

Florida's Constitutional Amendment to Reduce Class Size: What Can Research Tell Us?

Dominic J. Brewer, Ph.D.

RAND Education

Introduction

In November 2002, the voters of Florida passed a constitutional amendment affecting the lives of millions of children, teachers and parents. With 52% approval, the measure amended Article IX, Section 1 of the state constitution, which specifies that the state has an obligation to provide a quality public education free of charge to its citizens. In this case “high-quality” is established through a very specific educational intervention—classes of a specified size. The state legislature is obligated to provide funding for all public schools to provide classes not exceeding 18 students for those in grades pre-K through 3, not exceeding 22 for students in grades 4 through 6, and not exceeding 25 in grades 9 through 12. The reductions from current levels will be phased in over several years.

What is this measure designed to achieve? What is the evidence on whether or not it will have the desired effects, or whether it is a cost-effective strategy for providing a high-quality education for all students? How could the legislature implement it in a way that maximizes the benefits and minimizes the costs? What have other states’ experiences been with similar approaches? Luckily, research evidence provides some—but inevitably not complete—answers to these questions. This policy brief provides an overview of that evidence.

Reducing class sizes is an approach to improving our schools that is simple and popular, but it generates furious debate among academics just as it does among policy makers. There is a large body of research on the topic, often by advocates or opponents, and much of it of dubious quality. Sifting through this evidence, however, does yield some fairly robust conclusions. The research evidence suggests that there is some basis for believing that small classes can lead to gains in student achievement, particularly class sizes in the 13-17 range for students in the early grades, with the biggest gains for at-risk students. However research also tells us that how class size reduction policies are designed and the conditions under which they are implemented are likely to be critical to whether or not they yield the expected benefits for students. And even if small class sizes lead to improved achievement, they may not be the most cost-effective means to bring this about.

The Research Evidence on Small Classes

What Is Class Size Reduction Intended to Achieve?

Improving public education has been a national priority for the past several decades for policy makers and educators at the federal, state and local level. The overriding goal since *A Nation at Risk* was published in 1983 has been to improve student achievement, to boost the “outputs” of schools and students. Although schools produce multiple outcomes, this typically translates into improvements in student scores on standardized achievement tests.

Research has repeatedly demonstrated that student achievement is determined by many factors. The most important of these are individual and family background characteristics (for example, the education and income levels of a child's household) and community characteristics (the composition of a child's peer group and neighborhood wealth, for example) that are typically outside the ability of policy makers to affect, at least directly. Schools are believed to have a much smaller influence, but are the main vehicle to effect improvement.

Many strategies directed toward boosting student achievement have been tried. These include: increasing resources for schools directly through smaller classes; smaller school sizes; more computers or higher teacher salaries; different pedagogical approaches and new curricula; improving teacher and leadership training and professional development; the establishment of clear curriculum standards and a system of accountability through student testing; and new providers of schooling such as charter schools and voucher schools that operate independently of the existing school district-operated public schools.

These alternative strategies have been applied in many different ways around the country. Some are quite complex and prove hard to actually implement. Most often, several types of approaches are introduced simultaneously or overlap. Often they do not last long, as policy makers change their approach in response to political considerations, or as funding ebbs and flows. Researchers have found it remarkably difficult to pin down which of the strategies is most effective.

Small class sizes are popular with students, their parents and with teachers. Reducing the number of students per class is easy to understand. There is no need for sophisticated measurements or for burdensome compliance mechanisms. Indeed, the approach has considerable "face validity"—i.e., it is easy to construct some plausible intuitive arguments for why class size might matter and therefore why reducing class sizes might improve student outcomes.

Why Are Small Classes Believed to Matter?

The number of students in a classroom with a teacher is believed to affect how much is learned in a number of ways. The primary idea relates to what teachers are able to do in small classes—either because they change what they do, or because some things they already do work better in small classes. Generally teachers use pedagogical strategies that devote most time to whole-class instruction. When there are fewer students per teacher, the teacher is able to focus greater attention on individual students, working with them one on one, understanding their learning needs and choosing appropriate pedagogical strategies to best serve them. Teachers can assign and examine more work and cover more material. It is believed that the result is fewer disruptions in the classroom.

Of course this view puts the teacher—and his or her ability to take advantage of the possibilities of a small class—at the center. Research on educational productivity has typically confirmed the importance of teachers in affecting student achievement (for example, see citations in Goldhaber & Brewer, 1997). Small classes may enable good teachers. Whether average or below-average teachers are more effective in smaller classes is an open question.

Interestingly, there is relatively little research that actually establishes which of these explanations is correct. Why? Establishing such relationships is inherently difficult. Observational studies—which often reveal little difference in what teachers do—tend to be small-scale and not generalizable. Larger-scale survey research is problematic because measuring teacher and student behaviors through this means is not always reliable. In addition, most research has focused on whether there is a relationship between class size and student achievement rather than why one might exist. Consequently, the research base on the mechanisms through which small classes might be important is thin.

In general, teachers seem to use the same instructional strategies when faced with different-sized classes, rather than change instructional content or the way the material is presented. There is some evidence, however, that teachers may spend more time on individualized instruction and less time on disciplining students in small classes. “Overall the weight of the evidence tilts strongly toward a conclusion that reducing class size, by itself, does not significantly affect the instructional activities that occur in classrooms” (Ehrenberg, Brewer, Gamoran, & Willms, 2001, p.23).

Do Small Classes Boost Student Achievement? The Tennessee STAR Study and Other Evidence

Mechanisms through which small classes affect student achievement may not matter so long as there is indeed a clear relationship. There have been hundreds of studies that have tried to establish this and they fall into two groups: STAR and non-STAR. STAR (or “Student Achievement Ratio”) was a demonstration project carried out in Tennessee in the 1980s, and the results of the effort changed the debate over the research evidence on class size. The reason was that STAR was the first and only randomized experiment on a moderate scale. Randomization is important because identical students are assigned to either a group that receives a clearly defined treatment or one that does not receive the treatment. This means that group differences observed later can reasonably be attributed to the treatment. Such conditions are very rare in educational research.

Non-STAR studies are typically of two types.

- First, there are a large number of small-scale experimental studies. These are often poorly designed (few have true randomization), have too few participants, are of short duration and have no rigorous evaluation upon which to have much faith in the

results. Even so, an oft-cited meta-analysis of this work conducted by Glass and Smith (1979) concluded that there were some positive benefits of classes below 20 particularly for early grades.

- Second, there are a large number of non-experimental studies. These try to exploit “naturally occurring” variation in class sizes and student performance and try to tease out the effect of the former on the latter, while holding constant all the other factors that influence student performance. Needless to say, these studies are particularly susceptible to data quality and statistical issues. (For example, they often rely not on a measure of class size but on teacher-pupil ratios, which are not the same thing.) Ehrenberg et al. (2001, p.15) in their review of this evidence concluded: “there are simply too many statistical problems...in these non-experimental studies to place great faith in any of the findings” (p.15).

STAR, in contrast to this other research, was a relatively well-done, randomized study that has produced a number of fairly robust conclusions. Indeed, it was the emerging evidence from STAR that led to an uptick of interest in class size-related policies in the 1990s.

STAR was a state-funded program, operated 1985-89, that randomly assigned students entering kindergarten to regular classes of 22-26 students, small classes of 13-17 students, or regular classes with a teacher's aide. Students remained in these small classes for four years. Schools and districts volunteered and were selected to participate but students and teachers were randomly assigned within a school. Teachers received no extra training or materials. Initially some 70 schools and 46 districts participated; by the end of study, the number of students had grown from 6,400 to 12,000.

An original study team and numerous researchers subsequently have examined data from STAR. Most agree that the teachers' aides made no difference to student outcomes. Most also agree with the view of Finn and Achilles (1999) that STAR showed “an array of benefits of small classes, including improved teaching conditions, improved student performance and, after the experimental years, improved student learning behaviors, fewer classroom disruptions and discipline problems” (p. 98). The consensus view is that:

- The benefit of a class in the 13-17 range versus a class in the 22-26 range is about a fifth of a standard deviation gain in student achievement for students in K-3. (If two students at the 50th percentile—i.e. half of the students were better and half were worse—were assigned to the two classes, one would be at the 58th percentile after being in the small class.)
- The benefit appears for students by the end of first grade.
- The benefit persists, maybe for many years afterwards; but is not cumulative (i.e., gains do not increase over time).
- The benefits are greater for minority students—two fifths of a standard deviation in

student achievement, which is generally considered a large educational effect.

The precise magnitude of the estimated effects differs somewhat from study to study based on the methodology the researchers use. There is also some disagreement about important issues—for example, whether the benefits may or may not be cumulative. Some researchers suggest the more years spent in smaller classes the longer the benefits last but others have claimed that the effects are not cumulative beyond first grade (see Hanushek, 1999).

Although the STAR evidence is quite strong and is undoubtedly the best we have on the underlying relationship between small classes and student outcomes, it is only one study. It is also important to note several important caveats.

First, any experiment must be conducted in a careful controlled way that is independently verifiable—with a clearly defined and well-measured treatment, longitudinal tracking of multiple outcomes, accurate randomization, no crossover between treatment and control groups, minimal attrition, and so on. In general, the STAR case holds up pretty well, though it is not perfect. For example, there is some evidence that students assigned to large classes switched to small ones, and there was considerable attrition (of the initial experimental group less than half remained for all four years). STAR was also designed with non-random selection of schools. However, despite these problems, the overall results seem robust and have been confirmed by many researchers.

Second, the STAR findings are not by themselves sufficient for designing class size policies because some key questions are unanswered (see Grissmer, 1999). For example: Is the relationship between class size and student outcomes linear? Do class sizes have to fall below a certain level? Does the impact vary across grade levels or subjects or for kinds of students? How does timing and exposure affect the size of the effects? What are the long-run consequences? How important is class size relative to other factors?

Third, experiments may have limited external validity. This means that one has to be careful in making inferences about the results of an experiment to a real world, large-scale application. In particular, the results demonstrated in STAR apply given the conditions that prevailed—for example, classes of 13-17 in grades K-3, random assignment of teachers, no lack of facilities space, full funding of the additional teachers, and so on. This does not necessarily mean one can't use the results in STAR to justify small class policies, but it does suggest considerable care must be given to try to replicate the conditions of the demonstration as closely as possible if one hopes to mimic the benefits. We now turn to this issue in more detail.

The Research Evidence on Class Size Reduction Policies

Despite compelling evidence to support small classes, translating this concept into policy has proven difficult. Indeed, the experience in recent years across the United States has revealed two important aspects of class size reduction policies. The first is design. There is not one policy that can be labeled “class size reduction.” Indeed, there are a multitude of approaches, reflecting the many dimensions upon which decisions can be made. The second is implementation. How a policy is implemented, and the conditions that exist when implementation occurs are important to understanding the likely impact of the policy. It is therefore important to realize that it is quite reasonable to conclude that the research evidence supports the notion that small classes may help student performance, but simultaneously be skeptical—again based on research evidence—that specific class size reduction policies are effective.

The Design Dimensions of Class Size Reduction Policies

There is no such thing as “a” class size reduction policy. It can be designed in many ways. A quick scan of the many states that have adopted class size reduction policies reveals a considerable amount of variety. A number of dimensions on which the policy can vary include the following, all of which may have important implications for benefits and cost and ease of implementation (Brewer, Krop, Gill & Reichardt, 1999).

- **Actual class size.** The most obvious difference is in the specified target class size. As noted above, the STAR study suggests gains in student achievement occur in the 13-17 range. Typically, however, practical constraints have led policy makers to adopt class sizes above this level. California is 20:1; the Florida initiative is 18:1 for grades K-3. We don't know if the same benefits can be achieved at these higher levels.
- **Measurement of class size.** Class size formally refers to the actual number of students in a given classroom, interacting with one teacher, on a given day. But the number of students in a classroom varies across time from the beginning of the school year to the end, and indeed even during the school day itself due to absences, pull-out programs and so on. Given this, if the target class size is enforced, inevitably students will typically be in classes below the target.
- **Flexibility in enforcement.** Policies can vary according to whether they require that every class in every grade in every school in every district be at the specified level or whether it is acceptable for the average class size to be at the target level across grades in a school, or across schools in a district. (If the policy is literally enforced class by class, a school with 19 students in each of grades 1, 2 and 3 would be forced to create a new class if it had an 18:1 target for each grade group—i.e., would have six classes total. A more flexible policy might allow the average across the grades to be 18 or less, which could be accomplished with just 4 classes.) Policies also can be mandatory or voluntary.

- **Grade levels and subjects covered.** Most class size reduction policies apply to the early grades, but some target other grades or specific subjects (e.g., literacy, mathematics).
- **Kinds of students targeted.** Some policies have been applied to all students (e.g., California) while others have been targeted to at-risk students, following the STAR results. Statewide policies of course do not give schools flexibility over what strategies they can use to meet individual students' instructional needs. In addition, the number of students included in the policy greatly affects the scale of the implementation and consequently the need for new teachers and facilities.
- **Funding.** Policies can be funded in many different ways. Given the multiple sources of funds that schools use to provide services to students, any funding at less than full cost (which is in any case difficult to specify in advance) may risk substitution of funds and programs.

Implementing Class Size Reduction Policies

Smaller classes are in one sense easier to implement than many educational reforms. It's relatively easy to count the number of students in a classroom, compared to figuring out whether a teacher is teaching a new curriculum, for example. However implementation is not necessarily straightforward because smaller classes require more resources—more teachers and more space. The success at obtaining these resources depends on many factors including the timing of the policy and the pre-existing conditions under which it is introduced.

Timing

First, since hiring new teachers and finding new spaces for classes takes time, an important element in class size reduction policies is the extent to which they are phased in. A gradual phase-in allows schools to locate suitable teachers, modify existing facilities (rather than use portables) or build new ones. Minimizing the disruption is likely to reduce harm caused to students and unintended consequences. Second, once teachers are hired and facilities are built, there is some rigidity associated with these resources. Teachers typically get tenure after 3 years and are hard to fire if enrollment drops or if a small class policy has to be reversed. Similarly, accommodating new classes increases facilities needs but closing schools tends to be unpopular. Class size reduction is hard to undo when state budgets turn sour. Recent efforts by states such as Oregon, Utah, Washington, Massachusetts, Tennessee—and now Florida—exemplify how difficult these reversals are politically.

Pre-Existing Conditions

Additional resources depend on how many new classes are needed which in turn depends on the class size target and the starting point. If one school is already at the

target, it can comply with the policy with no additional resources; a school with an average class size several students above the target will need to hire more teachers and find more space. Further, it is a fact of public education that all schools and all school districts do not begin from the same starting point. Specifically, districts that are further above the class size target and therefore need more resources to adopt the policy are likely to be those that are poorer and have more at-risk students (precisely the ones who may benefit most according to the STAR results). Similarly, these schools are likely to have the hardest time attracting high-quality teachers and sometimes have the most overcrowded facilities, making implementation of smaller classes problematic.

Design and Implementation:

The Contrasting California and Wisconsin Cases

These design and implementation factors have played out in many states over the past few years as many different versions of class size reduction have been tried. Two of the most interesting and contrasting cases have been well documented by researchers. Here is a brief summary of the findings.

California

In 1996 the state legislature introduced a class size reduction program affecting almost two million students in grades K-3 with a target of 20:1. Before the policy, the average class size in these grades in the state was about 28, so this represented a significant reduction. Priority was given first to grades 1 and 2, then to K and 3. The policy was passed just two months before the start of the school year. The program was voluntary—it provided funding of \$650 per student (now \$800) if they were in a “small” class—but almost all districts implemented the policy, which was popular with parents and educators. Over the past 5 years the state has spent more than \$8 billion on the program.

What have been the consequences of this huge effort? The program has been evaluated for the state by a consortium of independent research organizations with the following findings (Bohrnstedt & Stecher, 1999, 2002; Stecher & Bohrnstedt, 2000; Stecher, Bohrnstedt, Kirst, McRobbie & Williams, 2001):

- Gains in student achievement in reading, writing and mathematics have been found, although they are typically very small, on the order of one twentieth to one tenth of a standard deviation difference in third grade between small and large classes, and are the same for all types of student. This result has been found in each of the three years of the evaluation. But these achievement results are problematic because the program was implemented very rapidly, the state embarked on multiple reforms simultaneously, and it was introduced before statewide student testing, i.e., there were no baseline data. Researchers studying the achievement effects of the policy,

therefore, have to rely on variation in the implementation of the reform and statistical methods to tease out achievement effects. At best, then, it is probably reasonable to say that the evidence to date has not shown student achievement improvements that were hoped for, following STAR.

- There has been a dramatic decline in teacher qualifications statewide. Because the policy was across the board, there was a huge boom in teacher hiring in a state where there were already teacher shortages. Not surprisingly, the percentage of credentialed teachers, average years of experience, and so on, have declined in the early grades since the program began, particularly in the first two years.
- Since districts started from different class sizes (some at 20 and some as high as 33) and the policy was implemented uniformly, the rapid implementation, existing shortages, and the funding formula (which essentially provided a cash windfall to the richer districts that were already at or close to the target), greatly exacerbated the inequalities between rich and poor districts. The latter struggled to find qualified teachers.
- Since the funding was inadequate to cover full costs, there is evidence of some reallocation of resources away from other programs and services. For example, survey data suggest two thirds of California's districts have been forced to cut back funding from programs such as libraries and professional development to pay for smaller classes. Further, many districts experienced growing class sizes in grades 4 and 5 to compensate. Biddle and Berliner (2002) have described the California case as a "lesson in what not to do."

Wisconsin

In almost complete contrast, Wisconsin adopted in 1996 a highly targeted, phased-in class size reduction program as part of a reform called SAGE (Student Achievement Guarantee in Education). This five-year pilot program began in just 14 schools that had high concentrations of poor children, and reduced class sizes from 21-25 to 12-15. Molnar et al. (1999) have demonstrated that this program had effects on student achievement not dissimilar to STAR (a gain of about 0.2 standard deviations in achievement) with minority students gaining significantly more. Compared to STAR, this study was much smaller in scale and less rigorous (students not randomly assigned), but its findings are more recent and generally confirmed the STAR results.

**The Florida Initiative:
Applying Lessons
Learned From
Research**

What are we to conclude from the research on small classes and on class size reduction policies, for the Florida case? It seems reasonable to conclude based on STAR, that it is possible for small classes to produce achievement gains—but only under certain conditions. If, in a non-experimental setting, class size reduction is introduced, the Wisconsin SAGE program has shown it is possible to generate these gains. But the California case also points to a slew of potential negative effects that can unfold without careful attention to design and implementation.

Let us consider the various elements of Florida's amendment to reduce class size (2002).

First, the targeted class sizes do not match well with the research evidence on effects on student achievement. Very little is known about effects in grades 4 and above. For K-3, class sizes *below* 17 are believed to have an effect. At the proposed level of 18 in K-3, we cannot be sure of the likely effects.

Second, design details are critical—how precisely will class size be measured? How will the policy be enforced? And so on. All these design details could affect the likely benefits and costs of the policy.

Third, the constitutional amendment applies to all schools statewide and is written in a way that seems to imply very little flexibility—i.e., every class must be at the target level. As of 2002, the average Florida classroom held about 23 students through the 5th grade. Reducing down to 18 represents a significantly smaller reduction than occurred in California, but it is still large. In addition of course, the starting point varied tremendously across the state, with fast-growing counties like Broward and Miami-Dade in the 25-27 range. These districts already cannot find enough teachers or space, and the 2002 constitutional amendment will undoubtedly exacerbate these problems.

Fourth, the amendment provides a mechanism for phasing in smaller classes. This provides an opportunity to allow adjustments over time and should mitigate the likely negative effects of a hiring boom and facilities crunch. Florida's 8-year lead time is radically different from the California case.

Finally, the amendment is effectively silent on funding mechanisms because it puts the burden for full funding on the state legislature. If the state fully funds the implementation of the smaller classes, one might expect to see cutbacks in other programs or troubling distributional consequences. It is unknown whether political and economic realities, or the difficulty of accurately estimating the true full costs, will lead to funding that mitigates or promotes implementation difficulties.

Conclusions

The smaller classes envisaged by Florida's constitutional amendment (2002) seem unlikely to produce dramatic gains in student achievement, but may yield modest gains if safeguards ensure that enough high-quality teachers are available to all students. It is possible that the initiative could cause some of the problems that occurred in California's case, although given the way the amendment is drafted (e.g., with phasing in and potentially full funding), some of the worse consequences could be avoided.

The main issue for Floridians, however, should probably be what alternatives there are to class size reduction. Given the high costs associated with small classes—originally estimated by the Revenue Estimating Conference at \$20.0-27.5 billion cumulatively through 2010—it seems likely that there are more cost-effective strategies of achieving equivalent gains in student achievement (see Levin and McEwan, 2002). The dilemma for policy makers, and the reason class size reduction policies will remain a popular option, is that there are relatively few simple or well-established alternatives that can easily be pointed to. In this context, the idea of smaller classes is likely to be on the table for many years to come.

References

- Bohrnstedt, G. W., & Stecher, B. M. (Eds). (1999). *Class size reduction in California: Early evaluation findings, 1996-1998*. (CSR Research Consortium, Year 1 Evaluation Report). Palo Alto, CA: American Institutes for Research.
- Bohrnstedt, G. W. & Stecher, B. M. (Eds). (2002). *What we have learned about class size reduction in California*. (CSR Research Consortium, Capstone Report). Sacramento, CA: California Department of Education.
- Biddle, B. J., & Berliner, D.C. (2002). *What research says about small classes and their effects*. San Francisco: WestEd. Retrieved November 30, 2004, from http://www.wested.org/online_pubs/small-classes.pdf
- Brewer, D., Krop, C., Gill, B. P., & Reichardt, R. (1999). Estimating the cost of national class size reductions under different policy alternatives. *Educational Evaluation and Policy Analysis, 21*(2), 179-192.
- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest, 2*(1), 1-30.
- Finn, J. D. (2002). Small classes in American schools: Research, practice and politics. *Phi Delta Kappan, 83*(7), 551-560.
- Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*(2), 97-109.
- Florida's amendment to reduce class size*. (2002). Retrieved December 8, 2004, from <http://election.dos.state.fl.us/initiatives/initiativelist.asp>
- Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and student achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2-16.
- Grissmer, D. (1999). Conclusion - Class size effects: Assessing the evidence its policy implications, and future research agenda. *Educational Evaluation and Policy Analysis, 21*(2), 231-248.
- Goldhaber, D., & Brewer, D. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources, 32*(3), 505-523.
- Hanushek, E. A. (1999). 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), 143-163.
- Levin, H. M., & McEwan, P. J. (2002). *Cost Effectiveness and Educational Policy*. Larchmont, NY: Eye on Education.

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE Program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis, 21*(2), 165-177.

Stecher, B. M., & Bohrnstedt, G. W. (Eds.) (2000). *Class size reduction in California: Summary of the 1998-99 evaluation findings*. Sacramento, CA: California Department of Education.

Stecher, B., Bohrnstedt, G., Kirst, M., McRobbie, J., & Williams, T. (2001). Class size reduction in California: A story of hope, promise, and unintended consequences. *Phi Delta Kappan, 82*(9), 670-674.