

CROWD

An international look at the relationship between class size

REDUCING CLASS SIZES IS ONE OF TODAY'S MOST popular education reform strategies. The Education Commission of the States estimates that such efforts cost states \$2.3 billion during the 1999–00 school year alone. The federal government contributed another \$1.6 billion in 2000–01 toward meeting the Clinton administration's goal of decreasing class size nationwide in the early grades to no more than 18 students. During the past year or so, the deteriorating condition of state budgets and the Bush administration's new emphasis on accountability have made class-size reduction less of a priority. Yet it remains popular among parents, teachers,

and the teacher unions, which often promote it as an alternative to vouchers.

The motivation for reducing class size is intuitive: with smaller classes, teachers should be able to devote more time to each student, both in the classroom and in giving feedback on homework and tests. The concern is at least threefold. First, reducing class size is remarkably expensive, since it requires hiring more personnel. There may be less costly reforms that are at least as effective as class-size reduction. Second, hiring more teachers may dilute the quality of the workforce, thereby negating any gains among the students of good teachers.

PHOTOGRAPHS BY GETTY IMAGES



Control

and student achievement by MARTIN R. WEST & LUDGER WOESSMANN

Finally, the intuitive relationship between class size and teachers' effectiveness may not actually hold true—teachers may be no more successful with 18 students than with 23.

The most persuasive evidence of the benefits of class-size reduction has come from the Project STAR (Student/Teacher Achievement Ratio) experiment in Tennessee, where students were randomly assigned to classrooms of varying size. Smaller classes appeared to yield substantial gains among kindergartners and possibly 1st graders in the first year of the program—gains that were maintained throughout their school years. However, a large body of research literature on

class-size reduction contradicts the findings from Project STAR.

To lend a fresh perspective on this issue, we use data from the Third International Mathematics and Science Study (TIMSS) to compare the effects of class size around the world. While Americans squabble over whether class size should be 18 or 25 students, teachers in Korean schools routinely face classrooms of more than 50 students. These and other differences, such as the quality of a nation's teachers, can be valuable tools in discerning where, if ever, class-size reductions are likely to be beneficial.



Two Strategies

Ascertaining the effect of class size is less straightforward than it might appear. The central problem is that students are not assigned to classrooms randomly. For instance, schools often establish small remedial classes for lagging students or small enrichment classes for the so-called gifted and talented. In addition, school systems may direct students into schools with different average class sizes on the basis of their performance.

Parents also may influence their children's class sizes. They may work hard to move their children to schools with smaller classes, where they are likely to receive more attention. Thus variation in class size may be simply the result, rather than the cause, of differences in student achievement. Estimating the true effect of class size on student performance requires a strategy that looks only at variations in class size that are unrelated to students' previous achievement.

In principle, two such strategies are available. The first is to conduct a randomized field trial along the lines of Project STAR in Tennessee. Unfortunately, while it used a powerful research design, the Tennessee study was flawed in its imple-

In just two countries, Greece and Iceland, did smaller classes appear to elicit superior student performance.

mentation. For one thing, no data were collected on students' performance before they were assigned to their classrooms, making it impossible to know whether the assignment was truly random. In addition, the teachers were aware of their participation in Project STAR, as in almost any true experiment. This has led some to question whether its findings can be expected to hold under more typical conditions. It is also worth noting that the evidence here comes from an experiment conducted in a single U.S. state during the mid-1980s, in which classes were reduced from 22–25 students to fewer than 17. In that sense,

the findings may not apply to school systems in other parts of the world.

The second strategy, quasi-experimental research, relies either on special types of variation in class size or on econometric techniques to make appropriate comparisons. However, the conditions that must be met in order to use this approach make credible quasi-experimental studies possible for only a small number of school systems. For example, Princeton economists Anne Case and Angus Deaton used data on black students in South Africa during apartheid to measure the effects of class size. They argued that the black population of South Africa during this time lacked the power to influence class sizes, making the assumption that students were randomly assigned to class-

rooms of different size more plausible. But the South African school system under apartheid was obviously unique; in some districts, the average class size reached 80 students.

While Case and Deaton found that smaller classes were modestly beneficial, Harvard economist Caroline Hoxby's careful quasi-experimental study of elementary schools in Connecticut suggests that Case and Deaton's results may not be relevant for more developed countries. Hoxby analyzed variation in class size due to random fluctuations in the number of births and restrictions on maximum class sizes. She found no evidence of even trivial class-size effects. However, her approach requires a long panel of rich data and has yet to be applied in other contexts.

International Evidence

Taking data from TIMSS, we used a quasi-experimental design to take a broader look at how class size affects student achievement in different nations around the world. Conducted in 1994–95, TIMSS was the largest international study of student

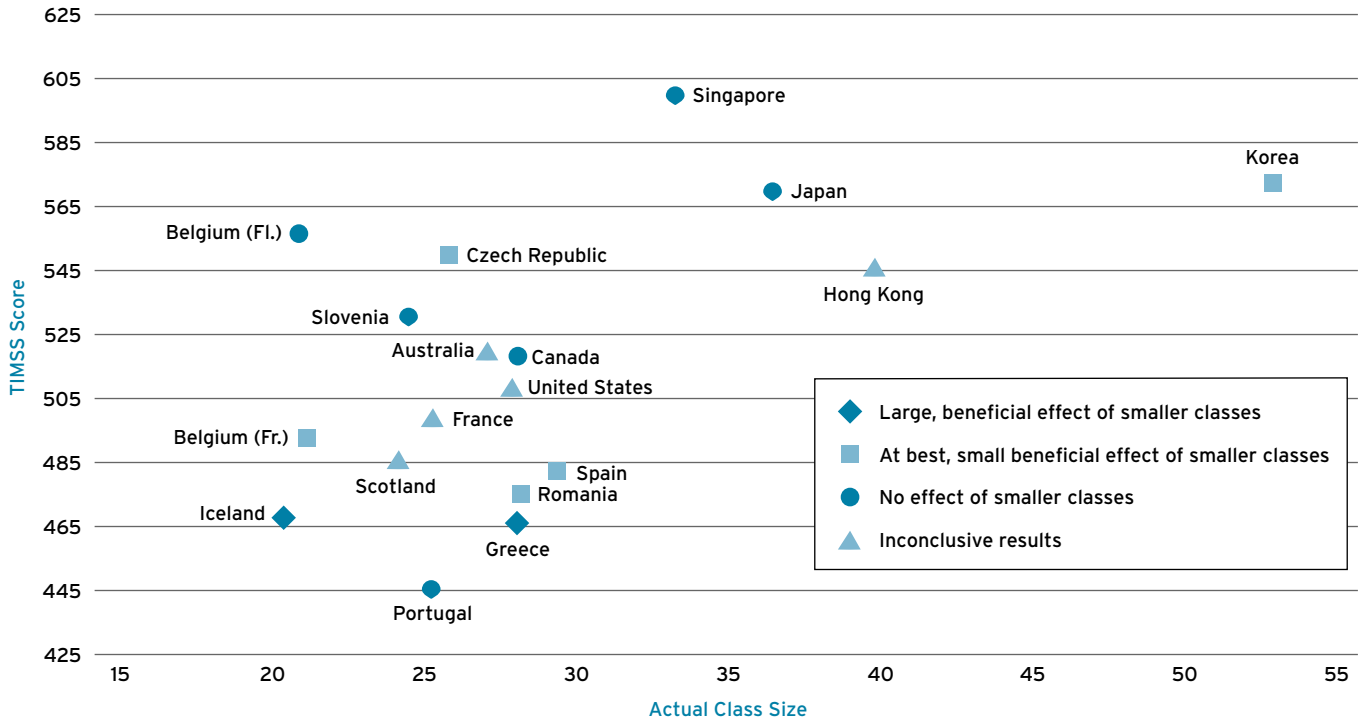
performance ever, with more than 40 countries participating initially. Each country administered the test to a nationally representative sample of middle-school students, defined as those students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing (grades 7 and 8 in most countries).

Our strategy takes advantage of the fact that data were collected on both actual and average class sizes and on students' performance and socioeconomic backgrounds for more than one grade level in each school. We looked at whether 7th

Few Benefits from Smaller Classes (Figure 1)

In just 2 of 18 countries, Iceland and Greece, were there large, beneficial, statistically significant effects of smaller classes. Large effects can be ruled out in 11 of the 18 nations.

Average Class Sizes and Test Scores in Grades 7 and 8



SOURCE: TIMSS

graders in a particular school performed better than the same school's 8th graders (relative to the national average for their respective grades) when, on average, the 7th-grade classes were smaller than the 8th-grade classes. With this strategy, the variation in class size we considered is strictly a consequence of fluctuations in the cohort size from one grade to the next. This excludes variation in class sizes within the same grade and from school to school, both of which can be subject to the influence of parents and school-system policies that tend to sort students into classrooms by their performance. The remaining differences should be essentially unrelated to student performance.

This approach forced us to restrict the sample to schools in which both a 7th-grade and an 8th-grade class were actually tested and in which data on the actual class sizes and average class sizes were available for each grade. We ultimately conducted our analysis on the 18 countries in which data for at least 50 schools in both mathematics and science remained after applying these criteria.

As shown in Figure 1, Portugal exhibits the lowest average combined test scores in math and science among the 18 countries in our sample, Singapore the highest. Iceland has the

smallest average class size, with just 20 students per classroom. At nearly 53 students per class, Korea has by far the highest average. The other East Asian countries also feature large classes, with an average of more than 30 students. In general, the countries with the smallest classes tended to be the worst performers. The reverse is also true: high performers tend to have larger classes. While this does not say much about the effectiveness of reducing class sizes in various environments, it does demonstrate that it is possible to have a high-achieving school system with relatively large classes.

Results

Let's look first at the results of a straightforward comparison that adjusts the data on student performance for students' socioeconomic background and grade level (since 7th and 8th graders were tested), thereby attempting to isolate the effects of class size. This initial analysis is of interest primarily because it is analogous to the approach used in most research on class size. Comparing these results with those obtained by a more reliable strategy will provide an indication of what biases may



East Asian countries feature large classes, with an average of more than 30 students.

exist in other studies.

In 11 of the 18 nations, the estimate of the effects of class size were positive and statistically significant, suggesting that students in larger classes perform significantly better than students in smaller classes. In other words, a naïve strategy that does not account for the ways in which students are sorted into classes of different size leads to the counterintuitive result that students fare better in larger classes. Moreover, this result seems universal: it emerges in western Europe (Belgium, France), eastern Europe (Czech Republic, Romania), Australia, and East Asia (Hong Kong, Japan). No country showed students in smaller classes outperforming their peers in larger classes.

Let's turn now to the preferred strategy, which controls for the fact that students performing at different levels may be

to be substantial: Students scored just over two points (or 2 percent of the international standard deviation) higher for every one student fewer in their class.

Precision Testing

What can be learned from the 16 countries where the results were statistically insignificant? Does this suggest the lack of a causal relationship between class size and student performance? Or is it merely the result of statistical imprecision? In four of the countries, Australia, Hong Kong, Scotland, and the United States, the standard error of the estimated effects of class size was extremely large, indicating that little confidence should be placed in the results. The lack of precision in these cases seems to be a direct consequence of our research strategy's rather demanding data requirements. These school systems simply exhibit little variation in average class size from one grade to the next—the type of variation on which our strategy relied.

The remaining 12 countries can be further distinguished by comparing their results with those from other studies. We chose first to compare our results with those reported by Princeton economist Alan Krueger in his reanalysis of the Project STAR data from Tennessee, which produced some of the highest estimates of class-size effects among credible studies. Krueger performed a very rough cost-benefit analysis, in which the eco-

This evidence suggests that capable teachers are able to promote student learning equally well regardless of class size.

sorted into smaller or larger classes both between and within schools. The first notable feature of this approach is the disappearance of the counterintuitive result that students do better in larger classes. In 16 of the 18 countries, none of the results was statistically different from zero. In the other two countries, Greece and Iceland, smaller classes did appear to elicit superior student performance. Moreover, the benefits appear

Low-Quality Teachers? (Table 1)

Iceland and Greece maintain both low class sizes and low school spending by paying their teachers poorly, suggesting that smaller classes may be beneficial only when teachers are of low quality.

	Countries*	Average Class Size	Average Score on TIMSS	Per-Pupil Spending	Average Teacher Salary	Teacher Salary/Per-Capita GDP	Share of Teachers with Each Level of Education				
							a.	b.	c.	d.	e.
Countries with large, beneficial effects of smaller classes	Greece Iceland	24	467	\$2,374	\$16,311	1.0	16%	5%	45%	32%	1%
Countries where smaller classes evince, at best, a small, beneficial effect	Belgium (French) Czech Republic Korea Romania Spain	31	514	\$3,478	\$27,496	1.8	2%	19%	15%	35%	27%
Countries where smaller classes have no effect	Belgium (Flemish) Canada Japan Portugal Singapore Slovenia	28	537	\$5,667	\$29,038	1.8	2%	32%	7%	55%	5%

*The results for Australia, France, Hong Kong, Scotland, and the United States were inconclusive

a. Training but No Secondary Education b. Secondary Education c. B.A. d. B.A. + Training e. M.A. + Training SOURCE: TIMSS and OECD

conomic benefits of class-size reduction, in terms of the increase in future earnings due to higher test scores, appeared to approximate the costs.

Krueger’s results indicate that students in kindergarten classrooms that had 7 to 8 fewer students than regular-sized classes performed about 3 percent of a standard deviation better for every one student fewer in their class. Converted to international scores on TIMSS, this is equivalent to three test-score points. This is greater than the two-point gain we found in Iceland and Greece, but it is within the standard error of these estimates, suggesting that the actual effect of reducing class size in Iceland and Greece could be as large as Krueger found in the United States.

For 11 of the 12 countries with relatively precise yet statistically insignificant estimates, the possibility of class-size effects of the same size as Krueger found can be rejected with at least 95 percent confidence. There could still be class-size effects in these nations, just not of the magnitude estimated by Krueger. Note, however, that Krueger’s effects were found in

kindergarten and 1st grade, while these estimates are for students in 7th and 8th grades.

We further tested to see whether a one-student reduction in class sizes would increase TIMSS scores by just one point, or 1 percent of an international standard deviation. An effect of this size would be so small as to be essentially negligible from the standpoint of public policy; a one-point gain is too little to justify the expense of class-size reduction. Regardless, even the possibility of this small an impact can be rejected with at least 90 percent confidence in 6 of our 12 school systems with reasonably precise results.

In short, the effect of class size on student performance varies across the 18 countries in our sample (see Figure 1). We can rule out even a minimal relationship between class size and TIMSS scores in the middle grades in six school systems: those of Flemish Belgium, Canada, Japan, Portugal, Singapore, and Slovenia. In an additional five school systems, we can rule out the possibility of large class-size effects: French Bel-



The evidence suggests that capable teachers are able to promote student learning equally well regardless of class size.

gium, the Czech Republic, Korea, Romania, and Spain. These results cast doubt on the desirability of class-size reduction in the middle grades as a reform strategy in many countries. In Greece and Iceland, by contrast, smaller classes were clearly beneficial. (In five countries—Australia, France, Hong Kong, Scotland, and the United States—our strategy led to inconclusive estimates that do not allow for any confident assertions about the effects of differences in class size.)

Quantity versus Quality

Why would class-size reduction elicit improvement in Greece and Iceland but not elsewhere? One might expect class-size effects to be related to such characteristics as a nation's overall level of resources. For instance, it is feasible that countries with relatively large classes would glean substantial benefits from reducing class sizes. However, there is no clear pattern in countries' average class sizes that distinguishes the two countries where substantial class-size effects exist from either the six countries where we ruled out any noteworthy class-size effects or from the five countries where we ruled out at least large class-size effects. Greece's average class size is similar to the mean class size among the nations where no class-size effects were found, and Iceland's average class size is substantially lower (see Table 1).

One possibility is that class-size reduction has a large impact in relatively ineffective school systems. Both Greece and Iceland performed considerably below the international average on TIMSS, while the countries where class-size reduction did not have even a small effect performed above the average. Also, even though Greece's class sizes are roughly at the mean and Iceland's were substantially lower than the mean, education spending per student in both countries is substantially below the average of the two comparison groups. This suggests that Greece and Iceland spend rather little per employed teacher, which is reflected in the data on teachers' salaries. Teachers' salaries in Greece and Iceland are below the mean of the other countries in absolute terms, in terms of salary per teaching hour, and relative to the country's per capita GDP (see Table 1).

A low average salary for teachers suggests that a country may

be drawing its teaching population from a pool of less-skilled workers. If this is the case, different countries appear to be making different tradeoffs between the quantity and quality of their teachers: with class sizes low, Greece and Iceland employ many teachers of low quality. The countries where class-size effects were not observed appear to employ relatively fewer teachers, but of higher quality.

This assumption is borne out by the available data on teachers' educational attainment. In Greece, the highest level of education reached by the vast majority of teachers is the equivalent of a bachelor's degree without any teacher training. In Iceland, about one-third of the teachers surveyed by TIMSS had not even completed secondary education, with only some basic teacher training. Meanwhile, about 60 percent of the teachers surveyed in the other countries held either a bachelor's or a master's degree in addition to their training as teachers.

This evidence suggests that capable teachers are able to promote student learning equally well regardless of class size (at least within the range of variation that occurs naturally among grades). Less capable teachers, however, do not seem to be up to the job of teaching large classes.

This interpretation is corroborated by teachers' responses in TIMSS when they were asked to what extent their teaching was limited by a high student-to-teacher ratio in their classroom. In Greece and Iceland, 45 percent of teachers reported that their teaching was limited "a great deal" by a high student-to-teacher ratio. The comparable statistics averaged only 19 percent and 25 percent among countries where no class-size effects and no large class-size effects were found, respectively. This is despite the fact that average class sizes in Greece and Iceland were lower than in either comparison group.

In short, our evidence suggests that the existence of class-size effects is related to the quality of the teaching force. Smaller classes appear to be beneficial only in countries where average teacher quality is low. If teacher quality is a key input in education, this interpretation can explain why class-size effects exist in some countries but not in others and at the same time why the countries in our sample where we did find sizable class-size effects also exhibit poor overall performance. Greece and Iceland exhibit class-size effects and poor performance because they employ a population of relatively less capable teachers, while other countries exhibit no class-size effects but high overall performance because they employ good teachers. This suggests that it may be better policy to devote the limited resources available for education to employing more capable teachers rather than to reducing class sizes. The merits of this admittedly speculative conclusion are a promising topic for future research.

—Martin R. West is a research fellow at the Harvard University Program on Education Policy and Governance and the research editor of Education Next. Ludger Woessmann is a senior researcher at the Ifo Institute for Economic Research in Munich, Germany.