classroom teaching), provisions should be made for monitoring class size. One of the reasons why parents continue to complain about their children's class sizes even though the pupil-teacher ratio has fallen is that class size has apparently fallen more slowly. Figure 3 shows pupil-teacher ratios and median class size for American public schools from 1955 to 1997. In 1955, the average pupil-teacher ratio was approximately the same as median class size: about 27 students. But, by 1994 (the last year for which accurate estimates can be computed), median class size was 23 students even though the pupil-teacher ratio was 17.3.[9] Special education accounts than about one-quarter of the "missing" teachers.[10] The other "missing" teachers are accounted for by non-classroom duties and reductions in the number of hours per day that teachers teach.

Figure 3
Pupil-Teacher Ratios and Class Sizes in American Schools, 1955-1997



VI. Do Teachers Matter?

Smaller classes provide opportunities for teachers to improve achievement, but a teacher must be effective if she or he is to actually improve achievement. This leads to the question, "Do teachers matter?" The answer has two parts. Yes, teachers as *individuals* do matter. A good teacher produces significantly greater learning in class after class of students. This has recently been definitively demonstrated with extraordinarily comprehensive data from the state of Texas.[11] The best teachers raise student achievement by *at least* 7 percent (compared to the worst teachers). Unfortunately, there is far less evidence that supports the idea that policies, such as salary increases and teacher testing, actually can deliver better teachers by guaranteeing that people with the ability to be good teachers engage in and stay in teaching. Studies of increases in teacher salaries and increases in teacher testing have produced almost no credible evidence of effects.[12]

Again, the difficulty is that most evidence is not based on either an explicit or a natural experiment. Higher teacher salaries are generally found in districts that serve students from affluent families and sometimes found in districts where students are hostile. Thus, it is difficult to disentangle the effects of higher salaries from the effects of growing up in an affluent family or being a hostile student. Teachers who score well on examinations are found in districts that serve affluent families, so that estimates of higher teacher scores on student achievement are biased upward. Studies that analyze states that have raised their passing standards for teaching credentials tend to find little or no effect on student achievement, even though states usually combine higher standards with teacher salary increases.

We will only obtain good evidence of what policies (if any) affect teacher quality when

researchers find data from explicit experiments or natural experiments. Any government that funds such an experiment and produces data would certainly receive considerable attention from the research community.

In summary, there is strong evidence that teachers, *as individuals*, matter, but it is hard to find evidence that salary and testing policies matter. These facts suggest that it is worthwhile designing policies to get better potential teachers to enter and stay in teaching, but that it would be advisable to have evaluation, accountability, and incentives built into these policies.

Endnotes

1. See Table 64, *Digest of Education Statistics*, published by the U.S. Department of Education, 1997.

2. See E. Hanushek, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24 (1986), 1141-77. See also J. Betts, "Is there a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature," in Gary Burtless, ed., *Does Money Matter? The Link between Schools, Student Achievement, and Adult Success*, Washington, DC: The Brookings Institution, 1995.

3. C. Hoxby, "The Effects of Class Size and Composition on Student Achievement: New Evidence from Natural Population Variation," National Bureau of Economic Research Working Paper No. 6869, 1998. The study can be obtained online at www.nber.org or www.economics.harvard.edu/faculty/hoxby/papers. The study may also be obtained by writing to the author at choxby@harvard.edu or Professor Caroline Hoxby, Department of Economics, Harvard University, Cambridge MA 02138.

4. Also, there was a 9th grade test prior to 1986, and there is currently a 10th grade test. Students are compared to other students from the same school who took the same test.

5. The actual cohort size can be instrumented by the size that the cohort would be if all the students whose birthdates made them eligible for a cohort actually did attend public school in that cohort. See the study for details.

6. National Assessment of Educational Progress (NAEP) scores are on a scale that ranges from 0 to 500. On the reading test "a score of 200 implies an ability to understand, combine ideas, and make inferences based on short uncomplicated passages about specific or sequentially related information" and "a score of 300 implies an ability to find, understand, summarize, and explain relatively complicated literary and informational material." On the math test, "performers at the 200 level have considerable understanding of two-digit numbers and know some basic multiplication and division facts" and "performers at the 300 level can compute decimals, simple fractions and percents[;]....,can identify geometric figures, measure lengths and angles, and calculate areas of rectangles." The source is the U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress, NAEP Trends in Academic Progress, by Educational Testing Service.

7. See H. Parsons, "Hawthorne: An Early Organizational Behavior Management Experiment." in *Pay for Performance: History, Controversy, and Evidence*, B. Hopkins and T. Mawhinney, eds., New York, London, and Sydney: Haworth Press, 1992, p 27-43.

8. An Israeli system that makes schools compete for funds is described by V. Lavy, "Evaluating the Effect of Teacher Salary Bonuses on Student Achievement", Hebrew University mimeo.

9. Class size must be estimated based on teachers' reports or administrators' reports. For recent years (1986 to the present), I base the estimates on teachers reports of their class sizes in the *Schools and Staffing Survey*, published by the U.S. Department of Education. For earlier years, I base the estimates on administrators' reports that are summarized in the biennial reports published by individual states' departments of education.

10. See E. Hanushek and S. Rivkin, "Understanding the Twentieth-Century Growth in U.S. School Spending," *Journal of Human Resources*, 32.1(1997), pp. 35-68.

11. See E. Hanushek, J. Kain, and S. Rivkin, "Teachers, Schools, and Academic Achievement," National Bureau of Economic Research Working Paper No. 6691, 1999. See also NBER Working Paper No. 7082 by the same authors. These papers are available online at www.nber.org.

12. An excellent general reference on the effects of salary and credential policies is R. Murnane, J. Singer, and J. Willet, *Who will teach?*, Cambridge: Harvard University Press, 1991. Unfortunately, although this book contains evidence on how changing salary and testing policies affect career decisions of teachers, it does not present evidence on how student achievement is affected.