



DOCUMENT REVIEWED:	“The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida’s Statewide Mandate”
AUTHOR:	Matthew M. Chingos
PUBLISHER/THINK TANK:	Program on Education Policy and Governance, Harvard University
DOCUMENT RELEASE DATE:	May 14, 2010
REVIEW DATE:	July 14, 2010
REVIEWER(S):	Jeremy D. Finn, The University at Buffalo-SUNY
E-MAIL ADDRESS:	finn@buffalo.edu
PHONE NUMBER:	(716) 645-1126
SUGGESTED CITATION:	Finn, J. (2010). <i>Review of “The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida’s Statewide Mandate.”</i> Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved [date] from <a href="http://epicpolicy.org/thinktank/review-class-size-florida">http://epicpolicy.org/thinktank/review-class-size-florida</a>

### Summary

In 2002, voters in Florida approved a constitutional amendment limiting class sizes in public schools to 18 students in the elementary grades, 22 students in middle grades, and 25 in high school grades. Analyzing statewide achievement data for school districts from 2004-2006 and for schools in 2007, this study purports to find that “mandated [class-size reduction] in Florida had little, if any, effect on cognitive and non-cognitive outcomes.” The study has four flaws that, taken together, invalidate it as an evaluation of class-size reduction:

- 1) The data used are drawn from grades 3 and 4 to 8, where the likelihood of finding class size effects is small.
- 2) The differences in class sizes of two comparison groups (treated and untreated) range from about 0.5 to about 3.0 students, all too small to make a difference educationally.
- 3) School and district average class sizes are used in the analysis rather than the actual sizes of classes in which students were enrolled.
- 4) The comparison is between two sets of districts, both with small classes, differing only in whether state funding was used in a focused or general way. This study actually found that administrative discretion in spending state class-size reduction funds did not affect students’ academic performance.

## Review

### I. INTRODUCTION

Despite a strong research base that supports the effect of class-size reduction (CSR) on student achievement,<sup>1</sup> arguments persist about the efficacy of reducing class sizes.

In 2002, voters in Florida approved a constitutional amendment to limit class sizes in public schools to 18 students in the elementary grades, 22 students in the middle grades, and 25 in high school grades. The policy was to be phased in over a period of years but implemented fully by 2010-2011.<sup>2</sup> *The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida's State-wide Mandate*, by Matthew Chingos,<sup>3</sup> purports to evaluate the academic outcomes of the policy as implemented through 2007.

The study consists of two parallel analyses of statewide achievement data, one focusing on school districts' average test scores on the Florida Comprehensive Assessment Test (FCAT) in reading and math for 2004-2006, and another using school-level average test scores for 2007. The district-level analysis was limited to scores in grades 4 through 8, and the school-level analysis to grades 3 through 8. Two groups of students were compared in each analysis: those in districts or schools with average class sizes smaller than the size required by the constitutional amendment (the "untreated" group), and those in districts or schools with average class sizes above the required number (the "treated" group).<sup>4</sup> Districts and schools with average class sizes that were already smaller than the state-mandated class sizes were free to use state class-size reduction funds as they saw fit. The other schools and districts were required to use the funds to reduce class sizes. The per-pupil funds provided to

both sets of districts were the same.

Two sources of information about average class sizes were considered: "official statistics from the Florida Department of Education (FLDOE)...in core classes in grades four to eight (p. 5)" and the study author's calculations from data provided by the State Education Data Warehouse (EDW). By and large, the two data sets yielded the same conclusions.

### II. FINDINGS AND CONCLUSIONS OF THE REPORT

The study addressed three major questions:

1) Was the constitutional amendment followed by a reduction in average class sizes? The Florida Department of Education reports that "average class size in core classes in grades four through eight ... fell from 24.2 in 2003 to 18.6 in 2009" with class sizes falling slightly more in regular than in special education classes (p. 5).

2) How much did class size vary between the treated (those starting with larger than mandated class sizes) and untreated (those starting with smaller than mandated class sizes) groups? In the *district-level analysis*: In Year 1 (2004), the average class size in treated districts dropped by 0.9 students in grades 4 through 8. In untreated districts, the average dropped by 0.1 students. By year 3, the average class sizes had dropped by 3.0 and 1.4, respectively. That is, untreated districts dropped by 1.6 students fewer than did treated districts on average.<sup>5</sup> In the *school-level analysis*: In Year 1 (2007), the average class size in treated schools dropped by 2.0 students. In untreated schools, the average dropped by 0.5 students; 1.5 students fewer

than the decrease in treated schools on average. There was some variation from grade to grade and some variation between official FLDOE reports and the author's calculated averages that had no noteworthy impact on the results.<sup>6</sup>

3) How different was the academic achievement of the treated and untreated groups? In the *district-level analysis*, the report concludes that "mandated CSR did not have a positive effect on student achievement above and beyond the effect of equivalent additional resources" (p. 21). This overall finding of no effect was robust: in other words, it did not vary by grade level, when Stanford Achievement Test scores were used instead of the FCAT scores, or, for the most part, with a variety of approaches to the statistical analysis. In the *school-level analysis*, the report concludes that for both math and reading scores "even small effects can be ruled out" (p. 23). In general, "The results from both the district- and school-level analyses indicate that mandated CSR in Florida had little, in any, effect on cognitive and non-cognitive outcomes" (Abstract).

Comparing the treated and untreated groups, however, was not the same as comparing the achievement of students in small and large classes (see Section VI below).

### III. THE REPORT'S RATIONALE FOR ITS FINDINGS AND CONCLUSIONS

On the surface the approach taken in this study seems logical. All districts in the state, and later all schools in the state, were required to reduce class sizes. Thus no cross-sectional comparison of small classes with large classes was readily available. Academic performance could have been compared before and after class-size reduction took effect, examining achievement differences

in classes, schools, or districts as a function of the magnitude of the reduction, but, as the report acknowledges, this may have been confounded by other changes in school funding, curricula, or new programs over the same time period. Thus, the author took advantage of the fact that some districts had class sizes that met the state requirements prior to the constitutional mandate. Although these districts received the same additional funding as other districts, they were not required to use it for class-size reduction. Funds to these "untreated" districts were in essence "block grants" to the districts to be used at their own discretion. According to the report, "Some surely used it to reduce class size, although the class size numbers suggest this behavior was modest and did not compromise the difference in changes in class sizes between the groups"<sup>7</sup> (p. 9).

Districts with average class sizes above those mandated were required to use the funds for class-size reduction ("treated" districts). Approximately 42% of districts were classified as treated in grades 4 to 8, and approximately 24% of schools. Per-pupil spending was the same in treated and untreated districts. In comparison, per-pupil spending was increased by 11.7% in treated schools and 7.6% in the untreated schools. The report states correctly that "school-level results have a modestly different interpretation than the district-level results" (p. 14), since the groups also differed in the additional resources available.

These two groups of districts (and then schools) were compared in terms of their average class sizes and test scores in reading and mathematics with the intention of estimating the impact of the state CSR initiative. The assumption is made that students in treated districts, and then schools, experienced greater class-size reductions and

thus were “more affected by the policy” (p. 6) than were students in untreated districts and schools.

#### **IV. THE REPORT’S USE OF RESEARCH LITERATURE**

The study is presented as an analysis of the effects of small classes on student achievement,<sup>8</sup> and the literature reviewed pertains to class-size reduction. This background research, however, is not used to inform the study and gives an unbalanced picture of class size research to date.

Since 1978 or earlier, it has been found consistently that benefits of small classes accrue mainly in the early grades (K-3) and primarily are associated with substantial reductions in class sizes.<sup>9</sup> These earlier findings are neither cited nor incorporated into the study. This study began with data from grade 3 (schools) and grade 4 (districts) without attention to the earlier grades. The findings that multiple years in a small class bring greater benefits than do fewer years<sup>10</sup> are not mentioned at all. No analyses of class size research in the upper grades—the grades studied in this evaluation—are cited or described.<sup>11</sup>

The report expresses a bias toward studies that conclude class-size reductions have little or no impact on student achievement. On page 1 the report asserts that “there is little empirical evidence on this question.” Indeed there is as much published and unpublished research on this topic as on most topics in education—much of it showing positive effects. The introduction summarizes Alan Krueger’s secondary analysis of the Project STAR experiment in one sentence, giving no attention to primary analyses or to other secondary analyses of the data. No research is cited regarding other large scale class-size initiatives except for the California pro-

gram,<sup>12</sup> which yielded positive effects but also created a reduction in teachers’ qualifications. The positive effects in California were obtained *despite* the problem of a large number of new and emergency-certified teachers.<sup>13</sup>

Of greater significance, the report fails to consider the strength of conclusions from Project STAR (or other evaluations). STAR was a large within-school randomized experiment—the “gold standard” of empirical research methods. Instead the report compares the results with one flawed analysis of non-experimental data by another scholar that yields negative results.

#### **V. REVIEW OF THE REPORT’S METHODS**

For the analysis of academic achievement, the study used multiple regression analysis with student-level data. Several features were added to the usual multiple regression analysis. In the district-level analysis, districts were weighted by enrollment; dummy codes were included for districts so that student achievement was studied relative to the achievement of other students in the district; standard errors were adjusted for clustering (p. 11), although the type of adjustment is not stated. The school-level analysis is described as “essentially identical to the district-level analysis” (p. 13).

The most important methodological feature is the use of enrollment and achievement trajectories over the years prior to the treatment. In each analysis, the achievement trend for three years prior to the treatment was estimated and the treatment was considered to have an impact if the *change in the trend* was different for the treated and untreated groups, for example, if the slope of the achievement trend became increased for the treated districts (or schools) in compari-

son with the untreated districts.<sup>14</sup> Three years of post-treatment achievement data were available for the district analysis (2004-2006) and one year of data for the school-level analysis (2007).

The negative findings resulted from the conclusion that the post-treatment trends for treated and untreated districts, and for treated and untreated schools, were parallel (i.e., the slopes are the same). This appears to be a correct conclusion based on correct analyses. To the author's credit, the analysis also considered a number of alternative ways of viewing the data and used robust standard errors in tests of significance—all producing consistent results.

## **VI. REVIEW OF THE VALIDITY OF THE FINDINGS AND CONCLUSIONS**

From a statistical standpoint, this study is well done. However, it is not an evaluation of Florida's class-size reduction policy and, even if it were, it was not based on an adequate conceptual foundation. There are four reasons, in order of increasing importance:

1) Prior research has shown that benefits of small classes are greatest in kindergarten or first grade<sup>15</sup> (an argument the report cites on page 2) and decrease in magnitude when students enter small classes in second and third grades, respectively.<sup>16</sup> Consistent with developmental theory, in the STAR experiment students who entered small classes in these years had larger and longer-lasting academic benefits than did students who entered in second or third grade.<sup>17</sup> Effects were strongest for students who entered small classes early and remained in them for several consecutive years.

If effects are smaller for students who begin small classes in second grade than in first, smaller in third grade than in second, and so

on, then there is little likelihood of finding effects in higher grades, especially in a study that averages all class sizes in a school or a district. That is, by studying upper grades, the analyses had limited chances of finding differences in trajectories because the differences may have been small. These certainly would not tell the full story implied by the report's title, "The Impact of ... Florida's Statewide Mandate."

2) The report finds that class size in treated districts dropped by an average of 1.6 students more than in untreated districts. On average class sizes in treated schools dropped by 1.5 more students than in untreated schools. It's difficult to imagine that instruction would change or that the academic or social environment would change with the addition or subtraction of 1, 2, or even 3 students. Moreover, it would be extremely difficult to detect an effect statistically based on such a small difference.<sup>18</sup> The finding of "little, if any, effect" could have been anticipated at the outset from the design of the study.

3) Actual class sizes are hidden when school or district averages are reported. This report provides information about how average class sizes were calculated in this study (in the form of two footnotes on page 5), but none about variability within schools or districts. In most schools, third and fourth (and fifth) grade classes are mostly self-contained, but still there is variability from class to class. In the years covered by this study, individual classes were not constrained to a particular size, and some may have been smaller and some larger. Self-contained (small) special education classes were included in the averages, which suggests that some regular classes may have been quite large. In higher grades, students usually move from room to room for different subjects, and may participate in classes

of very different sizes. If a small class is more efficacious than a large class, the use of averages may well obfuscate this fact.

4) All districts (and then schools) studied in this analysis were required by state mandate to have small average class sizes. Those “treated” received targeted funds to reduce average class sizes down to the mandated sizes. Those “untreated” had small classes prior to the constitutional amendment and received a block grant that was not necessarily used for class-size reduction. Despite some speculation—the results “raise the possibility that comparison districts were able to spend the additional resources more productively than the treated districts” (pp. 21-22)—no information was available about how the funds were spent. After the constitutional amendment was implemented, both groups had small average class sizes.<sup>19</sup>

Instead of a study of the impact of class-size reduction, this is actually a study of the impact of providing resources to districts that are earmarked for class-size reduction versus providing the same amount of resources that districts could spend as they wish (an accurate paraphrasing of the author’s own

statement on page 9). These results confirm other research showing that providing a small amount of administrative budgetary discretion has no impact on student achievement. They do not show whether small classes are beneficial academically because the study was not a comparison of smaller with larger classes.

## **VII. USEFULNESS OF THE REPORT FOR GUIDANCE OF POLICY AND PRACTICE**

Despite its title, this report does not address the issue of class-size reduction. By being presented as an evaluation of Florida’s mandated class size limits, it may lead parents, educators, or policy makers to draw faulty conclusions about the impact of the program.

It is possible that Florida’s class-size reduction, at least in the upper grades, is having little or no impact on student achievement. It is also possible that—consistent with the beliefs of teachers and parents—small classes in the upper grades provide academic benefits. The report does not answer this question one way or the other.

## Notes and References

- <sup>1</sup> Class size initiatives have been widespread across the U. S. Among those have been widely publicized are Tennessee's Project STAR, Wisconsin's Project SAGE, the statewide class-size reduction in California, the county-wide initiative in Burke County, NC. These are described in multiple publications including Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2004). The "why's" of class size: Student behavior in small classes. *Review of Educational Research*, 73, 321-368. Numerous other local, district, and statewide evaluations have also been conducted.
- <sup>2</sup> The maximum class sizes needed to be attained by 2004 at the district level and 2007 at the school level.
- <sup>3</sup> Chingos, M. (2010). *The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida's Statewide Mandate*, Working Paper PEPG 10-03. Cambridge, MA: Harvard University, Kennedy School, Program on Education Policy and Governance.
- <sup>4</sup> This classification made an evaluation of class-size reduction in grades P to 3 difficult or impossible because most districts would have been in the treated group.
- <sup>5</sup> The regression analyses produced slightly different estimates of the difference in class-size reductions, which also varied by grade from little or no difference in grades 4 and 5 to a 3.0-student difference in grade 7. Based on the regression results, the report notes "Class sizes in grades four and five were reduced by similar amounts in the treated and comparison districts" (p. 17).
- <sup>6</sup> The author reports that "when instead I use the official FLDOE class size averages, I obtain modestly different results." (p. 9, footnote 16)
- <sup>7</sup> Actual information on how these funds were used by untreated districts was not available.
- <sup>8</sup> This is a misleading portrayal of the study; see "The Validity of the Conclusions" below.
- <sup>9</sup> Glass and Smith's (1978) meta-analysis yielding these results is cited in the report but without reference to its findings. See also
- Robinson, G., & Wittebols, J. (1986). *Class size research: A related cluster analysis for decision making*. Arlington, VA: Educational Research Service.
- Slavin, R. E. (1989). Achievement effects of substantial reductions in class size. In R. E. Slavin (Ed.), *School and classroom organization* (pp. 247-257). Hillsdale, NJ: Erlbaum.
- <sup>10</sup> See Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, 103, 145-183.
- Krueger, A. B. & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school tests results: Evidence from Project STAR. *The Economic Journal*, 11, 1-28.
- <sup>11</sup> For example, Boozer, M. & Rouse, C. (1995, June). *Intraschool variation in class size: Patterns and implication* (Working paper no. 344). Washington, DC: National Bureau of Economic Research, Industrial Relations Section.
- Deutsch, F. M. (2003, June), How small classes benefit high school students. *National Association of Secondary School Principals Bulletin*, 87, 35-44.
- <sup>12</sup> It is not clear why the report cites the Jaspens-Rivkin evaluation rather than the original evaluation conducted by the California consortium of policy makers and scholars: Bohrnstedt, G. W. & Stecher, B. M. (Eds.). (2002). *Capstone report: What we have learned about class size reduction in California*. Palo Alto, CA: CSR Research Consortium. American Institutes for Research
- <sup>13</sup> This problem may have self-corrected in the ensuing years.
- <sup>14</sup> In the district analysis, the three years were 2001-2003, which immediately preceded the treatment. In the school-level analysis, it appears that the three years were also 2001-2003 although the post-treatment data were collected in 2007. This information is given only as table footnotes, which include the phrases "[scores] were standardized...based on the distribution of scores in 2001 to 2003" and "Data cover period from 2001

to 2006.” The only post-treatment year in the school analysis was 2007, the first year in which class-size reduction was mandated for individual schools..

<sup>15</sup>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, 143-163; Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, 111, 1-28.

<sup>16</sup> Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, 103, 145-183.

<sup>17</sup> The long-lasting effects included academic achievement in the upper grades, taking college entrance examinations and graduating from high school:

Krueger, A. B. & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school tests results: Evidence from Project STAR. *The Economic Journal*, 11, 1-28.

Nye, B., Hedges, L., & Konstantopoulos, S. (2000). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 21, 127-142.

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology*, 97, 214-223.

<sup>18</sup> Recall that these are also upper grades where the effects may be small.

<sup>19</sup> Indeed, Figure 2 shows that from 2004 to 2006 (after the constitutional mandate was implemented), eighth-grade math and reading scores increased monotonically from year to year for both groups!

---

The Think Tank Review Project is made possible by funding from the Great Lakes Center for Education Research and Practice.