AMERICAN Educator
SUMMER 1999
AMERICAN FEDERATION OF TEACHERS

PROJECT STAR

The Story of the Tennessee Class-size Study

BY JAYNE BOYD-ZAHARIAS

Project STAR is "one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools."[1]

IN 1945 when Helen Pate-Bain, the driving force behind Project STAR, graduated from George Peabody College for Teachers, in Nashville, Tenn., and started her teaching career, the average class size in Tennessee schools was thirty. For the next twenty years, she taught English and speech classes in grades 7 to 11. English classes always had thirty to thirty-five students, but speech classes, because they were an elective, were considerably smaller—about fifteen to twenty students per class.

Pate-Bain noticed that she accomplished much more in her smaller classes, and she began thinking about the impact of class size on student learning. Smaller classes, she decided, would make a big difference for all her students, but they would probably be even more important for children in elementary school.There were always several students in her seventh-grade classes who were not prepared to do the work. Pate-Bain believed that this resulted from the children's not getting an adequate educational foundation. Reducing class size in the early grades so teachers could give youngsters the individual help they need in learning the basic skills of reading, writing,and beginning mathematics could pay dividends throughout their school years.

When Pate-Bain began her crusade for smaller class size in the early grades, she was always met with the same response:"You can't *prove* that class size makes a difference in learning."This answer, though legitimate, was also very convenient for local school boards,and state and national legislators because lowering class size would have a big impact on school budgets.

In time,there were a number of small studies on the effect of reducing class size. Pate-Bain read Lynne M. Johnson's pilot study that was carried out in South Carolina[2], and she visited Indiana's Prime Time demonstration project in 1982.[3] She found Gene Glass's synthesis of the many small studies of reduced class size[4] particularly useful. Basing her view on his analysis, she concluded that one teacher with fifteen students in the classroom would produce much better learning *and* be financially feasible.

In 1984, when she was an associate professor at Tennessee State University, Pate-Bain received a small grant to

Jayne Boyd-Zaharias was research director of Project STAR. She is now the director of HEROS (Health and Education Research Operative Services), which has been funded by the Tennessee legislature to continue Project STAR follow-up studies (heros@telalink.net; 615/449-7904).

study reduced class size (one teacher per fifteen students) in grades 1 through 3 at a Nashville elementary school. The results were positive, but student mobility reduced the sample size, and by the end of third grade, it was so small that Pate-Bain knew researchers,administrators, and politicians would not place much credence in her results.

How could she finance a study with a large sample? Pate-Bain went to the main source of money for public education in Tennessee:the legislature.She spent a year making personal visits to Tennessee legislators,presenting her findings, and trying to sell them on the need for a large-scale study of class size. Legislators had already been discussing the class-size issue in the context of Gov. Lamar Alexander's Better Schools Program. Money was the crucial consideration for the legislature, and it was money that tipped the balance in favor of STAR. Legislators were convinced that a large-scale study of class size would at least let them know if smaller classes actually would benefit students,and would help them decide where to spend future education dollars. So they agreed to fund Project STAR (Student/Teacher Achievement Ratio), a study that would consider the effects of class size on students in grades K-3.The legislature provided $12 million,the majority of which—more than $9 million—was to go for what Pate-Bain calls "the most important piece of equipment"in classrooms: teachers. In this era immediately following *A Nation at Risk,* many states started to make changes in their education systems, but none made a substantial contribution to education research like the one that Pate-Bain led in Tennessee.

The Study Design

The STAR experiment was designed by a group of researchers including Pate-Bain and other academics and members of the Tennessee Department of Education. Some of its key features are:

1. *All Tennessee schools with K-3 classes were invited to participate. Giving every school a chance to join the study helped to ensure a diverse sample.* It also ruled out the possibility that critics could attribute class-size effects to STAR researchers' having "chosen" certain schools.

2. *Each school included in the study had to have a large enough student body to form at least one of each of the three class types—small (thirteen to seventeen students), regular (twenty-two to twenty-six students), and regular with a full-time teacher aide (twenty-two to twenty-six students)—in order to accommodate the within-school design.* The within-school design provided a built-in control for differences among schools in areas like resources,leadership,and facilities.As a result, class-size effects could not be attributed to these differences.

3. *Seventy-nine schools in forty-two systems met the within-school design requirement, and the original STAR sample comprised more than 6,000 students per grade level.* The large sample lent credibility to the results. It also allowed for the inevitable reduction in the size of the sample because of student mobility.

4. *Schools from inner-city, rural, urban, and suburban locations were included in the experiment.* This feature of the study, which the legislature mandated, guaranteed that the sample would include children from various ethnic backgrounds and income levels.

5. *Students and teachers were randomly assigned to their class type.* The random assignment made certain that differences in the students' test scores could be confidently attributed to class size.It would not be possible to assert that the researchers had placed all the smart children within a particular class type, or that the best teachers were given a particular class size.The random assignment of classes to one of the three categories was one of the strongest features of the STAR study.

6. *Investigators followed the standard procedures for confidentiality and human subjects' research.* Only principal investigators and their staff had access to individual student information. Results were always reported at an aggregate level so that no individual child's demographic or test-score data could be discerned.

7. *No children were to receive fewer services than normal because of the experiment.* This, too, was required by the legislature, but it was an easy condition to fulfill: Without STAR, all of these children would have been in class sizes ranging from twenty-two to twenty-six (or larger).Therefore,the study did not "harm" any children.

8. *Student achievement was to be tracked by standardized tests, which were carefully monitored.* During testing,monitors ensured that test instructions were followed and that teachers did not coach or help students taking the tests.

9. *An outside consultant was contracted to perform all primary statistical analyses.* Jeremy Finn, a professor of education at the State University of New York, Buffalo, served as the primary statistician. An expert in the field, he had not been involved with the study or the principal investigators before the Tennessee Department of Education contacted him.This additional safeguard guaranteed impartial results.

On paper, these nine key factors may appear fairly obvious.In fact,having them all in place was (and still is) a rarity in education research. Experiments that do so are expensive and take a long time to bring to fruition.And as medical researchers also know, an experiment involving human subjects raises special problems. In our case, we also had to get the schools involved in the study to agree to the terms of the research,and that was not always easy. For example,some schools that planned to participate had to drop out because their principals would not agree to the random assignment of students and teachers.These administrators had a system for assigning children and teachers, and they wanted to stick to it. Other schools did not want to administer the achievement tests, so, again, they could not participate in the study. The strong controls maintained by the consortium meant that all statistical variables, except for class size, could be factored out of the analyses.Therefore, any achievement effects could be attributed to class size.

The consortium decided to measure student achievement using the Stanford Achievement Test (SAT),a nationally normed standardized achievement test, and the Basic Skills First (BSF) test, a criterion-referenced test designed to measure areas that matched the Tennessee state cur-

riculum. Both of these measures were already being used by some school systems across Tennessee. The primary STAR analyses used the total math and reading scores from the two tests.

Carrying out the study involved its own set of complications. For example, at the beginning of the study, laws requiring school children to have Social Security numbers were not in place.If we were to follow students from year to year, we had to have an identification number, so the consortium decided to request that students put their birth certificate numbers in the spaces designated on the test booklets for the ID number. However, there was a problem.Birth certificate numbers were eleven digits,and test spaces for ID numbers only allowed for nine digits. We tried to head off potential confusion by sending instructions with the tests, telling teachers to cut off the first two digits of their birth certificate numbers, but the instructions were not always followed. This created a nightmare for the data processing staff. Student data had to be checked and matched not only by ID, but also by name, birth date, gender, and ethnicity.

Then, at the end of the first grade, the state decided to switch to Social Security numbers for the primary student ID. This amplified our problems. Now, the data processing team was faced with tracking two sets of IDs and matching up the modified nine-digit birth certificate number with the Social Security number. Tracking the two ID sets was a difficult process, but it ultimately resulted in a better method for following the STAR students: STAR researchers would now be able to track students through high school and beyond.

Early Results

In the fall of 1985, with the STAR design in place, 6,328 kindergarten children and 329 kindergarten teachers were randomly assigned to one of the three class sizes. Children were to remain with their initial class assignment through the end of the third grade, the 1988-89 school year.[5] Except for the fact that some of these children were in classes of fifteen,all followed the normal routines established by their specific schools and school systems.

There was one aspect of the study that researchers could not control for—students coming into or leaving a class. Because kindergarten was not mandatory in Tennessee during the 1985-86 school year, the first year of the study, the enrollment of first-graders during year two of STAR,brought an influx of new students to the study. Similarly, when students moved out of STAR schools, they were no longer part of the longitudinal sample. However, STAR researchers had foreseen these problems. When students moved out of STAR schools, they were no longer part of the study. The new students entering STAR at the beginning of first grade were included in the study, and the research proceeded as planned. The influx of new students throughout STAR increased the total number of STAR participants to more than 11,000.

During spring of each year, the STAR students took the SAT and BSF tests. Kindergarten results showed that the small-class students outscored their peers from the larger classes, and the differences in scores were statistically significant. During the following years of the experiment, in grades 1 through 3,test results continued to show statistically significant differences between small and regular-size classes.The outcomes on both the SAT and BSF always favored the small classes. (See Figures 1 and 2.)

Smaller classes made the biggest difference for inner-city, low-income minority children. However, *all* students benefited from the experience, regardless of their ethnicity, gender, socioeconomic status, or the location of their school.

The state of Tennessee took seriously the findings about the benefit of small classes for low-income and minority children. In 1989, it established Project Challenge, which provided funds to the sixteen poorest counties (based on per-capita income) for reduced class sizes in kindergarten through third grade. This project was not an experiment like STAR;it was a policy application of the STAR findings, and it got excellent results. Charles Achilles, a member of the consortium that created the original STAR design, followed student achievement in these counties (1997)[6] and found that "on average, the Challenge systems that started the 1:15 treatment in 1989 ranked well below the state average. By 1995 they ranked near or above the state average."

Figure 1
STAR Reading Percentile Ranks
Kindergarten–Grade 3, 1985-89

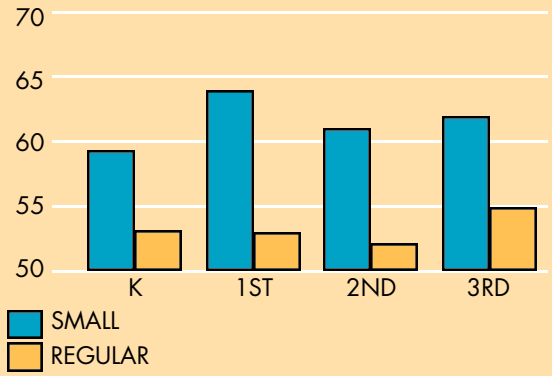
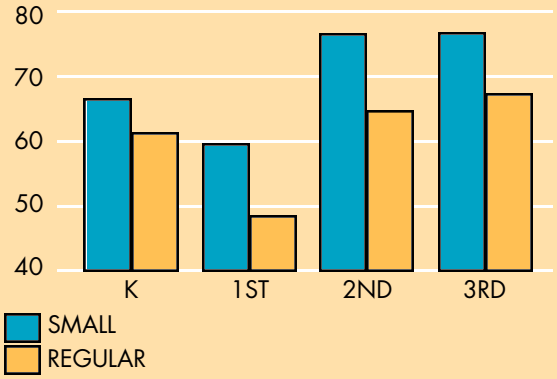


Figure 2
STAR Math Percentile Ranks,
Kindergarten–Grade 3 1985-89



There were never strong research findings favoring the regular-size classes with full-time teacher aides.This component had been added at the request of legislators, who reasoned that it costs less to employ an aide than a teacher. However, Tennessee at that time, had few standards or requirements and little training for aides, and Pate-Bain believes that this is probably why the STAR research did not discover significant differences favoring the classes with aides.

At the end of each year of the study, in order to keep the Tennessee legislature informed, Pate-Bain visited members individually to give them an update on our findings. She liked presenting the research in language that could be easily understood by parents, teachers, legislators, and school board members, and she would ask her staff to convert scores into grade equivalents. However, until this year, STAR findings were never published in this format.

We now have results from recent re-analyses of the effects of small classes in grades K-3, presented in terms of months of schooling. Students in small classes exceeded their counterparts in regular classes in every grade and were about a half year (from 2.8 to 4.7 months) ahead in their schoolwork by the end of grade 3.[7] (See Table 1.)

Table 1
Advantages of Attending a Small Class in K-3
Reported in Months of Schooling

	Reading	Math	Word Study Skills
Kindergarten	.5 months	1.6 months	.5 months
Grade 1	1.2 months	2.8 months	.8 months
Grade 2	3.9 months	3.3 months	5.7 months
Grade 3	4.6 months	2.8 months	4.7 months

Pate-Bain and other STAR researchers devised procedures to continue monitoring the progress of STAR students when the children went on to fourth grade and into regular-size classes.At the same time, Jeremy Finn,the program's independent statistician, began a related study in which he investigated the effects of smaller classes on student participation. He found that fourth-grade students from the smaller classes were more engaged in class and school activities than the students who had attended the larger-size classes in kindergarten through third grade.The small-class students showed greater initiative, participated more willingly, and put forth greater effort in their fourth-grade classes.[8]

Follow-up

What happened to the STAR students as they proceeded through school? Pate-Bain thought that the benefits that came from being in a small class would last, even through high school. But it was also possible that, in the end, students who had small classes in the K-3 years wouldn't do any better than the children in the standard-size classes. The STAR experiment would not be completed, and its results certainly would not be conclusive, without this follow-up. But gathering the records needed to carry it out was laborious and expensive. In 1995, Pate-Bain,who had "retired," and I formed Health and Education Research Operative Services (HEROS), Inc., to start collecting and analyzing data about students who had been part of STAR. Again, with the help of Tennessee legislators and the Tennessee Department of Education, along with funding from private foundations, we were able to continue collecting data from grades 5 through 12 on the STAR students and enter these data into the master database.

First we had to determine what types of data were available, how they could be obtained, and some of the questions that could be answered from these data. In 1997, Pate-Bain, Elizabeth Word (director of the STAR consortium), and I conducted a pilot study as preparation for a major follow-up study. We found that positive results continued for STAR small-class students through grade10.[9] This was encouraging. However, the biggest benefit of the pilot study was that it gave us a good idea about the procedures we would need to use in collecting the various types of data on a large scale. Although standardized test scores were readily available from the Tennessee education department, we could get other crucial data only from individual school districts (e.g., type of high school diploma and grade-point average) or even individual schools (e.g., participation in extracurricular activities and reasons for dropping out of school).

In 1997, the Tennessee education department provided us with the standardized test score data for STAR students from grades 5 through 12, and Finn and Achilles conducted analyses of the long-term effects of small classes using the new data.They found that in grades 4,6,and 8—after all pupils had returned to regular-size classes—STAR students who entered small classes in kindergarten had better long-term outcomes than those who began in first grade.Also, there were statistically significant differences in achievement between students who attended small classes for one,two,three,or four years.Long-term effects were significant on some tests in some grades (4,6,and/or 8) for pupils who attended small classes for *three years*, and on all tests in all grades for pupils who attended small classes for *four years*.[10] (See Table 2.)

Table 2
Long-term Advantages of Attending a Small Class for Four Years (K-3)
Reported in Months of Schooling

	Reading	Math	Science
Grade 4	9.1 months	5.9 months	7.6 months
Grade 6	9.2 months	8.4 months	6.7 months
Grade 8	1 yr. 2 mo.	1 yr. 1 mo.	1 yr. 1 mo.

The new data provided by the state department of education also included information about retention for STAR students.They show that the children who attended small classes were approximately 2.5 percent less likely to have been retained during grades K-8 than those who attended the larger classes. If money is the issue—as the people holding the purse strings always told Pate-Bain—these results suggest that smaller class size pays off: The average cost of one year of schooling in Tennessee is about $4,600, so each time a child fails a grade we're spending

an extra $4,600 on that particular child.

Pate-Bain and I have also been collecting information and records about STAR students from their high school years. As we realized from the 1997 pilot study, the process is laborious and involves phone, fax, mail, e-mail, and personal visits to schools.At this point we have collected more than 3,000 student records from individual schools and school districts,and over 8,000 from the state department of education. We will continue collecting high school records on STAR students through 1999, and if funding permits, we will try to track down students who have moved out of state. Since the database does not yet include all the records we eventually hope to have, findings about STAR students'performance in high school are not yet final. However, given the thousands of records already in the database, we feel confident that the conclusions we reach will be reliable.

Our preliminary findings show that STAR students who attended small classes in K-3 were more likely, as high school students, to be enrolled in advanced classes and honors courses (e.g., foreign languages, geometry, and honors English) than STAR students from the larger classes. They were also more likely to rank in the top 10 percent of their graduating class and to receive honors diplomas. The most statistically significant results from our preliminary findings relate to graduation and dropout rates: Approximately 72 percent of the students from smaller classes and 66 percent of the students from larger classes graduated on schedule in spring 1998. (See Figure 3.)

This goes hand in hand with our finding that students from small classes are less likely to drop out of high school.There is approximately a 5 percent difference between the number of students from small classes who dropped out and the number of students from larger classes who dropped out. (See Figure 4, page 36.) This is another area where we can show the financial advantage of small classes—this time for the students themselves.According to the U.S.Census Bureau,a high school graduate earns over $5,000 per year more than a high school dropout. And this figure does not take into account dropouts who do not become employed and/or who receive welfare.

Do smaller classes in the early grades have any impact on college attendance? Alan Krueger, a professor of economics at Princeton University, recently conducted a study to determine which STAR students took Scholastic Assessment Test (SAT) or American College Testing (ACT) college entrance exams and were, therefore, planning to attend college.He found that approximately 44 percent of students in the smaller classes did so, in comparison with 40 percent of students from the larger STAR classes.The effect of smaller class sizes, as measured by students taking college entrance exams, was most dramatic for minority students. African-American students who had the advantage of small classes in their K-3 years took these exams far more often than black students who had attended the larger classes (40.2 percent for students from small classes, compared with 31.7 percent of students from standard-size classes).Further, small classes appear to close the black/white ACT/SAT testing gap by more than 54 percent.[11]

Our research on class size is not complete. We will continue collecting high school data for STAR students through 1999. These data will allow us to fine-tune our preliminary work and to take a good look at many other small-class outcomes, such as type of diploma—regular, special education, or honors—attendance, type of coursework,and retentions. We also plan to follow these students after high school to look at factors such as college graduation, employment, welfare, and incarceration rates. Perhaps,when these results are in, we will be able to attach a true monetary value to small classes that will put an end to critics' continuing questions about their cost. But our research has implications beyond financial issues, important though they are.

The U.S. has been in the midst of a serious and long-term effort to reform our schools. Our biggest challenge is to improve learning for all but especially for minority students. Their achievement continues to lag behind, creating a situation that is terribly inequitable and full of danger for them and for our society. The evidence from STAR shows that small classes in the early years help all children, but that low-income, minority children benefit especially. To date, there is no other plan for reform that can offer this kind of

Figure 3
Percentage of STAR Students Graduating on Schedule

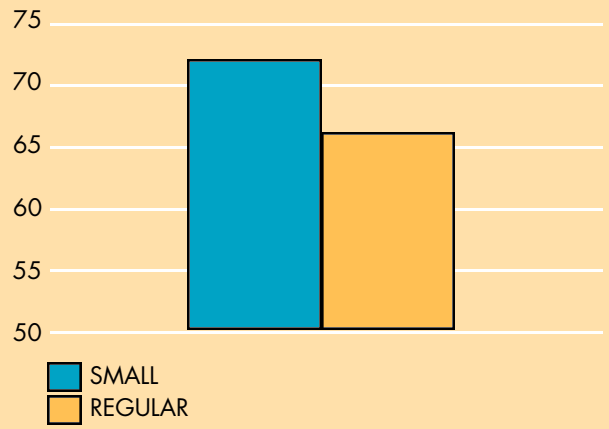
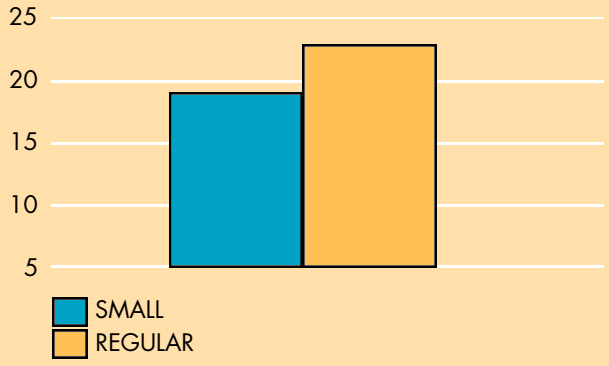


Figure 4
Percent of STAR Students Not Completing High School



assurance. We do not contend that smaller class size is a panacea. But it is a powerful and proven way to start. □

Endnotes

[1] Mosteller, F. (1995).The Tennessee study of class size in the early grades. *The Future of Children: Critical Issues for Children and Youths,* 5(2), pp. 113-127.

[2] Johnson,L. (1978) South Carolina first-grade pilot project 1976-77: The effects of class size on reading and mathematics achievement. Office of Research Report Series Vol. 1, No. 35, Columbia, S.C.: South Carolina Department of Education, p. 112 (ED 021 252).

[3] Mueller, J.D., et al. (1987). *Project Prime Time: Final report of Indiana's class-size experiment.* Bloomington,Ind.: Indiana University.

[4] Glass, Gene V., and Smith, Mary Lee. (1979). Meta-Analysis of Research on Class Size and Achievement, *Educational Evaluation and Policy Analysis,* vol. 1, no., 1, pp. 2-16.

[5] Because of teacher and parent complaints about student compatibility, many students from the regular and regular-aide classes were randomly reassigned between these two class types at the beginning of first grade. Since there were no regular-aide effects in kindergarten,this did not contaminate the study. During the entire study, only 100 small-class students (out of more than 11,000) were reassigned to one of the other class types.

[6] Achilles,C.M.(March 1997).What does it take? Presentation at the Plain Talk About KIDS Conference.The Los Angeles Center for Development and Learning (LACDL).

[7] Finn, J.D.,Gerber, S.B.,Achilles,C.M.,Boyd-Zaharias, J.,(May 1999). Short and long-term effects of small classes. Paper presented at the conference on Economics of School Reform.Copies available from J.D. Finn (finn@acsu.buffalo.edu).

[8] Finn, J.D. (1998). Class Size and Students at Risk:What is known? What is next? Washington, D.C.: U.S. Department of Education, OERI, National Institute on the Education of At-Risk Students; Finn, J.D. (August 1993).School Engagement and Students at Risk. Washington, D.C.: National Center for Education Statistics (93-470); Finn, J.D.,and Cox, D. (Spring 1992). Participation and withdrawal among fourth-grade pupils. *American Educational Research Journal, 29*(1), 141-162.

[9] STAR Follow-up Studies (Sept. 1997).

[10] Finn et al. (1999). *Op. cit.*

[11] Krueger, A.B. and Whitmore, D.M.(April 1999). Executive Summary: The Effect of Attending a Small Class in the Early Grades on College Attendance Plans (akrueger@princeton.edu).

Project STAR required a great deal of effort and cooperation from everyone who made up the STAR consortium, including its director Elizabeth Word of the Tennessee Department of Education; principal investigators C.M. Achilles, currently a professor at Eastern Michigan University; John Folger of Vanderbilt University; Fred Bellote of the University of Memphis; and John Johnston, who replaced Professor Bellote when he retired; primary statistician Jeremy Finn of SUNY Buffalo; and research associate Nan Lintz of the University of Tennessee at Knoxville.

Those making STAR follow-up studies possible include Elizabeth Word; Deborah Gilliam of the Tennessee Department of Education; and HEROS research associate and database expert DeWayne Fulton, who has worked on STAR since the beginning.

Tennessee Department of Education personnel who have been essential include Ben Brown and Lynisse Patrick.

SAGE: A Small-class Experiment in Wisconsin

In 1996-97, Wisconsin initiated Student Achievement Guarantee in Education (SAGE), a five-year program that seeks to improve the academic achievement of children living in poverty by reducing the student-teacher ratio in K-3 classes to one teacher for every fifteen students. Like STAR, the Tennessee class-size experiment, the Wisconsin program is a controlled experiment; unlike STAR, it calls for changes in addition to the reduced student-teacher ratio. SAGE schools had to agree to provide students with a rigorous academic program and after-school activities, and teachers with professional development. SAGE schools also agreed to institute accountability measures. SAGE is in effect in 30 schools throughout Wisconsin. Schools selected to participate receive up to $2,000 for each low-income student enrolled in a SAGE classroom.

SAGE students are tested in reading, language arts, and math using the Comprehensive Test of Basic Skills (CTBS). Results thus far, are promising.The following findings come from a report by principal investigator Alex Molnar (*Smaller Classes and Educational Vouchers*, Harrisburg, Pa.: Keystone Research Center, June 1999):

- In 1996-97 and again in 1997-98, SAGE first-graders scored significantly higher in all areas tested, with effect sizes in the range of 0.1 to 0.3.
- From spring 1997 to spring 1998, SAGE second-graders' scores increased more than those of students in comparison schools but not by statistically significant amounts. Over the first two years, SAGE second-graders showed statistically significant gains in language arts, mathematics, and total score but not in reading.
- The benefit of the SAGE program is especially strong for African-American students. In 1997-98, African-American students in SAGE classes increased their average total score by 52 points, compared with 33 points for African-American students in control schools.These higher scores in SAGE schools narrowed the achievement gap between white and African-American students; at the same time, the gap in comparison schools widened.
- In 1997-98, there was no significant difference between student achievement in SAGE first-grade classes with one teacher and up to fifteen students and another SAGE option—first-grade classes with two teachers and up to thirty students. If this finding is sustained in subsequent years, it would mean that school districts lacking the resources to build new classrooms could get the benefits of smaller classes by adding teachers to their larger classes.
- Interviews of teachers and principals, classroom observations, and other qualitative comparisons of teachers in SAGE and regular schools suggest that SAGE teachers know each of their students better, spend less time managing their classes, have more time for instruction, and are more likely to individualize their instruction.

More information about SAGE and copies of preliminary reports are available on the SAGE web site (http://www.uwm.edu/SOE/centers&project/sage/). See also Molnar, Smith et al., "Wisconsin's SAGE Class Size Reduction Program," *Educational Evaluation and Policy Analysis*, Summer 1999.