How Small Classes Help Teachers Do Their Best

Edited by

Margaret C. Wang
Distinguished Professor and Director
Temple University Center for Research
in Human Development and Education

and

Jeremy D. Finn
Professor, Department of Counseling and Educational Psychology
State University of New York at Buffalo

Jointly published by the
Laboratory for Student Success at
Temple University Center for Research
in Human Development and Education
and the
U.S. Department of Education

9

Wisconsin's Student Achievement Guarantee in Education (SAGE) Class Size Reduction Program: Achievement Effects, Teaching, and Classroom Implications

Alex Molnar, Philip Smith, John Zahorik, Amanda Palmer, Anke Halbach, and Karen Ehrle
University of Wisconsin-Milwaukee

This chapter reports the impact of reduced class-size on student achievement and on classroom events during the first two years of Wisconsin's Student Achievement Guarantee in Education (SAGE) program.

BACKGROUND

In December 1993, the Wisconsin Superintendent of Public Instruction appointed a 34-member Urban Initiative Task Force. The group was charged with the responsibility of recommending to the state superintendent ways of improving the academic achievement of children attending urban schools in the state.

On July 1, 1994, the task force issued its report. Instead of focusing solely on schools in urban areas, the task force looked at schools throughout the state, as long as those were schools serving children living in poverty. The task force made four recommendations:

- Reduce class size to 15 students per teacher beginning with kindergarten and first grade in the 1996–97 school year.
- Establish "lighted schoolhouses" open from early in the morning until late in the evening.
- Develop rigorous curricula.
- Create a system of staff development and professional accountability.

The program envisioned by the task force was to be voluntary. Any school that wished to participate had to agree to implement all four recommendations and enter into a contract with the Department of Public Instruction (DPI) detailing how it intended to implement each of the recommendations. The performance of each school was to be reviewed annually by the DPI. Any school that repeatedly failed to meet the terms of its contract could have its program funding cancelled. The task force also recommended an independent 10-year longitudinal study of the program's impact on student achievement and a citizens implementation and evaluation committee to monitor program implementation and assess its success.

The task force class-size recommendation called for reducing class size in Grades K-5 and Grades 6-8 in core subjects. All schools in the state serving 30% or more children living below the poverty level were to be eligible. Approximately 454 schools in 127 districts would have met this poverty standard. The first-year cost of the program was estimated to be as high as $28.6 million.

The state superintendent accepted the task force recommendations and in his fiscal year 1995–97 biannual budget submission proposed to fund the program using reallocated monies. The superintendent's budget recommendations were not made part of the budget submitted by the governor to the legislature. The legislative members of the Urban Initiative Task Force, however, were successful in adding the task force's proposed program to the state budget, albeit in a considerably scaled back form known as The Student Achievement Guarantee in Education (SAGE) program. SAGE was designed as a five-year pilot project that began with the 1996–97 school year. The SAGE legislation required and funded annual evaluations of the program.

All school districts in Wisconsin that had at least one school with 50% or more children living below the poverty level were eligible to apply for participation in SAGE. Within those districts, any school that had 30% or more students below the poverty level was eligible to become a SAGE school. All eligible districts were allowed one SAGE

school, with the exception of the Milwaukee school district that was allowed up to ten SAGE schools. Districts could apply for participation during the 1995–96 school year. No schools were to be allowed to enter the program after the initial eligibility period. Funding was set at a maximum of $2,000 per low-income student enrolled in the grade levels (K-3) affected by SAGE. No school district that applied for SAGE was rejected, and during the 1996–97 school year, 30 schools (7 in Milwaukee) in 21 school districts began the program in K–1. Second grade was added in 1997–98. Third grade was added in 1998–99.

In 1997 the legislature authorized the expansion of SAGE and in the autumn of 1998, an additional 49 schools entered the program. The SAGE evaluation, however, involves only the original 30 schools that entered the program in the autumn of 1996 and a group of 14-17 non-SAGE comparison schools. The 1996–97, 1997–98, and 1998–99 first-grade cohorts in these schools will be followed through third grade (see Table 1).

EVALUATION DESIGN

Although SAGE has four requirements, reduced class size is clearly at the heart of the program. There is considerable variation in the way in which SAGE schools implement the rigorous curriculum, lighted schoolhouse, and staff development provisions of the program. Reduced class size is, by comparison, the standardized and most dramatic reform in SAGE and thus the intervention that can most clearly be linked to test performance. The 1996–97, 1997–98, and 1998-99 SAGE evaluation

Table 1
Cohort CTBS Testing by Grade Level 1996-2001

1996–97	1997–98	1998–99	1999–00	2000–01
K	K	K	K	K
1 (fall & spring)	1 (fall & spring)	1 (fall & spring)	1	
	2 (spring)	2 (spring)	2 (spring)	2
		3 (spring)	3 (spring)	3 (spring)

reports contain considerable data on the three non-class size SAGE interventions (Maier, P., Molnar, A., Smith, P., & Zahorik, J., 1997; Molnar, A., Smith, P., & Zahorik, J., 1998; Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K., 1999). Only the impact of class-size reduction on student achievement and classroom events is considered here.

To determine the impact on student achievement of SAGE class-size reductions, the SAGE evaluation uses a quasi-experimental, comparative change design. This design was used because it was not possible to randomly assign students and teachers to classrooms, to keep classroom cohorts intact from year to year, and to control the class-size requirement in other ways. The comparison schools selected had normal size classes but resemble SAGE schools in family income, achievement in reading, K–3 enrollment, and racial composition. Fiscal constraints and the lack of incentives available prevented the use of matched-pair comparison schools, although comparison schools are from districts participating in the SAGE program. In 1996–97 there were 17 comparison schools, and in 1997–98 there were 14.

Students

The gender, race, and other characteristics of students in SAGE and comparison schools are displayed in Table 2. Many students withdraw from SAGE and comparison schools during the year, while others enroll. Those students who remained in their schools for an entire year are labeled "ongoing." As Table 3 shows, enrollment in comparison schools was slightly more stable than in SAGE schools. The number of ongoing and newly enrolled students recorded during spring data collection is reported in Table 4.

Classrooms

SAGE schools reduced class size in several ways in order to meet statutory requirements. The SAGE legislation defines class size as "the number of pupils assigned to a regular classroom teacher." In practice, reduced class size has been interpreted as a 15:1 student-teacher ratio (number of students per teacher in one classroom). Implementation occurred in the following ways:

Table 2
Characteristics of SAGE and Comparison Students 1996–97 and 1997–98

Characteristic	Percent of Students 1996–97		Percent of Students 1997–98	
	SAGE	COMPARISON	SAGE	COMPARISON
Gender				
Female	48.6	49.4	48.4	48.5
Male	51.4	50.6	51.6	51.5
Race/Ethnicity				
African American	24.8	32.9	26.3	24.7
Asian	5.7	5.5	5.2	5.6
Hispanic	6.6	8.0	6.5	10.0
Native American	11.7	1.4	10.3	1.5
White	48.8	49.0	43.8	52.2
Other	1.6	2.7	2.0	2.3
Subsidized Lunch Eligibility				
Free	57.7	49.4	54.0	43.4
Reduced	10.9	9.9	10.6	8.9
Not Eligible	31.4	40.7	35.4	47.7
Repeating Grade	3.2	2.6	2.7	2.0
English as Second Language	8.2	4.9	7.9	6.4
Referred to M-Team	13.6	9.2	9.6	6.8
Exceptional Education Need	13.1	9.7	10.0	7.1
Individualized Education Plan	8.2	5.5	8.8	5.6

Note: From "Evaluating the SAGE Program: A Pilot Program in Targeted Pupil-Teacher Reduction in Wisconsin," by A. Molnar, P. Smith, J. Zahorik, A. Palmer, A. Halbach, & K. Ehrle, 1999, *Educational Evaluation and Policy Analysis, 21*(2), p. 166. Copyright 1999 by the American Education Research Association. Reprinted with permission by the author.

- A *Regular* classroom refers to a classroom with one teacher. Most *regular* classrooms have 15 or fewer students, but a few exceed 15.
- A *Shared-Space* classroom is a classroom that has been fitted with a temporary wall that creates two teaching spaces, each with one teacher and about 15 students.
- A *2-Teacher Team* classroom is a class where two teachers work collaboratively to teach as many as 30 students
- A *Floating Teacher* classroom is a room consisting of one teacher and about 30 students, except during reading, language arts, and mathematics instruction when another teacher joins the class to reduce the ratio to 15:1.

Table 3
Enrollment Changes in SAGE and Comparison Schools by School Year (Number of Students and Percentage of Students)

	SAGE				COMPARISON			
	1996–97		1997–98		1996–97		1997–98	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Ongoing	2943	81.4	2455	42.3	1706	85.3	1402	44.3
Withdrew	397	11	1093	18.8	178	8.9	589	18.6
Enrolled	274	7.6	2262	39	115	5.8	1175	37

Note: From "Evaluating the SAGE Program: A Pilot Program in Targeted Pupil-Teacher Reduction in Wisconsin," by A. Molnar, P. Smith, J. Zahorik, A. Palmer, A. Halbach, & K. Ehrle, 1999, *Educational Evaluation and Policy Analysis,21*(2), p. 167. Copyright 1999 by the American Education Research Association. Reprinted with permission by the author.

Table 4
Number of Students in SAGE and Comparison Schools by Grade and School Year

	SAGE		COMPARISON	
	1996–97	1997–98	1996–97	1997–98
Kindergarten	1494	1524	820	676
First Grade	1723	1567	1001	985
Second Grade	NA	1541	NA	868
Totals	3217	4632	1821	2529

Two other types of classroom organization were also used in the SAGE program but to a very limited extent. They are the *Split Day* classroom consisting of 15 students and two teachers—one who teaches in the morning and one who teaches in the afternoon—and the *3-Teacher Team* classroom where 45 students are taught collaboratively by three teachers.

The types of classrooms are displayed in Table 5. SAGE classes range in number of students from 7 to 38. A few SAGE classrooms exceed the 15:1 student-teacher ratio, but only by one or two students. The student-teacher ratio for SAGE and Comparison Classrooms is shown in Table 6.

Data Collection Instruments

To provide information about the processes and products of the SAGE program for 1996–97 and 1997–98, a number of instruments were used as part of the evaluation. A description of the test and non-test instruments used in 1996–97 and 1997–98 follows.

1. *Comprehensive Test of Basic Skills (CTBS).* The Comprehensive Test of Basic Skills (1991) complete Battery, Terra Nova edition, Level 10, was administered to first-grade students in SAGE schools and comparison schools in October 1996 and May 1997. In 1997–98, first-grade students were subject to Level 10 testing in October and Level 11 testing in May. Second-grade students were tested only in May, at Level 12. The purpose of the first-grade October administration of the CTBS was to obtain baseline measures of achievement for SAGE schools and comparison schools. The complete battery includes subtests in reading, language arts, and mathematics. The CTBS was chosen as an achievement measure because it is derived from an Item Response Theory (IRT) model that allows comparison of performance across time. Moreover, it is one of a few instruments that attempts to minimize items biased against minorities and educationally disadvantaged students. Kindergarten students were not tested because of (1) concerns over the reliability and validity of standardized test results for kindergarten-aged

Table 5
Number of SAGE Classrooms by Type, Grade, and School Year

	Regular		Shared Space		2-Teacher Team		Floating Teacher		Split Day		3-Teacher Team	
	1996–97	1997–98	1996–97	1997–98	1996–97	1997–98	1996–97	1997–98	1996–97	1997–98	1996–97	1997–98
Kindergarten	50	89	2	4	24	22	3	2	0	0	1	0
Grade 1	61	84	8	8	18	23	7	2	2	0	0	1
Grade 2	NA	82	NA	6	NA	21	NA	3	NA	0	NA	1

Table 6
Teacher-Student Ratio for SAGE and Comparison Classrooms

Number of Students Per Teacher	Percent of SAGE Classrooms			Percent of Comparison Classrooms		
	1996–97 Grade 1 %	1997–98 Grade 1 %	1997–98 Grade 2 %	1996–97 Grade 1 %	1997–98 Grade 1 %	1997–98 Grade 2 %
7-13 Students	18	28	30	9	12	17
14-16 Students	62	64	63	0	7	6
17+ Students	20	8	7	91	81	77

children and (2) the view expressed by many kindergarten teachers that standardized tests would have a traumatizing effect on their students. The effects of SAGE on kindergarten students will be determined when they are tested as first-grade students the following year. All inferential results for the Terra Nova are reported in standard score format.

2. *Student Profiles.* This instrument, completed in October and May, provided demographic and other data on each SAGE school and comparison school student.
3. *Classroom Organization Profile.* Completed in October, this instrument was used to record how SAGE schools attained a 15:1 student-teacher ratio.
4. *Principal Interviews.* These end-of-year interviews elicited principals' descriptions and perceptions of effects of their schools' rigorous curriculum, lighted-schoolhouse activities, and staff development program, as well as an overall evaluation of the SAGE program.
5. *Teacher Questionnaire.* Administered in May, this instrument obtained teachers' descriptions and judgments of the effects of SAGE on teaching, curriculum, family involvement, and professional development. It also was used to assess overall satisfaction with SAGE.
6. *Teacher Activity Log.* This instrument required teachers to record classroom events concerning time use, grouping, content, and student learning activities for a typical day three times during the year.
7. *Student Participation Questionnaire.* In both October and May, teachers used this instrument to assess each student's level of participation in classroom activities.
8. *Classroom Observations.* A group of first- and second-grade classrooms representing the various types of 15:1 student-teacher ratios and a range of geographic areas were selected for qualitative observations to provide descriptions of classroom events.
9. *Teacher Interviews.* Although in-depth teacher interviews were not part of the original SAGE evaluation design, they were added in 1997 because it became apparent that teachers had important stories to tell about their SAGE classroom

experiences. The interviews dealt with teachers' perceptions of the effects of SAGE on their teaching and on student learning.

ANALYSIS OF STUDENT ACHIEVEMENT OUTCOMES

Statistics Utilized

The SAGE evaluation design uses descriptive statistics and multivariate inferential statistics, including linear regression and hierarchical linear modeling. Regression models were used at the individual level of analysis to control certain variables and entering the SAGE participation variable last, thus isolating its effects from the other variables. In addition, hierarchical linear modeling was used to assess the class effects of SAGE; that is, these analyses specifically assess classroom effects rather than those of individuals. The classroom effects examined by this approach are of primary importance to the SAGE evaluation.

The inferential analyses use the scale score for the CTBS Terra Nova. For these tests, a significance level of .05 was used, and significant results are denoted by an asterisk (*). SAGE versus comparison analyses are divided into two major sections: (1) First-Grade Results and (2) Second-Grade Results. Within each of these sections, the following are presented: (1) descriptive statistics (pre- and posttest), (2) ordinary least squares regressions, (3) analyses of the scores of African-American students, and (4) hierarchical linear modeling. In addition, a "within SAGE" analysis relating to class organization is presented.

First Grade Results

Descriptive Statistics

Valid Test Scores. The number of first graders for whom the valid test scores are available was less than the total number of first-grade students. There are four reasons for this. First, the evaluation team presented schools with the option of allowing EEN and ESL students to take the test, even though the test may be inappropriate for these students. These scores were invalidated based on a Nonvalid/Missing Test Report, developed by the evaluation team and completed for all first-grade classes. Second, given withdrawals and enrollments during the school year, a number of students had valid pretest scores, but no posttest scores and vice versa. Third, some students took the reading and

language arts components of the CTBS, or the mathematics component, but not both. Consequently, total scores are unavailable for these students. Finally, some of the students did not complete the pretest, posttest, or both the pre- and posttests. The number of valid test scores for the first-grade students in both cohorts are presented in Table 7.

Table 8 provides descriptive statistics for the CTBS from the pre- and posttest results for both first-grade cohorts. Inferential tests comparing pretest (fall) scores for the SAGE and comparison groups showed no significant differences for either cohort. Tests comparing posttest (spring) scores for the SAGE and comparison groups revealed significant differences ($p < .05$) on all test scores for both cohorts

Regression Analysis

Regression Models. The effect of the SAGE program on student achievement, controlling for other factors, was tested through a series of ordinary least squares regression models for each subtest and for total scale scores. Control variables were entered into the models in blocks, with the SAGE participation student variable entered into the models last.

The first block of control variables included student scores on the pretest and school attendance, measured as number of days absent, as reported by teachers in the spring of each year. The second block of control variables included dummy variables for race/ethnicity, coded 1 if a student was of a certain race/ethnicity, and 0 if not. Dummy variables were included for African American and White. A residual race category, "other," was included in the constant term in the regression equations. Eligibility for subsidized lunch, as an indicator of family income, was also included in the second block of control variables. This variable was coded 0 if the student was ineligible, 1 if the student was eligible for reduced price lunch, and 2 if the student was eligible for free lunch (this variable was assumed to be interval level). In the final block, a dummy variable for SAGE participation was entered. This variable was coded 0 if a student was from a comparison school and 1 if a student was from a SAGE school. The subsequently presented hierarchical linear models show that SAGE had no effect after class size was taken into account. Therefore, any "SAGE" effect actually represents a "class size" effect. This may suggest that the other SAGE interventions (i.e., rigorous curriculum, lighted schoolhouse, and staff development) have had little impact on achievement.