

# Evaluation Brief

---

Public Schools of North Carolina, State Board of Education, Phillip J. Kirk Jr., Chairman  
Department of Public Instruction, Michael E. Ward, State Superintendent

Volume 2, Number 6  
August 1999

## Update on Class Size Research Since 1998

---

Class size reduction (CSR) is seen as a key reform strategy at both state and national levels. Federal funds for CSR are now available to states with a focus on grades K-3, based on research to date that shows positive outcomes at the early grades. California initiated a statewide CSR initiative in 1996-1997 to reduce class size to 20 in kindergarten through third grade. Other states have initiated class size reduction focusing on lower-income schools (e.g., Indiana's *Prime Time* and Wisconsin's *Student Achievement Guarantee in Education [SAGE]*). Perhaps the best-known and most well-designed large-scale CSR study was the Tennessee *Student-Teacher Achievement Ratio (STAR) Study*, running from 1985-89 for 79 schools in grades K-3.

Because a number of reviews of class size research already exist, this Brief is designed to update the most recently published information. Finn (April 1998) and Pritchard (March 1999) have provided two useful and concise reviews for the US Department of Education. The Finn review has been posted on the Evaluation Section's WebPage. New information is provided by the 1998-99 report of the Wisconsin SAGE evaluation data, the 1998-99 report of California's class size reduction program, and a paper session at the 2000 Annual Meeting of the American Educational Research Association that addressed the most recent follow-up data for the STAR students in Tennessee.

Class size versus student-teacher ratio. A key issue in class size research is the confounding of class size (i.e., the actual number of students in a class) with student-teacher ratio (the number of students in a school divided by total number of teachers). It is easier to get a lower number with the latter, as special teachers, counselors, teacher assistants, and other staff may be included in the denominator. Many earlier studies used student-teacher ratio, and some reviews include studies with both types of indicators. However, the more recent studies have tried to clarify the issue by focusing more clearly on class size per se.

## Summary of Recent Research <sup>1</sup>

### Tennessee STAR Study.

Because of the significance of the STAR study and the fact that students have continued to be followed through middle and high school, a brief description is provided here. The study included 7,500 students in grades K-3 in over 300 classrooms and 79 schools during the 4-year longitudinal study. In each school, K-3 students were randomly assigned to one of three class configurations: small class (S) with enrollments of 13-17 students, regular class (R) with an enrollment of 22-27 students, and regular class with a full-time teacher aide (RA) with 22-26 students. Teachers were also randomly assigned to classes each year. Students remained in the same type of class until they were in grade 3. Students returned to regular class sizes in grade 4. Standardized achievement tests (Stanford Achievement Tests - SAT) and curriculum-based tests (Tennessee's Basic Skill First - BSF) were administered each year, along with a measure of motivation and self-concept.

Initial Analyses. The following results were found across all four years of the initial study (Finn, 1998; see also Finn & Achilles, 1999):

- Differences among the three types of classes were highly statistically significant, with small classes consistently outperforming either regular classes or regular classes with aides. In fact, there was no difference between the (R) and (RA) classrooms.
- These results held for all students - white, minority, inner city, rural, or urban. However, there was an even greater benefit for minority students and those attending inner-city schools. The effect size for minorities typically was about twice as large as for white students, resulting in reduction of the achievement gap.
- There were no differences in motivational/self-concept scales among the class types.

Lasting Benefits Study. Follow-up studies of the students in the STAR study were initiated in 1989. Students were returned to regular-sized classes in grade 4, and standardized tests, criterion-referenced tests, and a new measure of student engagement were administered. At the end of grade 4, results indicated that:

- Students who had been in (S) during K-3 grades had higher achievement in all academic areas compared students who had been in larger (R) and (RA) classes.
- Students who had been in (S) were better behaved than students from larger classes.
- Students from (S) were rated as expending more effort in the classroom and taking more initiative with regard to learning activities.

At least through grade 8, significantly higher academic achievement for the students from the smaller classes persisted, although the difference had become smaller.

Recent Star Follow-up Research. At the 2000 Annual Meeting of the American Educational Research Association, Finn, et al. (2000) presented additional analyses of the Tennessee STAR data that included (1) reanalysis of previous data using alternative analysis techniques (i.e., Hierarchical Linear Modeling [HLM]), (2) a look at the effect of the number of

years in a small class on achievement, (3) examination of carry-over effects to higher grades, and (4) estimated strength of effects in terms of number of months of schooling.

1. HLM Analyses. New analyses confirmed statistically significant differences in favor of small classes on every test in every grade. There were still no significant differences between (R) and (RA). There were fewer significant interactions with specific groups, but inner-city students still realized greater benefits than did suburban/rural students.
2. Number of Years in Small Classes. Students who attended (S) in K-3 performed better than students in (R) or (RA) on all subjects in Grades 4, 6, and 8. However, number of years in small classes was important. Thus, for third-grade students, one year in at grade 3 was not significant, two years in grades 2 and 3 were marginally significant, and three and four years were significant - with four-year effects greatest.
3. Carryover Effects to Grades 4, 6, and 8. One year in (S) during grades K or 1 was not sufficient to produce *long-term* effects, even in Grade 4. Two years was somewhat better, but there were no noteworthy carryover effects. Carryover effects were consistently significant for students with three and four years in (S). All test scores in grades 4, 6, and 8 (with the exception of Grade 6 science) were significantly higher for students who were in (S) for four years. [The authors note that attending (S) for four years is confounded with kindergarten entry, since that is the only way that students could have participated four years.]
4. Impact in Terms of Months of Schooling.
  - During the K-3 years, the effects (or impact) typically increased from grade to grade (e.g., Total Reading increased from a 0.5-1.3 to 3.9-4.5 month advantage from kindergarten to Grade 3). Additional years in (S) generally increased the effects (e.g., For Grade three, Total Mathematics increased from a 1.3 to 3.7 month advantage with one versus four years in [S]).
  - The impact of (S) for students in subsequent grades increased as well with each additional year students spent in small classes. The advantage of three years in (S) was approximately 4-5 months in reading and mathematics for Grades 4, 6, and 8. Four years in (S) yielded an effect of 5-6 months (half a school year) in Grades 4 and 6, and even higher in Grade 8.

*This research demonstrates that spending three or four years in small classes in grades K-3 has an enduring effect for at least five years after the intervention.*

Boyd-Zaharias and Pate-Bain (2000) are conducting an even longer-term follow-up study of Project STAR students. High school data have been collected on more than 5,000 STAR students, and students have also been matched with Tennessee college enrollment files. Since many students may attend college or live out of the state, the latter numbers are considered an underestimate. Although data are preliminary, initial analyses show that STAR students who were in small classes were significantly more likely than the other students to graduate from high school on schedule and to receive an honors diploma. They were less likely to dropout of high school. *Very preliminary* analyses suggest that students in (S) have a somewhat higher college attendance rate.

## SAGE Second-Year Evaluation <sup>2</sup>

Wisconsin initiated a class size reduction initiative in grades K-3 in 1996-97 called *Student Achievement Guarantee in Education (SAGE) Program*. Districts with high numbers of students from low-income families were eligible for this program. Target class sizes of 15 were implemented in grades K and 1 in 1996-97, grade 2 in 1997-98, and grade 3 in 1998-99. Some classes had 30 students with two certified teachers, while most classes had 15 students and one teacher. SAGE schools were matched to comparison schools for evaluation purposes. First-year evaluation results are reported in Molnar, et al. (1999). Second year evaluation results have since been issued (Molnar, Smith & Zahorik, December, 1999), and selected results are summarized here.

- As in the first year results, SAGE schools show statistically higher performance on the Comprehensive Tests of Basic Skills (CTBS) than comparison schools in language arts, math, and total scores in 1998-99.
- At the classroom level analysis, small classrooms tended to score significantly higher in language arts, mathematics, reading, and total scores after adjusting for individual pre-test results, socio-economic status and attendance.
- In both years, Black students scored lower on CTBS pre-test scores than Black students in comparison schools, but made significantly higher gains, outperforming comparison school students on post-tests.
- In both years, Black SAGE students made greater gains on the total scale score than White SAGE students, closing the achievement gap. In comparison schools, Black students achieved smaller gains than White students, widening the achievement gap.
- Results suggest that significant effects of SAGE at the end of the first grade were maintained in second and third grades.
- Classrooms with 30:2 student-teacher ratios demonstrated achievement equal to classrooms with 15:1 ratios, with the exception of language arts and mathematics sub-tests in second grade where 15:1 classrooms were significantly higher.
- It should be noted that generalizations are based on aggregate results, where achievement scores are pooled for small and large classrooms respectively. Using these aggregate results, a small class is expected to gain two or three months more than a large class in early grades. Although not every small class made such gains, they did form the majority of top-performing classrooms over a two-year period.

## California's Statewide Class Size Reduction <sup>3</sup>

In 1996-97 California began a *Class Size Reduction Program* to reduce the student-teacher ratio throughout state for Kindergarten through third grade to 20 to 1. In the first-year report for 1997-98 (Bohrnstedt & Stecher, 1999) results indicated that: (1) implementation of CSR was proceeding rapidly, (2) the students most in need (e.g., low-income, minority) were the least likely to have benefited from it, (3) 12 percent of K-3 teachers in 1997 were not fully credentialed compared to one percent prior to CSR implementation, and (4) third-grade students in smaller classes had a small positive, but not significant, gain in achievement compared to students in larger classes.

1998-99 Findings. In the second evaluation report for the 1998-99 school year, the CSR Program was almost fully implemented, with 92 percent of California's K-3 students in classes of 20 or fewer. Those classes not yet reduced in size were concentrated in districts serving high percentages of minority, low-income or English Language learners (ELL). Other findings are summarized below.

- The average qualifications of California teachers (e.g., full credentials) declined during the past three years for all grade levels, but declines were worse in elementary schools.
- Qualifications for K-3 teachers continued to decline, albeit at a slower rate than the previous year. Schools serving low-income, minority, or ELL students had fewer well-qualified teachers than other schools.
- There was a small positive gain in student achievement for reduced class sizes in grade 3, and this gain was realized equally by all students regardless of background characteristics. For students who had moved to Grade 4 in 1998-99, a small positive effect persisted even after they returned to larger classes.

The authors note that these results are smaller than those found in the STAR study, and this result may be partially due to the fact that no student in California has yet had four years in small classes. They also note that no state has initiated such a wide-scale CSR reform before. Additional time and experience are needed to judge the success of the program. It should also be noted that the target class size in California is 20, whereas the class sizes in the STAR study ranged from 13-17. It is unknown whether this size difference might also effect the difference in results.

## **Instructional Practices and Other Outcomes in Smaller Classes**

Understanding why smaller classes make a difference is essential to maximize results from this strategy. To what extent does reducing class size result in changes in teacher behaviors and to what extent are those changed behaviors responsible for any changes in student performance? Studies of instructional processes in reduced and non-reduced size classes are preliminary at present but do offer guidance to schools and districts. The STAR and SAGE studies, as well as the California evaluation, have looked at instructional strategies and/or behavior in the classroom. Clearly, reducing class size impacts numerous variables in the classroom. Mitchell, et al. (1989) suggest that CSR changes the access to available educational resources, including reducing the numbers of possible student-to-student interactions and increasing the available instructional time between student and teacher. In a study in Burke County, NC, teachers found they had more physical space to work with once they had fewer students (Egelson et al., 1996).

Finn (1998) explores "student engagement" as an overarching explanatory concept. Studies of student engagement show that it is correlated with both academic and behavioral outcomes and can explain why some students perform well in school in spite of disadvantages faced as members of high-risk populations. It has also been shown that minority and low-income students are not as engaged, either academically or emotionally, with the schooling process. If less time is spent on classroom management and more on interaction between students and

teachers, academic engagement should be enhanced. Thus, if small class size positively impacts student engagement, it should logically have the biggest payoff for minority and at-risk students.

In a study of math classes drawn from a nationally representative sample of middle and high schools (the Longitudinal Study of American Youth), Betts and Shkolnik (1999) found that class size variations resulted in little change in how teachers allocated time. There were small increases in the amount of time teachers spent on review of materials when size was reduced, but not on time spent on new material - especially for classes described as "below average." Middle school teachers were somewhat more likely to change time allocations than high school teachers. The authors note that the positive impact of CSR might be stronger if teachers adjusted instructional practices to a greater extent to take advantage of smaller classes. However, these results are based on grades 7-12, and results might vary considerably for elementary school teachers.

Most of the major CSR studies at the elementary grades do show improvements in student behavior, less disruption, and/or more time-on-task. Students in one STAR study of grade 2 classes (Evertson & Folger, 1989) found that in small classes:

- higher levels of performance were attained with less expenditure of teacher time,
- students initiated more contacts with the teacher for clarification and seeking help in mathematics,
- students were more task-oriented and spent less time waiting for the next assignment in reading, and
- teachers were rated as better monitors of students' understanding of class material.

In the Burke County study, trained observers found a greater percentage of on-task events and fewer institutional (i.e., discipline) events in small classes compared to regular classes. On-task behaviors increased as a percentage of all behaviors over time and decreased over the same time span in larger classes (Achilles, et al., cited in Finn 1998).

The SAGE evaluation has focused on classroom processes through use of classroom observations, teacher questionnaires, teacher and principal interviews, and teacher logs. They found the predominant characteristic of teaching in smaller classes is *individualization* by one-on-one tutoring, small group instruction, and whole-class instruction where each child received attention. Increased individual attention is reportedly gained by less time spent on discipline and more on instruction, greater knowledge of students, and greater enthusiasm for teaching.

## **Cost-Benefit of Class Size Reduction**

While research shows smaller classes in the primary grades leads to positive achievement effects for those immediate grades as well as future grades, reducing class size to an average of 20 or less has considerable cost implications. Brewer, et. al. (1999) estimated operational costs (i.e., non-facility costs) of hypothetical national CSR programs to reduce class size to 20, 18 and 15 respectively for grades 1-3 under various policy parameters. They point out that the many variables that go into estimating costs for any CSR strategy include size of class; grade levels affected; whether the size must be met at a school, district, or state level; how flexible the policy is; and whether it includes all students or targeted students (e.g., schools with high percentage of

low-income students) to name a few. In addition, their policy analysis, as well as California's early experience with CSR, shows that a dramatic increase in number of classes - and therefore in the demand for new teachers - may well raise teacher salaries and/or reduce teacher quality. In sum, they provide a base estimated annual cost for a national class size reduction to 18 students in grades 1-3 of approximately \$5 billion. Additional factors could add another \$600 million annually. To reduce class size to 15, the average size in the STAR Study, would cost more than twice as much (\$11 billion). And these figures do not include kindergarten.

The effectiveness or benefits side of the equation is difficult to assess, including what other strategies we forego (opportunity costs) when we invest only in CSR. Is CSR more effective than other interventions? For example, Brewer, et al. point out that \$5-6 billion per year is enough to raise salaries of every teacher in grades 1-3 nationally by \$10,000 per year. Assuming a link between teacher quality and salary and that only one choice were possible, is it more effective to have smaller classes and less qualified teachers or larger classes and more qualified teachers? In addition, how much is enough, or "What benefits are associated with what levels of investment (Finn, 1998)?" If STAR Study results continue to be confirmed, levels of achievement and other positive long-term academic outcomes may well be worth the investment. Yet comparisons to other interventions, which may not be implemented at a similar grade level, intensity, or duration, are difficult. And none of these issues addresses the need for more classroom space, facility costs, or ability to find new teachers.

A lower cost strategy would be to target CSR to schools and students most in need (e.g., high minority and high poverty schools). Odden (1990) concluded that targeted CSR for low-achieving students was worthwhile, especially if coupled with other strategies. The promise of CSR - especially in the early grades - must be considered along with financial implications, opportunity costs, and implementation feasibility. Nevertheless, the promise of CSR in the early grades and the potential longer-term academic outcomes are compelling.

## **Conclusions About Class Size Research to Date**

- Research at the early grades supports the positive impact of smaller class sizes on student behavior and achievement. Reductions in class size to 20 and below appear to be a minimal requirement, with even greater reductions likely to produce greater benefits. Smaller class sizes (average 15) in the STAR study suggest both better immediate and long-term gains, but research does not provide clear answers on specific benefits of classes of various sizes. Fewer definitive claims can be made at present regarding the impact of CSR at higher grade levels. There is ample logic as well as developmental theory that support targeting any CSR strategy at grades K-3. If these early years form the foundation for learning, basic skills, and attitudes toward school and if they represent an important developmental period (as do the preschool years), then targeted CSR in these grade levels *may* ultimately be most cost-efficient and have the most payoff. However, more research is needed in these areas.
- Follow-up to the original STAR experiment suggests that long-term benefits accrue the more years a student spends in a small class in grades K-3. Students need to be exposed to CSR for three - and especially four - years to have maximum long-term benefit.

- The positive results of CSR appear to be particularly important for minority and low-income students. If CSR needs to be phased in, starting with schools serving students most at-risk would seem to be a desirable strategy with the greatest pay-off.
- Reducing class size does not automatically change teaching behaviors, although in the elementary grades more individualization and more time-on-task are evident. However, "...the very feature of 'smallness' sets the stage for greater student engagement in learning. Teachers must attend to all students regardless of any inclination to the contrary (Finn, et al., 2000)." Initiatives to reduce class size may be most effective if explicit strategies to help teachers better utilize the available resources of time and fewer students are concurrently implemented.
- Rapid reduction of class size has obvious implications for finding an adequate number of qualified teachers, as seen in California's initiative. Other considerations include adequate classroom space, increased costs of facilities, and lost opportunities for other interventions.

---

<sup>1</sup> A 1999 special issue of *Educational Evaluation and Policy Analysis* was devoted to reviewing the major class size research as of 1998. That issue is Volume 21, Number 2, Summer 1999.

<sup>2</sup> The 1998-99 SAGE evaluation report is available on the University of Wisconsin-Milwaukee WebPage at <http://www.uwm.edu/SOE/centers&projects/sage/>.

<sup>3</sup> This summary is based on the Executive Summary for the 1998-99 California evaluation results. The full report was not available at the time this Brief was written. The Summary and, ultimately, the full report will be available on the WebSite for the CSR Research Consortium at <http://www.classsize.org>.

## References

Achilles, C. M., Kiser-Ling, K., Owen, J., & Aust, A. (1994). *Success starts small: Life in a small class. Final Report*. Greensboro, NC: University of North Carolina.

Betts, J. and Sckolnik, J. (1999). The behavioral effects of variations in class size: The case of math teachers. *Educational Evaluation and Policy Analysis*, Vol. 21 (2), 193-213.

Bohrnstedt, G. & Stecher, B. (P.I.s) (1999, June). *Class size reduction in California 1996-98: Early findings signal promise and concerns*. Palo Alto, CA: CSR Research Consortium.

[Available on the CSR Research Consortium Website at: <http://www.classsize.org>]

Bohrnstedt, G. & Stecher, B. (P.I.s) (2000, June). *Class size reduction in California: The 1998-99 evaluation findings*. Palo Alto, CA: CSR Research Consortium.

[Available on the CSR Research Consortium Website at: <http://www.classsize.org>]

Boyd-Zaharias, J. & Page-Bain, H. (2000, April). *The continuing impact of elementary small classes*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

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.

Egelson, P., Harman, P., & Achilles, C. M. (1996). *Does class size make a difference? Recent findings from state and district initiatives*. Greensboro, NC: Southeastern Regional Vision for Education (SERVE).

Evertson, C. M. & Folger, J. K. (1982, March). *Small class, large class: What do teachers do differently?* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Finn, J. D., (1998, April). *Class size and students at risk: What is known? What is next? A Commissioned Paper*. Washington, DC: National Institute on the Education of At-Risk Students, U.S. Department of Education.

[Available on the Website of <http://www.ed.gov/pubs/ClassSize/title.html>]

Finn, J. D. & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*. 21 (2). 97-109.

Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2000, April). *The enduring effects of small classes*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Mitchell, D., Carson, C., & Badarak, G. (1989). *How changing class size affects classrooms and students*. Riverside, CA: California Educational Research Cooperative, University of California.

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.

[The full report is available on the SAGE website at:  
[http://www.uwm.edu/soe/centers&projects/sage/.](http://www.uwm.edu/soe/centers&projects/sage/)]

Molnar, A., Smith, P., & Zahorik, J. (1999). *1998-99 evaluation results of the Student Achievement Guarantee in Education (SAGE) Program*. Milwaukee, WI: University of Wisconsin-Milwaukee.

[Available at <http://www.uwm.edu/soe/centers&projects/sage/>]

Odden, Al (1990). Class size and student achievement: Research-based policy alternatives. *Educational Evaluation and Policy Analysis*, 12, 213-227.

Prichard, I. (1999, March). *Reducing Class Size: What do we know?* Washington, DC: National Institute on Student Achievement, Curriculum and Assessment, U. S. Department of Education.

[Available on the web at <http://www.ed.gov/pubs/ReducingClass/title.html> ]