FIRST YEAR RESULTS OF THE STUDENT ACHIEVEMENT GUARANTEE IN EDUCATION PROGRAM

EXECUTIVE SUMMARY

Submitted by the SAGE Evaluation Team
Center for Urban Initiatives and Research
University of Wisconsin - Milwaukee
Peter Maier
Alex Molnar
Stephen Percy
Phillip Smith
John Zahorik

Research Assistants
Greg Giglio
Sally Hochstein
Lisa Radtke
Laura Roskos
Mark Schill
Kathy Shields

DECEMBER 1997

For further information contact Alex Molnar, Center for Urban Initiatives and Research, University of Wisconsin - Milwaukee, P.O. Box 413, Milwaukee, WI 53201, (414)229-5916

I. INTRODUCTION

The Student Achievement Guarantee in Education (SAGE) evaluation is being conducted under contract to the Department of Public Instruction by the Center for Urban Initiatives and Research (CUIR) at the University of Wisconsin - Milwaukee. This is the first of five annual evaluation reports.

The purpose of the SAGE evaluation is to determine the effectiveness of the Student Achievement Guarantee in Education (SAGE) program in promoting academic achievement of students in grades K-3 in schools serving low-income children. The SAGE program was enacted by the Wisconsin legislature in 1995, with implementation in kindergarten and first grade beginning in the 1996-1997 school year. The SAGE statute [s. 118.43] requires participating schools to (a) reduce class size to 15 in grades kindergarten and one in 1996-97, grades kindergarten to two in 1997-98, and grades kindergarten through three in 1998-99 to 2000-2001; (b) keep the schools open from early in the morning to late in the day and collaborate with community organizations to provide educational, recreational, community, and social services (i.e., the "lighted schoolhouse"); (c) provide a rigorous academic curriculum to improve academic achievement; and (d) establish staff development and accountability mechanisms.

During 1996-97, the SAGE program was implemented in 30 schools located in 21 school districts throughout the state, as shown in Table 1. Over the course of the year it involved 3,614 students and 220 teachers in 190 kindergarten and first grade classrooms. The gender, race, and other characteristics of students in participating schools are displayed in Table 2.

Schools reduced class size in several ways. 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, implemented in the following ways:

- A *Regular* classroom refers to a classroom with 1 teacher. Most *regular* classrooms have 15 or fewer students, but a few exceed 15.
- A *2 Teacher Team* classroom is a class where two teachers work collaboratively to teach as many as 30 students.
- A *Shared Space* classroom is a classroom that has been fitted with a temporary wall which creates two teaching spaces, each with 1 teacher and about 15 students.
- A *Floating Teacher* classroom is a room consisting of 1 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.

Two other types of classroom organization were also utilized in the SAGE program, but to a limited extent. They are the *Split Day* classroom consisting of 15 students and 2 teachers, one who teaches in the morning and one who teaches in the afternoon, and the *3 Teacher Team* classroom where there are 37 students taught collaboratively by 3 teachers.

The types of classrooms and the enrollments in each are displayed in Table 3.
In sum, SAGE classes range in number of students from 9 to 37. A few SAGE classrooms exceed the 15:1 student-teacher ratio, but only by a few students. The average SAGE classroom contains 17.4 students.

The SAGE Evaluation

The SAGE evaluation plan for 1996-97 follows. Described are the purpose, design, instrumentation, and data collection plan.

Purpose

The main purpose of the SAGE evaluation is to determine if the SAGE program of 15:1 student-teacher ratios, rigorous curriculum, lighted schoolhouse, and staff development is of benefit to students in promoting academic achievement. The main questions that guided the evaluation effort for 1996-97 are the following:

1. What differences exist in student achievement between SAGE schools and comparison schools?
2. How was each of the four SAGE elements implemented?

    a) *15:1 student-teacher ratio* (type of classroom, teaching methods, student behavior)

    b) *Rigorous curriculum* (congruence with national standards)

    c) *Lighted-schoolhouse* (type and extent of before and after school programs)

    d) *Staff development* (type and extent of program)

The first question focuses on the ***product*** of the SAGE program, i.e., student achievement. The second question focuses on the ***process*** of the SAGE program, i.e., what happened in SAGE classrooms and schools that may, over time, help explain achievement variations and suggest future actions for teachers and administrators seeking to enhance student performance.

Research Design

A two-part formative evaluation is used to determine the effectiveness of SAGE. The first part focuses on reduced student-teacher ratio, the main variable of the SAGE evaluation, through a ***quasi-experimental, comparative change design***. The comparative change design utilizes a

treatment group (30 SAGE schools) which has implemented the 15:1 student-teacher ratio and a comparison group (16 non-SAGE schools) that is as identical as possible to the treatment group *except* for reduced student-teacher ratio. Changes in achievement over time, as measured by a standardized achievement pre-test (baseline) and repeated standardized achievement post-tests, are compared between the groups.

To carry out this design 16 comparison schools were identified. The selection of comparison schools was constrained by practical considerations. Originally, the evaluation research design called for "matched pairs;" that is, one comparison school for each SAGE school. However, because of limited incentives to encourage potential comparison schools to participate in the evaluation the matched pairs design was changed to a "matched group" design, which compares SAGE schools *as a group* to comparison schools *as a group*. Furthermore, the evaluation team intended to draw comparison schools from among all elementary schools in the state, but the lack of incentives foreclosed this strategy as well. Instead, comparison schools were selected from school districts participating in the SAGE program, for whom cooperation with the evaluation was a condition of participating in the program. Reliance on SAGE districts for comparison schools resulted in underrepresentation of rural schools in the comparison group, since rural districts have limited numbers of elementary schools from which to choose. Moreover, some of the rural schools in the comparison group have class sizes only marginally above the 15:1 ratio required in SAGE schools.

The specific method of identifying schools for the comparison group was to minimize Squared Euclidean Distance[1] between the following variables (Z-scored) for each school: percent scoring above standard on the Wisconsin Third Grade Reading Test; percent Asian, Native

[1] Distance $(X, Y) = \Sigma (X_i - Y_i)^2$

American, African American, and Hispanic; percent low income; and, total enrollment in grades K-3. Squared Euclidean Distances were computed for all SAGE schools and schools within SAGE districts. The first step was to check similarities among participating SAGE schools; a relatively homogeneous group of SAGE schools requires only a single matching school. Several relatively homogeneous groups of varying sizes were identified. The "best" matches were determined first by one non-SAGE match per SAGE group, and second, by pairwise matches for SAGE schools that did not fit a group. “Best” was defined as a combination of quantitative, research design, and practical considerations. Because some SAGE schools do not resemble any other schools in SAGE districts, particularly on racial composition. Squared Euclidean Distances were recomputed, omitting the variable rendering schools so dissimilar, or substituting the variable *percent white* in place of all other racial variables.

Difference of means tests between SAGE schools, as a group, and comparison schools, as a group, showed no statistically significant differences on any of the variables at the .05 level.[2] Similarly, difference of means tests between SAGE schools, as a group, and comparison schools, as a group, from student demographic data collected by the SAGE Evaluation Team showed no statistically significant differences on any of the variables at the .05 level. However, when these data were subjected to tests at the individual-level of analysis, as shown in the composite profile of SAGE and comparison schools in Table 4, the large increase in N yielded several statistically significant differences.

The largest difference is the percentage of Native Americans in the SAGE versus the comparison group. The over representation of African Americans in the comparison group

[2] The SAGE evaluation team selected, a priori, .05 as the critical threshold for all tests of statistical significance.

reflects the high proportion of Milwaukee schools in the comparison group, nearly 44 percent of all comparison group schools. Comparison group students are somewhat better off economically than SAGE students. Further, the comparison group has fewer Exceptional Education Needs (EEN) students and fewer English as a Second Language (ESL) students.

During the course of the 1996-97 school year records were compiled on 5613 students. Many students withdrew from SAGE and comparison schools during the year, while others enrolled. Those students who remained in their schools for the entire year are labeled "ongoing." As Table 5 shows, enrollment in comparison schools was slightly more stable than in SAGE schools. Moreover, in both SAGE and comparison schools, the number of students withdrawing exceeded the number of students enrolling during the year. Thus the number of ongoing plus newly enrolled students recorded during spring data collection totals 5038, distributed across schools and grades as shown in Table 6.

In addition to the comparative change design, the nature of the four SAGE program elements is being examined to explain variation in achievement among SAGE schools, classes, and students. This is accomplished through a **repeated measures, reflexive controls** design. SAGE schools, classes, and students are compared to themselves over time, as measured at the beginning of the treatment through an achievement pretest and other baseline measures and after the treatment begins through repeated achievement tests and other indicators.

Data Collection Instruments

To provide information about the process and product of the SAGE program for 1996-97, a number of instruments were created and administered as part of the evaluation[3]. These instruments are the following:

[3] See the *Evaluation Design Plan for the Student Achievement Guarantee in Education (SAGE) Program,*

1. *Student Profiles*. This instrument, completed in October and May, provided demographic and other data on each SAGE school and comparison school student.
2. *Classroom Organization Profile*. Completed in October, this instrument was used to record how SAGE schools attained a 15:1 student-teacher ratio.
3. *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.
4. *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.
5. *Teacher Activity Log*. This instrument, administered in October, January, and May, required teachers to record classroom events concerning time use, grouping, content, and student learning activities for a typical day.
6. *Student Participation Questionnaire*. In both October and May teachers used this instrument to assess each students' level of participation in classroom activities.
7. *Classroom Observations*. A group of first-grade classrooms representing the various types of 15:1 student-teacher ratios and a range of geographic areas was selected for qualitative observations in October and May to provide descriptions of classroom events.
8. *Teacher Interviews*. Although in-depth teacher interviews were not part of the original SAGE Evaluation Design, they were added because it became apparent that teachers had important stories to tell about their SAGE classroom experiences. The interviews, held in

---

August 13, 1996, for complete details.

May, dealt with teachers' perceptions of the effects of SAGE on their teaching and on student learning. The observed teachers served as the interview sample.

9. *Comprehensive Test of Basic Skills (CTBS).* The Comprehensive Test of Basic Skills (CTBS) Complete Battery, Terra Nova edition, Level 10 was administered to first-grade students in the 30 SAGE schools and the 16 comparison schools in October, 1996 and May, 1997. The purpose of the October administration of the CTBS was to obtain baseline measures of achievement for SAGE schools and comparison schools. The complete battery includes sub-tests 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 which 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 a) concerns over the reliability and validity of standardized test results for kindergarten-aged children, and b) 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.

The methods of data collection by type of school and grade are listed in Table 7.

The instruments identified in the SAGE Evaluation Design that are not used in this report, or used in only a limited way, were the Baseline Data Questionnaire, School Implementation Plan, Teacher Profile, and the Teacher Development Plan. These instruments and program requirements were either not completed, as in the case of the Teacher Development Plan, or were only useful in part, as in the case of the Baseline Data Questionnaire.

The remainder of this report provides the results of the evaluation of the 1996-97 SAGE program. Part II presents data about the effects of SAGE on student achievement in reading, language arts and mathematics. Part III describes what went on in SAGE classrooms. Part IV addresses rigorous curriculum, staff development and lighted schoolhouse programs.

II. EFFECTS OF THE SAGE PROGRAM ON STUDENT ACHIEVEMENT

The effects of the SAGE program on student achievement were evaluated by several methods. Analyses were conducted at both the individual-level and class-level of analysis. SAGE effects were assessed with both bivariate and multivariate statistical tests. Results are reported first for individual-level analyses, then for class-level analyses.

The number of first grade students for whom valid test scores are available is substantially less than the total number of students. First, the evaluation team presented schools with the option of allowing EEN and ESL students for whom the test may be inappropriate to take the test anyway. These scores were invalidated based on a "Nonvalid/Missing Test Report," developed by the evaluation team and completed by all first grade classes. Second, given withdrawals and enrollments during the school year, a number of students had valid pre-test scores, but no post-test scores, and vice versa. Third, some students took the reading and language arts components of the CTBS, or the mathematics component, but not both. Total scores are unavailable for these students. Finally, some students were absent for all of the pre-test, the post-test, or both. The number of valid test scores for the 1996-97 school year is presented in Table 8.

Pre-Test (Baseline) Results of Standardized Testing

Table 9 provides descriptive statistics on the **scale scores** from the pre-test, or baseline results. Scale scores can be used to measure student performance across all grade levels. Given the longitudinal nature of the SAGE evaluation, scale scores will serve as the primary measure of

student achievement.[4] To place the pre-test scale scores in context, **national percentiles** are also provided in Table 9. For example, the mean or average total scale score of 517.07 corresponds to a national percentile rank of 38.90. That is, the average first grade student in the SAGE evaluation scored as well on the CTBS as about 39 percent of students taking the test nationwide. Since the SAGE program was created in response to lower levels of achievement among low income students, this subaverage (below 50th percentile) performance on the baseline CTBS was expected.

The results from difference of means tests between SAGE and comparison student scale scores from the October CTBS are reported in Tables 10 through 13. Comparison school students scored slightly higher than SAGE school students on reading, language arts, and total score, and slightly lower in mathematics. However, none of the differences is statistically significant at the .05 threshold; we fail to reject the null hypothesis of no difference between SAGE and comparison school students on the pre-test. Since SAGE and comparison students are virtually equal in achievement at the beginning of the SAGE program, any subsequent differences in achievement tests that favor SAGE students may be more confidently attributed to the student-teacher ratio of 15:1 in the SAGE program.

Post-Test Results of Standardized Testing

As noted above, student populations varied in SAGE and comparison schools due to withdrawals and within-year enrollments. The post-test results are based only on those students

[4] CTB/McGraw-Hill uses a three parameter logistic model to create scale scores.The **total scale** score is computed by CTB/McGraw-Hill as the average of the three scale scores from reading, language arts and mathematics.

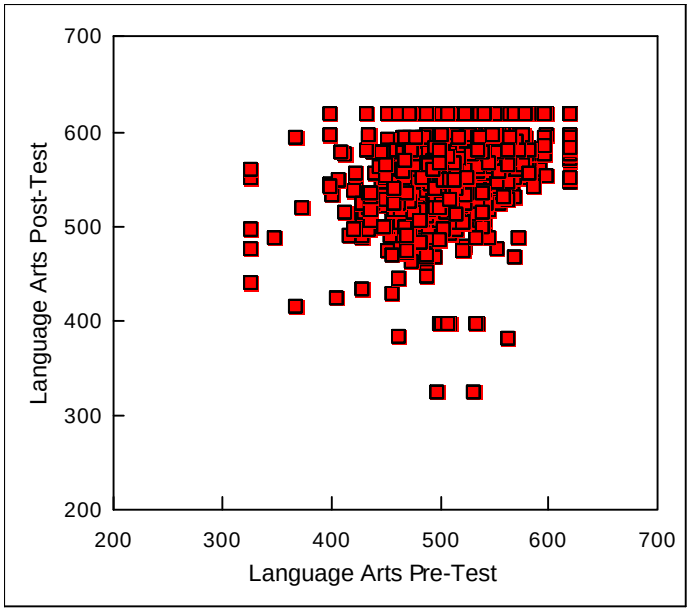
who remained in their schools for the entire 1996-1997 school year (88.2 percent of SAGE students and 87.0 percent of comparison school students took both pre- and post-tests).

Although the CTBS allows measurement of performance over time, with younger children different test levels can result in content-related invalidation. For example, when one attempts to compare students on level 10 and level 11 of the CTBS, the scores are comparable from a measurement point of view, but the contents of the two tests are not totally congruent. For this reason level 10 was used a both a pre-test and post-test measure. However, as a consequence of the decision to administer level 10 of the CTBS for both pre-test and post-test, a substantial number of students achieved perfect scores on the sub-tests of the CTBS.

Perfect scores introduce an element of uncertainty into comparative analyses; once a student reaches an achievement ceiling the extent to which a student *might have* achieved is unknown. This "restricted range" issue is, on balance, more problematic than the issue of content-related validity. Therefore, beginning in 1997-98 level 10 of the CTBS will be administered to first grade students in the fall, and level 11 will be administered in the spring.

As shown in Table 14, perfect scores are particularly prevalent in the Language Arts and Reading sub-tests. The ceiling effect on the CTBS Language Arts sub-test is portrayed graphically in figure 1. A perfect score in language arts equals 620, the point on the graph where a straight line appears on the post-test and, to a far lesser extent, on the pre-test.

Figure 1. The Ceiling Effect on the CTBS Language Arts Sub-Test



As Table 14 shows, students in SAGE schools disproportionately achieved perfect scores. Thus ***estimates of the effect of the SAGE program on student achievement are likely to be conservative.*** One approach to mitigating the ceiling effect is to conduct analyses first with all cases, then truncating the sample and repeating analyses with those cases performing at or below the 75th percentile on the pre-test. Students who achieved perfect scores on the post-test are predominantly those who scored in the top quartile on the pre-test, and whose change scores from pre-test to post-test are restricted. In language arts, for example, the average change in score from pre-test to post-test is 49.67. However, the average change score for students who scored in the top quartile on the pre-test is 22.59, whereas the average change score for the other 75 percent of students is 58.73. Thus the statistical tests of the effects of the SAGE program that follow are presented first for all cases, then for those cases scoring through the 75th percentile on pre-tests.

Difference of Means Tests

Tables 15 through 22 show the results of difference of means tests for each of the CTBS sub-tests and total scores. Unlike the difference of means tests for the CTBS pre-test, which showed no statistically significant differences between SAGE and comparison school students, ***statistically significant differences are found in favor of SAGE students for each sub-test, and for total scale scores*** on the post-test**.** These statistically significant differences are observed whether all students are analyzed, or the top scoring quartiles on the pre-test are excluded.

The largest difference in means is found on the mathematics sub-test, followed by language arts, and then reading. The largest gain in SAGE student scores from pre-test to post-test, relative to comparison school students, was in language arts, as shown in Table 23. The smallest relative gain for SAGE students from pre-test to post-test was on the reading sub-test.

Finally, the expectation that observed differences between all SAGE and comparison school students would be understated due to a ceiling effect was not met in all of the bivariate analyses. When the top scoring quartile on pre-tests were withheld from analyses, the differences between SAGE and comparison school students on the language arts and mathematics sub-tests were actually smaller than when all students were included. To reiterate, however, comparison school students are better off than SAGE students in terms of family income and potential impediments to educational achievement. It is therefore necessary to statistically control for some of these differences through multivariate analyses.

Multivariate Analyses

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 sub-test and for

total scale scores. Control variables were entered into the models in blocks, with the SAGE/comparison student variable entered into the models last.

The first block of control variables included 1) student score on the pre-test; 2) eligibility for subsidized lunch as an indicator of family income, coded 0 if student is ineligible, 1 if student is eligible for reduced price lunch, and 2 if the student is eligible for free lunch (this variable is assumed to be interval level); and school attendance, measured as number of days absent, as reported by teachers in May of 1997.

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, Asian, Hispanic, Native American, and White. A residual category, Other, is included in the constant term in the regression equations.

Finally, a dummy variable for SAGE or comparison school student was entered on the third block. This variable is coded 1 if a student is from a SAGE school, 0 if a student is from a comparison school.

Some limitations of the data should be noted here. First, some of the racial/ethnic variables create complications. The variable "Asian" is a gross indicator which fails to distinguish among various Asian sub-groups. For example, we are unable to distinguish Hmong students, who tend to be more disadvantaged, from other Asian sub-groups. Native Americans are only minimally represented among comparison school first grade classes (as few as 8 in one analysis). And many Hispanic students are limited in English proficiency and did not take the CTBS (including one entire first grade class in a Milwaukee Public School). Similarly, many exceptional education students did not take the CTBS, or completed the test but had their scores invalidated. Whether a

particular student took the CTBS, or had his or her scores invalidated, was left to the discretion of the teacher. Thus variables for exceptional education needs and limited English proficiency were not included in the regression models.

Regression Results

Results of the regression analyses are presented in Tables 24 through 31. For all analyses, for each of the CTBS sub-tests and total score***, membership in a SAGE school emerges as a significant predictor of student achievement*** on the post-test, while controlling for pre-test scores, family income, school attendance, and race/ethnicity. The magnitude of the effect of SAGE on student achievement, as denoted by the "**b**" coefficient, varies depending on the CTBS sub-test, and whether all students are analyzed, or the top scoring quartile on the pre-test is excluded.

Consistent with the difference of means tests, the largest effects of SAGE are found on the mathematics sub-test, while the smallest effects of SAGE are found on the reading sub-test. However, unlike the difference of means tests, regression results show the effect of SAGE is consistently larger when the top scoring quartile on the pre-test is excluded. For example, Table 28 shows the effect of SAGE on mathematics. The model predicts that a SAGE student will score 3.876 scale points higher than a comparison school student, after controlling for pre-test scores, family income, school attendance, and race/ethnicity. In Table 29, where the top scoring quartile on the mathematics pre-test is excluded, the model predicts that a SAGE student will score 4.63 scale points higher than a comparison school student, after controlling for pre-test scores, family income, school attendance, and race/ethnicity.

When all cases are analyzed the goodness-of-fit of the models (as denoted by the adjusted R^2 statistic), ranges from .24 in reading to .48 for total scale score. That is, in predicting total scale score on the post-test, the variables included in the model explain 48 percent of the variance in total scale scores. Most of the variance in the post-test scores is explained of course by the pre-test scores. Family income and absenteeism emerge as consistent and statistically significant predictors of performance on all sub-tests and total scale score. Race and ethnicity show some relatively large effects (as denoted by the b coefficients), but the effects are highly variable and are generally statistically insignificant (race is discussed further below)[5].

When the top scoring quartiles on the pre-test are excluded from analyses, the magnitude of the SAGE effect (b) increases for all sub-tests and for total scale score. In the case of total scale score, for example, the estimated effect of SAGE membership on post-test performance is +4.60, as opposed to an estimated effect of 3.30 when all cases are analyzed. The relationship between SAGE and post-test scores is more variable however, when the top quartile is excluded (as denoted by the lower values of T). Indeed, the goodness-of-fit of the models (adjusted R^2) is lower when the top quartile is excluded.

Whether all cases are analyzed, or the top scoring quartiles are excluded, membership in SAGE schools has a consistently positive, statistically significant effect on achievement on the CTBS.

African American Students

A precursor to the SAGE program is the Tennessee STAR experiment in reduced class size, a statewide initiative involving 7000 students over four years in grades kindergarten though

[5] Additional models were tested including interaction terms for SAGE student by race/ethnicity. The interaction terms were not statistically significant.

grade 3. One of the conclusions reached in the Tennessee experiment in reduced class size is that "the advantage of being in a small class is greater for minority students than for whites," (Finn and Achilles, 1990: 567). Note that no distinction is made among minority sub-groups. For reasons discussed earlier, analyses of SAGE effects by race and ethnicity are problematic, particularly for Asians, Hispanics, and Native Americans. Still, the "achievement gap" between white and minority students on standardized measures of achievement remains a source of considerable interest, among both scholars and policy makers.

Among minority students in SAGE and comparison schools, African Americans clearly comprise the largest sub-group -- roughly 25 percent of all SAGE students, and 33 percent of all comparison school students. The African American student population does not present the analytical problems of interpretation raised by other minority groups. In the analyses to follow, African American students are first compared across SAGE and comparison schools on CTBS sub-test and total scale scores. Second, African American students are compared to white students across SAGE and comparison schools on CTBS total scale scores.

Table 32 provides comparisons of means on the CTBS post-test, and change scores from pre-test to post-test. African American SAGE students scored higher than comparison school students on every sub-test, and on total scale score. The differences are, in the main, not statistically significant. The change scores, however, consistently favor SAGE students and are statistically significant. In other words, ***African American SAGE students scored lower on the CTBS pre-test than African American comparison school students, but made significantly larger gains than comparison school students from pre- to post-test, and surpassed African American comparison school students on the post-test***.

Concern over the minority achievement gap on standardized tests has occasionally been focused on African American male students. Table 33 further distinguishes African American SAGE and comparison school students by gender. A clear pattern emerged during the first year of the SAGE program. African American male SAGE students attained comparable or higher change scores from pre- to post-test than African American female SAGE students. At year's end African American male and female students scored virtually the same on the CTBS post-test. This result is quite unlike the scenario in comparison schools, where change scores for females exceeded change scores for males on every sub-test, and on total scores. Thus at year's end comparison school females scored substantially higher than males on the CTBS post-test.

African American and White Student Achievement on the CTBS

African American students, as a group, scored significantly lower than white students, as a group, on the CTBS pre-test total scale score, as shown in Table 34. This result holds for both SAGE and comparison schools, though the gap between African Americans and whites is larger in SAGE schools. When all cases are analyzed, African American SAGE students achieved greater gains on the CTBS total scale score than white SAGE students from pre- to post-test, closing the achievement gap (though the gap remains statistically significant). In contrast, African Americans in comparison schools achieved lesser gains and in comparison schools the achievement gap widened.

Given the ceiling effect discussed earlier, the analysis was repeated excluding the top scoring quartile on the pre-test total scale score. Regarding pre-test comparisons the achievement gap is narrower for both SAGE and comparison school students, though still statistically significant. Change scores, however, vary considerably between African American SAGE and

comparison school students. In SAGE schools, African American students who performed at, or below, the 75th percentile on the pre-test achieved the same change score as white students who performed at, or below, the 75th percentile on the pre-test. An achievement gap remained, but grew no larger over the course of the first year of SAGE. In comparison schools, the achievement gap widened, as was observed when all cases where analyzed[6].

Finally, the analysis was repeated excluding the top ***two*** scoring quartiles on the pre-test total scale score. The results are almost identical to those found when only the top quartile was excluded. Thus, even among the lowest scoring 50 percent of students on the pre-test, the achievement gap between African American and white students widened in comparison schools, but remained essentially unchanged in SAGE schools.

Hierarchical Linear Models

Many social science research endeavors involve hierarchical data structures. Hierarchical data structures are those in which individual units are nested within larger units, the latter being the unit of interest. The SAGE data are a prime example: students are nested within classrooms, and it is the classroom effect that is of particular interest to the SAGE project. Hierarchical data structures pose special analytical challenges in that data analysis at the individual level may result in a biased impression of the effect of the nesting unit (in the SAGE case, the classroom). At the root of this problem is the fact that different classrooms often contain a different number of students, thus those classrooms that contain a greater number of students have greater influence

[6] It is worth noting that African American SAGE students are disproportionately found in large, two teacher classes. The implications of this will become clearer in the next section, which uses a class size variable in hierarchical linear analysis.

over the results of analyses done at the individual level. In general, if the effects of the nesting unit, the class, is of interest this is not a desirable outcome. An analytical approach known as hierarchical linear modeling (Bryk & Raudenbush, 1992) is designed specifically to accommodate these types of data structures. Essentially, hierarchical linear modeling (HLM) estimates individual effects by analyzing data within each class and then provides a weighted average of these effects. The effects of the class are then estimated as if all classes contained the same number of students. HLM was used with the SAGE data to provide an alternative and less biased accounting of the initial effects of the SAGE experience on test scores. In these models variables associated with individual students are referred to as level-1 variables and those associated with the class are referred to as level-2 variables.

HLM Models

Analyses were conducted for each of the relevant criterion post-test scores: reading, mathematics, language arts, and total. For all analyses, the level-1 variables were pre-test score, socioeconomic status (SES) measured as eligibility for subsidized lunch, and attendance measured as number of days absent. The post-test scores were adjusted for these three variables at the individual level, therefore the effects may be thought of as being statistically independent of the effects of these variables. A number of different level-2 models, each containing different level-2 variables, were specified for each variable of interest.

Model A. *Class Size*
These models examined the effect of class size on the adjusted criterion score.

Model B. *SAGE*
These models examined the effect of SAGE participation on the adjusted criterion score.

Model C. *Class Size, SAGE*
These models examined the effect of SAGE participation on the adjusted criterion score after the classrooms were class size adjusted, viewed as the effect of SAGE participation beyond the class size effect.

Model D. *Class SES, Class Size*
These models examined the effect of class size on the adjusted criterion score after the classrooms were SES adjusted, be viewed as the effect of class size once the effects of the classroom SES are removed.

Model E. *Class SES, SAGE*
These models examined the effect of SAGE participation on the adjusted criterion score after the classrooms were SES adjusted, viewed as the effect of SAGE participation once the effects of the classroom SES are removed.

Model F. *Class SES, Class Size, SAGE*
These models examined the effect of SAGE participation on the adjusted criterion score after the classrooms were class size and SES adjusted, viewed as the effect of SAGE participation beyond the class size and SES effects.

It is important to note that the "class size" variable used in these analyses measures the **number of students** in each class, and not the student-teacher ratio. As discussed earlier, some SAGE classes contain a relatively large number of students (e.g., 30), and some comparison school classes contain a relatively small number of students (e.g., 16).

Table 35 provides a summary of the effects of each of the level-1 and level-2 variables for each of these analyses. Level-1 effects can be interpreted as the weighted average of the within classroom effects of the level-1 variables. Level-2 effects can be interpreted as the classroom effects of the level-2 variables of interest. Level-1 coefficients may be thought of as the average effect of the modeling variable on the criterion score at the individual level. For example, for the total scale score, each day absent resulted in a .205 point drop on the post-test (-.205 coefficient) for the individual. These effects vary from classroom to classroom, however. Results for all three sub-tests and the total score are fairly consistent. On average, attendance loss resulted in a drop

in post-test scores, lower SES resulted in lower scores and higher pre-test scores resulted in higher post test scores. No dramatic differences in these coefficients were observed across sub-test scores.

The coefficients associated with the level-2 variables can be thought of a classroom effects. For example in model A for the total score, an increase of one student in class size resulted in a drop of .879 points for the class average (-.879 coefficient). Likewise, SAGE participation resulted in a 6.397 point gain in the class average. A discussion of each model follows.

Model A. Depending on the test, an increase in class size of one person can be expected to produce a .5 to 1.2 point loss in average post test performance. The results for all scores shows this effect to be significant.

Model B. Participation in SAGE shows significant class average increases in post-test performance for the total score (6.4 points) and the math subscore (8.0 points). Results for the reading and language arts scores were somewhat below this and were not statistically significant.

Model C. Combining class size and SAGE participation in a single analysis isolates the effects that SAGE might have beyond those produced by lower class sizes. Again, with the exception of the language test, class size has a significant effect on class average post test performance. Once class size has been accounted for, SAGE has no significant effect on class average performance.

Model D. Since socio-economic status (SES) is known to have an influence on academic test scores, a surrogate for this variable was used as both a level-1 and level-2 predictor. The level-2 variable was the average SES for the class and estimates the effect of the overall class SES

level beyond that associated with the individual, which is accounted for in the level-1 model. This model combines class SES and class size and the results indicate that both have a significant effect on class average post test performance. The effect of a 1 point class average gain in SES equates to between 16 and 21 points on the average post test score, depending on the test (keep in mind that SES was measured on a three point scale - thus a one point difference on average would quite pronounced). The effect of class size in this context is not much different than when entered alone (see model A).

Model E. When class SES and SAGE participation are entered in the same level-2 model, both variables have a significant effect on class average post test performance. In this context, SES has a slightly greater effect that in Model D, possibly indicating that SAGE participation and SES are less highly correlated than class size and SES. The effect of SAGE participation on class average post-test scores beyond those produced by SES differences ranges from about 7 points to about 12 points depending upon the sub-test. In general, these effects are larger than when SAGE is the sole variable in the model (see Model B). The likely explanation for this is that, in general, SAGE classrooms have a lower SES that control classrooms, and once this is accounted for, the benefit of SAGE participation is amplified.

Model F. This model combines SES, SAGE participation and class size in a single analysis. For all sub-tests, SES once again has significant effects on the class average post test score. For the total score, both class size and SAGE participation have a significant effect of class average performance. Class size has a significant effect on class average performance for the mathematics sub-test. Neither variable made a significant contribution to the average for the

language arts and reading subscores once SES was accounted for. This is likely due to the fact that these two variables (and SES) are relatively highly correlated.

Analysis for Truncated Group

As noted earlier, the use of CTBS Level 10 used for post-test resulted in a significant proportion of ceiling effects for project participants. These effects likely had an undetermined influence on the results displayed in Table 35. As with the individual level analyses, the HLM models were also applied to the data after removing the top quartile of scorers on the pre-test. This procedure eliminated most of the cases that were observed with a ceiling effect, and therefore the results are expected to be free from any bias introduced by these effects. In addition, for purposes of the HLM analyses, classrooms with fewer than 5 students after the elimination procedure were dropped from the analysis. This was done to avoid having a very few individuals in the classroom determine the effects for the classroom.

It should be noted that the regression coefficients generated from this truncated sample may be biased. Assuming linearity, it can be shown (e.g., Linn, 1982) that the regression coefficient resulting from the variable used for truncation (in this case pre-tests) will be unaffected by this procedure. However, the coefficients associated with other variables, which were subjected to incidental selection (to the extent that they correlate with pretest) can be expected to be attenuated. In addition, in all cases the standard error of the coefficients can be expected to be higher, therefore statistical significance is more difficult to attain. In all cases, then, the results from the truncated sample can be thought of as conservative estimates.

Table 36 shows the HLM modeling results for the 75 percent sample. In the majority of cases, the regression coefficients in Table 36 are attenuated with respect to the corresponding

values in Table 35, as expected. There are several instances where the values actually increase, however. These differences are most likely due to sampling error. Even though the standard errors are expected to rise in the truncated situation, the pattern of significant coefficients is quite similar across the models in Table 35 and Table 36. Consequently, the interpretation of the results changes little from the full sample.

III. LIFE IN SAGE CLASSROOMS: THE REDUCED STUDENT-TEACHER RATIO

To accurately comprehend the SAGE program it is important to understand how SAGE schools structure classrooms and implement features of the SAGE initiative (i.e., 15:1 student-teacher ratio, rigorous curriculum, lighted-schoolhouse, and staff development). The focus of this section is on ***process***, *i.e,* what went on in SAGE schools and classrooms rather than on the effect SAGE had on student achievement. In particular, this section is centered on the reduced student-teacher ratio implemented by SAGE schools.

Contained in this section of the report is a description of teaching and learning in SAGE kindergarten and first-grade classrooms. Data collected from teacher interviews, classroom observations, teacher activity logs, and teacher questionnaires are reported.

Teacher Interviews

Thirty first-grade teachers from 13 schools in 8 districts were interviewed during May 1997. This sample consisted of 18 individual interviews, and 6 interviews of teacher teams who taught in 30:2 student-teacher ratio classrooms. The teachers selected to be interviewed were those who served as the observation sample of the SAGE evaluation effort, except for two teachers whose schedules did not permit interviews. Of the represented classrooms, 14 were *15:1 Regular* classrooms, including 3 classrooms that contained both first grade and kindergarten students, 4 were *15:1 Shared Space* classrooms, 5 were *30:2 Team Taught* classrooms, and 1 was a *30:2 Floating Teacher* classroom.

The interviews, which were tape recorded and transcribed, were 20 to 45 minutes in length and focused on three main questions. Teachers were asked to describe 1) the extent to which their teaching changed as a result of having fewer students, 2) the extent to which they believed their students' achievement improved as a result of being in a class with fewer students, and 3) changes they anticipated in their teaching for the 1997-98 school year. Findings regarding each of the questions follow.

Teacher Change

All of the interviewed teachers, except two teacher teams, indicated that their teaching had changed as a result of having a reduced class size. These two teams stated that their basic teaching style had not been altered, but they described many adjustments that they had made in teaching, which were consistent with the changes described by the other 22 teachers. The changes that the teachers described related to discipline, instruction, and personal enthusiasm.

Discipline. Although one teacher felt that the amount of time devoted to discipline had not changed from previous years when she taught a larger class, all of the other interviewed teachers said that they spent much less time in dealing with student misbehavior. Some teachers stated that misbehavior had nearly vanished from their classrooms.

Several explanations were given for the reduction in student misbehavior. With only 15 students they can get the attention of the class more easily, teachers indicated. They can see what every student is doing. They can have direct eye contact with students and can be physically close to students. This leads to identifying problems early and dealing with them instantly, teachers said. Further, because the class is small a family atmosphere develops in the classroom. A different relationship emerges as students come to respect each other. In addition, teachers who

team taught in 30:2 Team Taught classrooms remarked that during those portions of classroom time when all 30 students were being taught as a group by one teacher, the other teacher was able to focus exclusively on student behavior and take action if needed.

Examples

> *Well, it's wonderful not having to stop instruction to do discipline. I mean that's probably one of the biggest plusses, that learning still goes on while another adult deals with the problem.*
>
> *Behavior is probably not very much of a problem any longer .... It's basically because you got a small number and you're on top of them all the time. You're monitoring them all the time. So, they know how to behave now.*

Instruction. A result of greatly reduced need to discipline students was substantially more time devoted to instruction, teachers indicated. Every SAGE teacher interviewed remarked that he or she was able to devote more time to instruction this year. A few suggested that less "paper work" associated with small class size also contributed to increased instructional time. More instructional time, teachers stated, permitted them to be less rushed in their teaching. They could spend more time interacting with students, reteaching when necessary, and providing more and varied learning activities. The main consequence of increased instructional time, however, was an increase in individualized instruction.

Examples

> *There is definitely more time on instruction. Just having fewer bodies in the classroom, there are fewer, ah, fewer problems arise and so there can be more time devoted to instruction.*
>
> *It definitely changed, you know. I do have more time that we're spending, you know, specifically doing instruction.*
>
> *Now I feel as if I have time to really facilitate as well as interact with kids.*

When teachers talked about how having a student-teacher ratio of 15:1 affected their teaching, the topic of ***individualization*** was mentioned the most often and generated the most emotion. All of the teachers agreed that they now could turn to the needs of individual students.

A class with fewer students enabled teachers to diagnose the learning needs of individual students and to diagnose them earlier. The teachers remarked that they knew their students'abilities better and that they came to know each student as a person. In addition to diagnosis, having fewer students also permitted teachers to teach students on an individual basis. They were able to get around to work with each student and they could do it frequently. Students were not required to wait idly for the teacher's attention. Those students who understood the lesson were given accelerated tasks while those who had difficulties or problems were remediated.

Besides this type of tutoring individualization, small class size resulted in individualization in another sense, teachers indicated. With fewer students each student gets more turns, to share ideas, to answer teacher questions, to ask questions of others, and to read aloud. Increased participation of this sort permits teachers to see individuals' present level of understanding and to take needed action, and it permits students to clarify their thinking on the basis of the feedback they receive.

Examples

It was much easier to pinpoint what students need.

Oh my gosh. I get to return so many more times. I mean it could be the same lesson, but I come back to them more than once to see how they're doing. So, I might work with them and have the time to work with them one-on-one.

Most of the time everybody gets to have something they're really interested in brought out. I mean, even if we're just having a discussion on a topic everybody will get to say something about it because there's time for that, because there are only 14 kids.

Well, with comparing this year to last year, I think that this year I was able to get around to more kids and see the mistakes right away and address them right away instead of waiting until I pick the papers up.

In addition to individualization, another area that most teachers believed had changed was ***content emphasis***. All but one teacher said that because they had smaller classes they were able to teach more content and teach it in greater depth. Several mentioned that they had moved into the second-grade reading curriculum and books. A few also mentioned that they were able to introduce thematic teaching.

Examples

This year we finished up with grade one and we went through the second book.

I think the kids are getting so much further than I've seen first graders at this point in the year.

Another instructional change mentioned by some of the interviewed teachers was increased use of ***student-centered activities***, however, far fewer teachers mentioned this area than individualization. These teachers believed that smaller class size enabled them to provide interest centers, more hands-on activities and the use of manipulatives, give students choices in tasks, provide more opportunities to solve problems, and engage in more activities that require creativity.

Examples

I can do a lot of, like I said, hands-on and that type of thing, things I wouldn't dare attempt with a large class.

Now I have kids in cooperative groups ... learning from each other, working together, sharing each other's materials. And, manipulatives, I really really like them to work on manipulatives.

Personal Enthusiasm. An area related to instruction about which teachers had strong feelings was teacher enthusiasm as a consequence of having small classes. Teachers indicated they had a much more positive attitude toward teaching and had much more energy and motivation regarding teaching because they were able to develop personal relationships with students and they could see substantial educational growth in their students. Some teachers also mentioned that they experienced less stress because they had fewer students to whom they had to attend. This resulted in fewer papers to correct and less work to be done at home in the evening.

Examples

> *This year has been much more positive. Part of that is because of the success of the children because that is the goal. When they are successful, then that makes you want to teach. That success is an upper in itself, and that makes the whole experience more enjoyable.*
>
> *I think that it gives you less stress because when you're teaching and trying to do a good job, you're worried about the students. You're worried about them and trying to help them. It's a lot easier to give your attention and help to 15 kids than it is 30 kids, and that has to bring down the stress level.*

Student Learning

There was nearly unanimous agreement among the interviewed teachers that student academic growth was affected by reduced class size. Teachers remarked that their students achieved at a high rate in reading, language, and mathematics. Because a solid foundation of reading skills could be built early, students were able, as has been mentioned, to proceed to second grade materials. Some teachers also said that student writing ability improved considerably as did their problem-solving ability in mathematics.

Examples

> *Their writing skills are much higher than in the past. That's because of all of the extra practice that they got this year.*

I think that their skills are stronger because they just get more attention, all of them.

They're so far, it's like they're beginning second grade .... As for a like reading goes, as far as their comprehension of different things, or as far as math goes, all over.

Anticipated Change

The teaching changes teachers revealed that they plan to make during the 1997-98 school year are of two kinds. Some teachers, after a year of teaching small classes, have learned the demands of this new environment and are considering how they can improve the practices they used during the first year of the program. They indicated that they need to plan more carefully because students learn and progress so quickly in this setting. Further, as the new first-grade classes will contain large numbers of children who have experienced small classes as kindergarten students, they know that planning an accelerated curriculum will be necessary. Other teachers seem to be more satisfied with the teaching they employed during the first year but now want to expand their use of student-centered teaching. These teachers mentioned that they intend to use, for example, more guest speakers, curriculum integration, creative activities, problem solving, readers' workshop, and student decision making.

Examples

I think what I would do is let the kids make more decisions, let them be more problem solving.

I would probably plan more to bring more people into the classroom like speakers to talk about things that I haven't had a chance to plan.

Classroom Observations

Classroom observations were made in 26 first-grade classrooms, including 2 classrooms that also contained some kindergarten children, in October and April during the 1996-97 school year. These classrooms, taught by a total of 31 teachers, were selected to reflect the variety of types of 15:1 student-teacher ratio classrooms in the SAGE program, and a range of geographic areas. Of the 26 classrooms 15 were *Regular* classrooms, 15 were *Shared Space* classrooms, 5 were *2 Teacher Team* classrooms, and 1 was a *Floating Teacher* classroom. The 26 classrooms were located in 13 schools from 9 school districts.

The October observations lasted from one to three hours. The purpose of these observations was to become familiar with the classrooms and to record a reading or mathematics lesson to serve as a baseline measure of classroom events. Upon completion of each observation, the observer wrote an expanded narrative account based on the notes recorded during the lesson. These accounts were transcribed and analyzed using previous research on class size as well as constructivist teaching theory as a guide. Previous research suggests that achievement benefits of reduced class size may be related to less time needed for managing the classroom, fewer students with whom the teacher must interact, and greater homogeneity of student needs. Constructivist teaching theory suggests that for student understanding to occur instruction must be based on students' prior knowledge, make students active participants, provide feedback on students' constructions, and encourage student reflection. This procedure yielded a set of categories which was then used to complete the analysis of each observation. The main categories of ***individualization***, ***engagement***, and ***management*** as well as subcategories are displayed in Table 37.

The April observations were more focused observations. These observations, which lasted from 30 to 90 minutes, used the categories established during the first observation as a guide. Observers looked for instances of these categories of behavior but also recorded other prominent behaviors or events.

Data from the two sets of observations follow. First, the classroom behavior of the total group of observed classrooms is presented. Second, classroom behavior by type of SAGE classroom is discussed.

Total Classroom Behavior

The observed behaviors for the total group of classrooms from both the fall and spring observations are presented in Table 38. The findings are expressed in total frequencies and in mean percents. Because of varying length of observations, the frequency with which different teachers used particular behaviors, and the observation style of different observers, data for individual classrooms were standardized by converting behavior tallies for each category to percents of total behaviors used and then computing mean percents to facilitate total group analyses.

Table 38 reveals that few changes were observed in teacher classroom behavior from October to April in the three areas of individualization, engagement, and management. By comparing frequencies it appears that there was increased use of nearly all categories of behavior from October to April, but this increase is most likely a function of focusing on the categories revealed by the first observation.

Individualization. Table 38 shows that considerable individualization occurred in observed SAGE classrooms. Although the teachers do not regularly permit students to choose or

create their own learning activities, they frequently provide help to individual students and actively involve many, if not all students, in classroom events. Also, but to a lesser degree, they monitor student progress and subgroup the class or single out individuals for special attention. Changes that appear from October to April are that participation increased while both subgrouping and total group instruction decreased somewhat. An example of a lesson in which individualization is prevalent is the following:

Individualization Example

> In the first portion of this observation, the teacher introduced the lesson to the entire class. The class was working with different numbers which could be added together to make 11 and 12. The teacher individually called on students to come up with answers to her examples. When the second student that she called on struggled, the teacher offered assistance. When the third student responded with "4+7=11" the teacher said to him, "Right!" The teacher continued explaining what the class was going to be doing, including a discussion of math "fact twins." She continued to call on students to give examples, ending this portion of the lesson having called on half of the students present.
>
> The class then broke into groups of two to complete a cooperative worksheet using fact twins. The sheet involved using 11 chips in various groups of two. It was arranged that one student would attempt the first question while the other student checked the answer. The roles were then reversed for the second question, and so on. As the students worked, the teacher walked around the room, providing assistance when necessary. This ranged from helping them in their own management to specific discussion of the lesson to modifying the activity for those who completed it quickly. One group that could not decide who should go first and were arguing about it was reminded by the teacher of a way to settle the dispute. The students used the rock/paper/scissors game to make this determination. The teacher then worked with a pair, helping them complete both the initial number sentence and its fact twin. When one group finished the 11's quickly, the teacher got another chip so that they could work on the twelve's. She told this group "good job!" Finally, as groups began to finish, she encouraged them individually to work on something on their own. Some began reading. Others began looking at pictures of the field trip they had recently taken. The teacher continued to monitor the entire class, helping those that needed it, as well as checking all students' homework. The lesson ended as the bell rang and the students lined up for recess. (15:1 Regular classroom)

Engagement. Taken together, the analysis categories of listening, practicing, and responding represent more teacher-centered teaching because students are assigned a passive role in learning, while gaming, manipulating, creating, dialoging, problem solving, and the remaining engagement categories represent more student-centered teaching because students are more actively involved in learning activities. In both the October and April observations the observed classrooms were dominated by teacher-centered teaching. However, there is a slight decrease in teacher-centered teaching and a corresponding increase in student-centered teaching from October to April. In particular, listening and practicing are used less, on average, while responding is used more in terms of teacher-centered teaching. In student-centered teaching the average percents increase for dialoging and problem solving but decrease for creating.

Engagement (Teacher-Centered) Example

> In this team-taught classroom, the 30 children sat at their desks as the teacher directed their attention to the blackboard on which she had written the letter "*J*." The teacher asked the class to look at the Wordwall and find a word that began with that letter. A student went to the Wordwall and pointed out the word "*jump*." The teacher then directed all the students to write *jump* on their papers. As the students did this both teachers circulated among the tables helping students where needed. The process was repeated four more times with different students called on each time to provide a word from the Wordwall. After the students had written all the words the teacher told them they are going to put them in ABC order. She called on one student for the first word (*GO*) and then had all the students write that word. She continued the process until all five words were alphabetized. The last portion of the lesson consisted of adding "*ing*" endings. To begin this portion the teacher began hopping up and down, and asked the students what she was doing. "Jumping!" they called out. She then had them go through all five words adding the "*ing*" endings. As each word is written, both teachers circulated to check that all students have written it correctly. When a student was having difficulty, the teacher (*not the one leading the lesson*) remained at the student's desk providing the extra assistance needed in order to keep up.
>
> When this portion of the lesson ended, the class was divided into three groups. Six children left the room to meet with the reading resource teacher. Three children worked with one teacher on one side of the room, while the remaining children worked with the

other teacher on the side of the room. One group read a story from their readers with different students reading different pages. As they went through the story, the teacher complimented the children who read well and helped those who had difficulty. She made sure that all students understood the content of the story. At the same time, in another portion of the room, the other group of three students worked with the story of the three little pigs, cutting out pictures and pre-printed words which tell the story. The teacher of this group helped the students individually match the words with pictures, with the ultimate goal of having each student create a book of the story. In each portion of the lesson, the teacher drove the activity. She asked questions and the students responded to the questions directly. (30:2 Team-Teaching classroom)

Engagement (Student-Centered) Example

This lesson began with the teacher distributing bags of geometric shaped blocks consisting of triangles, squares, trapezoids, etc. The teacher encouraged the children to experiment with the blocks to make patterns of their own. The students immediately constructed their own patterns. The teacher then passed out sheets which had two patterns pre-printed on them: a star and an octagon. The students were then asked to arrange their blocks in these shapes, beginning with the star. The teacher monitored, providing help when needed, and encouraged students to use different blocks to form the same shape. The students worked on their own, raising their hands when they wanted the teacher to check their work.

The teacher then used the overhead projector and her set of blocks to show one way of ma king the star. She then invited students to come to the overhead and show the blocks that they used in creating the star. Four different children demonstrated their constructions on the overhead, as the teacher praised them by saying, “Good Jeffrey!” and “I like the way that you are trying to find a different way.”

The lesson continued as the students began working on the octagon pattern. This time the teacher checked the patterns as she monitored, rather than using overhead demonstrations. Finally, students were instructed to create their own patterns with the blocks, draw them on a sheet of paper, and then later, have other students try to make them. The students become engaged in making their own patterns using a variety of blocks. The teacher continued to walk around the room, monitoring and helping. The class session ended, however, before students could challenge each other with their newly created patterns. (15:1 Shared-Space classroom)

Management. The data regarding management are difficult to interpret because of the large increase in the average percent for the teaching category labeled "permits" in the April observation data. Undoubtedly, more instances of students sharpening pencils, taking bathroom

breaks, and engaging in other out-of-seat behaviors on their own volition occurred, but the occurrence of this type of behavior appears to be exaggerated by observer attention due to its addition to the observation guide for the April observation. It can be seen in Table 38, however, that observed SAGE teachers are much more positive than negative in their classroom management. Praising, reminding students about behavior that is expected, warming the classroom climate, and permitting students free classroom movement make up nearly all of the classroom management. Reproving students or cooling the classroom atmosphere through sarcasm or ignoring students is used very infrequently. From October to April, however, there is a decrease in several types of positive management. This could be a result of both the end of the school when students may become less attentive and also it may be an unavoidable consequence of the huge increase in the permit category of behavior. That is, as students more freely move about the classroom abuses of freedom may require a response from the teacher.

Management Example 1

> In one classroom, as the noise level of the students increased, one of the teachers said to the students, "Matt is really focusing on those dots, he isn't talking. That's what you need to do." Later, she said, "Nice job Rodney, he's working so quietly." Toward the end of this particular lesson the teacher said, "I'm checking for superstar behavior. Wow! I could pick a bunch of kids from each table. Superstars are always ready!" (30:2 Team-Teaching classroom)

Management Example 2

> In another classroom, the teacher asked, "who is ready?" When a few voices were still audible, she said to one student, "I like the way you are listening!" She went around the room praising students' good behavior. (15:1 Regular classroom)

Classroom Behavior in Different Types of SAGE Classrooms

The observed average percent frequencies for each of the four main types of SAGE classrooms are presented in Table 39. Because of the small number of classrooms observed it

would be a mistake to generalize from these data to other classrooms of a particular type of student-teacher ratio, but the findings displayed may suggest trends for future analyses.

In general the findings for the four main types of SAGE in each of the three areas of individualization, engagement and management echo the findings for the total group of observed classrooms, as would be expected. There are some observed differences among the types, however. In terms of ***individualization***, *15:1 Shared Space* classrooms use monitoring and provide help more and have students participate less than the others. The *15:1 Regular* classrooms subgroup for special attention less than the others except for the one *30:2 Floating Teacher* classroom. For ***engagement***, *30:2 Team Taught* classrooms use teacher-centered teaching the most, while *15:1 Shared Space* and *30:2 Floating Teacher* use manipulatives and dialogue more than the others. Regarding ***management***, the *15:1 Shared Space* classrooms use much more praise than the others, but permit students to move about the classroom less than the others. The *30:2 Team Taught* teachers use behaviors to warm the climate more than the others.

Teacher Activity Logs

Activity Logs designed to provide descriptions of typical school days were completed by SAGE kindergarten and first-grade teachers three times (October, January, and May) during the 1996-97 school year. The logs required teachers to record classroom activities at 15-minute intervals for a complete school day in four areas: time use, grouping, content, and student learning activities.

A total of 638 logs were completed and returned. Of these, 218 (K=97, 1st=155, combined grade=6) were completed in October; 200 (K=93, 1st=101, combined grade=6) were completed in January; and 220 (K=100, 1st=116, combined grade=4) were completed in May.

For this analysis, October logs were viewed as a baseline measure of classroom activity while January and May logs were combined to form a measure of classroom activity reflective of the effect of reduced class size.

Overall Results

Table 40 presents mean percent scores for time use, content, grouping, and student learning activities for all SAGE teachers. Mean percent scores were determined by converting frequency of category use for each teacher to percents based on total frequency and then computing an average percent use for each category for the total group or subgroup of teachers.

As can be seen in Table 40, time use for all teachers is dominated by ***instruction***. More time is spent on instruction than time spent on routines, planning, and personal matters combined. This finding for instructional time is stable from October to May, as are the results for routines, planning, and personal activities.

In terms of ***grouping***, whole class instruction is the most prevalent organizational form teachers use to develop content and skills, yet it only accounts for roughly half of the time spent on instruction. Small group and individual instruction account for most of the other half. Over the year small group instruction increased while whole class, individual, and combined grouping decreased slightly.

Time spent on ***content***, as revealed in Table 40, was constant over the year with twice as much time spent on reading-language arts compared to mathematics. Reading-language arts taught as separate lessons consume about thirty percent of the instructional time while mathematics consumes bout 20 percent. Altogether, including integrated lessons, reading-language arts and mathematics account for about 75 percent of the instructional time.

Overall, the main ***student learning activities*** used were practicing, listening, receiving help, creating, manipulating, and dialoging. The activities used less frequently were problem solving, answering, reporting, receiving critique, and reflecting.

Of the 12 student learning activities, 4 are most closely identified with teacher-centered teaching: listening, practicing, receiving help, and answering. These 4, which comprise 25 percent of the options, account for 50 percent of the activities reported.

Those activities that decreased slightly in use after October were receiving help and receiving critique. Practicing was the one activity that increased in use over the year.

Results for Kindergarten and First Grade

Table 40 also presents results for kindergarten and first grade separately. The results show that in main findings the two groups are quite similar, but that some differences exist. For example, kindergarten teachers spend comparatively less time in instruction and comparatively more time on routines. They also use whole class and individual grouping less and small groups more often than first-grade teachers, especially in the Fall. Reading and language arts was the dominant content area in kindergarten, but much more time is devoted to this content area in first grade. Kindergarten teachers used comparatively more integrated content than first-grade teachers,

In regard to student learning activities, the two grade levels are remarkably similar. Although both emphasize the use of practicing, listening, receiving help, creating, manipulating, and dialogue, they differ somewhat in the extent to which some of these activities are emphasized. First-grade teachers appear to use listening, practicing, receiving help, and answering, the set of

behaviors representing more teacher-centered teaching more, and creating and manipulating less than kindergarten teachers.

Results for combined kindergarten and first grade classrooms most often reveal a middle position compared to kindergarten and first grade findings. Two exceptions evident in the combined January and May logs are that they use individual grouping less but problem solving more than other classes.

Results for Different Types of SAGE Classrooms

Results for the four main types of classes are reported in Table 41 in the form of mean percents. Combined January and May logs for *Regular*, *2 Teacher Team*, *Shared Space*, and *Floating Teacher* classrooms are basically alike but some differences are evident as can be seen. The *Regular* classrooms differ from the others in that whole group instruction is used more often as is integrated content. The *2 Teacher Team* classrooms use small groups more and individual instruction less than the others. They also are comparatively high in their use of integrated content. The main features of the *Floating Teacher* classrooms are that they use planning and evaluation, individual grouping, receiving help, and answering more frequently along with more time spent on reading-language arts than the other classrooms. Integrated content and creating were used comparatively less. In relation to use of more teacher-centered teaching *(listening, practicing, receiving help, and answering*), the *Floating Teacher* group is the highest. The *Shared Space* classrooms spend the most time on instruction and the least on planning and evaluation. They also spent the least time in small groups, but the most in combined classes. In terms of student learning activities they are comparatively high in receiving help and

comparatively low in problem solving. In terms of activities most closely associated with teacher-centered teaching, they are the lowest.

Teacher Questionnaire: Classroom Teaching

The first section of the Teacher Questionnaire, which dealt with the effects of reduced student-teacher ratio on classroom teaching, reveals results similar to those of the other classroom measures. This section, which contains two parts, was completed by 206 kindergarten and first-grade teachers. The first part required teachers to rate their agreement with each of 11 statements concerning classroom practices. The second part asked them to rank the 16 statements by identifying up to three statements that represent the most significant ways their teaching had changed as a result of reduced class size.

As can be seen in Table 42 the results show that teachers felt that all of the teaching behaviors listed were affected positively by reduced class size. The behaviors that were seen as being affected most, based on ranking, were reduced management-more teaching, individualization, assessing progress, diagnosing learning problems, and covering content in greater depth.

Student Participation

Student participation in the classroom was measured by a Student Participation Questionnaire. SAGE kindergarten and first grade teachers completed the questionnaire for each of their students in both fall 1996 and spring 1997.

The Student Participation Questionnaire consists of 16 items, measured on a 1 to 5 scale, designed to assess student behavior and participation in SAGE classrooms (see Table 43 for a list of questions). Following principal component analysis (see Table 43), 15 of the 16 questions were

combined into two additive scales. As shown in Table 44, one scale comprises 9 questions and measures the extent to which student behavior is "On Task." A second scale comprises 6 questions and measures the extent to which students are engaged in "Active Learning." Further analysis shows the On Task and Active Learning scales to be highly reliable indicators (as measured by the alpha coefficients shown in Table 44).

Student participation in the classroom represents both process and product. The questionnaire measures the extent to which individual students are engaged, or actively participate in the learning ***process***. But one of the ways in which a smaller student-teacher ratio could raise academic achievement is by first increasing students' level of participation in the learning process. Student participation represents an intervening measure between reduced student-teacher ratios and academic achievement; increasing student participation is an achievement in itself.

Student participation has indeed increased over the course of the first year of the SAGE program. Descriptive statistics on the fall 1996 and spring 1997 student participation scales are provided in Table 45 (because the two participation scales comprise a different number of items, the scores were transformed to a common scale ranging from 0-100). The mean On Task student participation score increased by 8.9 percent from fall 1996 to spring 1997; the mean Active Learning student participation score increased by 12 percent from fall 1996 to spring 1997. Perfect (maximum) scores on the participation scale also increased from fall to spring. In fall, 143 (4.5 percent) SAGE students attained perfect scores in Active Learning, and 114 (3.6 percent) had perfect scores in On Task behavior. In spring, 345 (12.0 percent) students had maximum scores in Active Learning, and 201 (7.0 percent) attained perfect scores in On Task Behavior.

The Active Learning and On Task behavior scales were entered into OLS regression models (first grade only) to test their effect on achievement on the CTBS (see Tables 46-49). ***After controlling for pre-test scores, attendance, eligibility for subsidized lunch, and race/ethnicity, both Active Learning and On Task behavior emerge as significant predictors of achievement on all sub-tests and total score on the CTBS.*** Active Learning had the largest effect on reading scale scores (b=.46). An increase of a single point on the Active Learning scale (on a 1-100 scale) predicts nearly a half a point increase in the reading scale score. The largest effect of On Task behavior is found for reading and mathematics (b=.37 for both).

Since the Student Participation Questionnaire was not administered to comparison school students the impact of SAGE on increasing student participation cannot be determined. Some increase in participation may be expected to result naturally as a function of a child's maturation. Further research and analysis of the antecedents and covariates of student participation is necessary.

Conclusion

Taken together the teacher interviews, classroom observations, teacher activity logs, and the teacher questionnaires provide a picture of teaching and learning in a 15:1 student-teacher ratio classroom. What emerges after one year of the SAGE program is a classroom where discipline problems and classroom management are greatly reduced, and when classroom management is needed, it is overwhelmingly positive. The direct beneficiary of this reduced time spent on managing the class is increased time spent on instruction, i.e., on actually teaching. Further, the increased instructional time that is now available to teachers is used to attend to the learning needs of individual students. As much as half the time in classrooms teachers are helping

individual students with difficulties they encounter or extending their learning beyond minimal competencies, checking their work and monitoring progress, and providing opportunities for them to become actively involved in learning by articulating current understandings and receiving feedback. The increased instructional time also permits greater emphasis on quantity and depth of reading-language arts and mathematics content. Well over 50 percent of the instructional time is specifically devoted to these areas.

The type of instruction that students encounter in SAGE classrooms is predominantly teacher-centered. Listening, practicing, receiving help, and answering account for between 50 to 75 percent of the teaching-learning that occurs. Although teachers indicated that their use of more student-centered activities such as creating, manipulating, and problem solving increased because of reduced class size, and there is evidence of use of these behaviors from both the observations and logs, student-centered teaching only plays a supplemental role in most SAGE classrooms.

Several pictures do not emerge from a composite of the interviews, observations, logs, and questionnaires. Although all teachers said some changes had taken place in their teaching during the first year of SAGE and observational and self-report log data substantiated these changes, a major change in classroom events from October to May was not observed. A possible explanation might be that the October data, rather than representing a baseline, show adjustments that teachers had already made in their teaching. That is, they quickly made changes in their teaching during September and part of October prior to initial data collection. These changes may have been fine-tuned throughout the year, but the first-year response to reduced class size in

rudimentary form may have already been established by October when the first observation occurred.

Another picture that does not emerge is a large swing to student-centered teaching, a change that some might expect as an result of reduced class size. Most SAGE teachers appear to have made real, substantial changes in their teaching. However, these changes are not of the magnitude of substituting one set of coherent practices for another. Instead, SAGE teachers appear to have enriched their teaching and student learning with student-centered teaching, but the major change they have made is to use their teacher-centered teaching with individuals, a change that fewer students has permitted by reducing the need for management and increasing instructional time.

Still another picture that has not emerged is a clear difference among the four main types of SAGE classrooms. They differ in isolated behaviors and events, but generally they appear to reflect the patterns found for the total group of SAGE teachers. Differences among classroom types, however, as well as differences from the beginning to the end of the year in general and in relation to type of teaching may emerge over time. The present data represent only one year of the five-year SAGE program. Longitudinal data to be collected during the next four years are needed to bring all of the pictures of classroom life into sharper focus.

IV. OTHER SAGE INTERVENTIONS

During the first year of the SAGE program the focus of participating schools has clearly been on implementing the reduced student-teacher ratio. The other SAGE interventions -- rigorous curriculum, staff development, and lighted schoolhouse programs -- have been attended to by SAGE schools in varying degrees. This section of the report briefly describes the state of the other SAGE interventions near the end of the first year of the SAGE program.

Rigorous Curriculum

The Teacher Questionnaire and Principal Interviews, both completed in May 1996, are the sources of data regarding rigorous curriculum.

The Teacher Questionnaire contains a section on classroom curriculum designed to determine the congruence of SAGE classroom curricula with professional curriculum standards developed by the International Reading Association (IRA), the National Council for Teachers of English (NCTE), and the National Council for Teachers of Mathematics (NCTM).

Teachers were asked to indicate, on a five point scale, the extent to which items on the questionnaire described the curriculum in their classrooms. A classroom's curriculum would be, according to the teacher's self-report, considered perfectly congruent with the professional curriculum standards of the IRA, NCTM, and NCTE if the teacher responded "5" on all curriculum items on the Teacher Questionnaire.

A total of 211 SAGE teachers completed the curriculum section (K=88, 1st=109, mixed grade=4). The responses of each teacher yielded mean scores for reading/language arts and for mathematics. As can be seen in Table 50 the mean scores for reading/language arts ranged from

2.11 to 4.84 with an overall average score of 3.88 (78 percent of perfect congruence). Mean scores for mathematics items ranged from 1.62 to 4.83 with an overall average score of 3.61 (72 percent of perfect congruence).

The responses of first grade and mixed grade teachers indicates that their reading/language arts curricula are more congruent with professional standards than SAGE kindergarten curricula. The responses of kindergarten, first grade and mixed grade teachers resulted in no significant difference in the degree to which their curricula were congruent with professional standards in the area of math.

It is interesting to note the three reading/language arts items that produced the highest means - and were thus closest to the recommended professional standard. According to SAGE teachers, students in their classrooms were most likely to be: 1) taught to apply a variety of decoding strategies (mean score=4.65), 2) taught the names of the parts of books (mean score=4.74), and 3) encouraged to choose books they are interested in reading (mean score=4.78). In mathematics the most notable divergence from the pattern mean responses at the 3 or 4 level was in response to the item on calculator use. This item had a mean response of 2.17 which suggests that calculators do not play much of a roll in SAGE kindergarten and first grade math instruction.

All SAGE principals were interviewed. Their responses to curriculum related questions suggests that, for SAGE principals, a rigorous curriculum includes basic skills, problem solving, and higher level thinking. Only a handful of principals seemed to believe that the curriculum of their school was rigorous. However, most SAGE principals regarded parts of their curriculum as strong.

Staff Development

The section of the Teacher Questionnaire that dealt with staff development asked teachers about their individual level of professional development as well as the extent to which their school district provides staff development programs. The questions were derived from standards for staff development published by the National Staff Development Council, in cooperation eith the National Association of Elementary School principals.

With respect to individual professional development, the SAGE contract requires that teachers and administrators develop professional development plans that focus on how they will help improve student academic achievement. Despite of the contractual requirements, roughly 60 percent responded that they had no "personal, formal, written development plan." As can be seen in Table 51, those teachers who had individual plans were more likely to engage in activities aimed at professional development. Teachers were asked if, over the past school year, they had participated in any of twelve activities aimed at further developing their teaching skills. Of the twelve professional development activities, four were found to be significantly related to existence of an individual professional development plan. Teachers who reported that they had a formal individual development plan were more likely to have 1) collaborated with other schools and institutions, 2) conducted research connected to teaching, 3) attended a professional conference or skill building workshop, and 4) taken a course for graduate or CEU credit.

Other questions concerned the extent to which their school district had a staff development program. Using categories established by the National Staff Development Council, teachers were asked to identify the stage of their district's program. Twenty one percent answered that their district was at the "initialization" phase, 66.4 percent responded that they were in the

"implementation" stage, while 9.3 percent felt that their school district was currently "institutionalizing" their staff development program. Similar statistical analyses were conducted to determine the relationship between the presence of a school directed staff development program and professional development activities. Teachers who felt that their school district was either in the "implementation" or institutionalization" phase were significantly more likely to have participated in eight of the twelve personal development activities during the past school year, as shown in Table 52.

The professional development activities that were most highly related were those that occurred within the framework of the district's level of development. More simply, those school districts that were further along in implementing a staff development program were more likely to have their teachers participate in such activities as teacher-to-teacher mentoring programs, joint planning, collaborative teaching, collaborative evaluation of student progress, and school-wide instructional initiatives or themes. This finding indicates that individual professional development plans and school district staff development programs may differ in the impact that they have on individual actions aimed at professional development.

Individual, formal professional development plans seem to be related to those activities that are most directly controlled by the individual independent of the school district's commitment to professional staff development. For example, a teacher who has a formal professional development plan in place may be more likely to conduct research related to his or her teaching in spite of the fact that the school district may not have a well-developed staff development program. Additionally, the commitment of the school district toward staff development relates most directly to those activities that occur within the boundaries of everyday school life. In other

words, staff development programs seem to correlate most highly with those activities that are directly under the district's control.

Lighted Schoolhouse

Data regarding implementation of lighted schoolhouse activities were obtained from the Principal Interviews and year end reports required by DPI. In addition, data regarding lighted schoolhouse activities existing prior to SAGE were obtained from the Baseline Data Questionnaire administered in May, 1996, and the school contracts completed for DPI prior to enrollment in the SAGE program.

Most schools have continued the activities they offered in previous school years. Principal Interview data suggest that SAGE schools have taken responsibility for the conception and operation of the lighted schoolhouse activities (as opposed to activities initiated by parents or community volunteers). However, they have not tended to focus heavily on their lighted schoolhouse activities in the first year of SAGE implementation, Some SAGE schools have experienced financial and transportation difficulties as a result of their lighted schoolhouse activities.

Regarding the differences between the pre-SAGE and SAGE years, Table 53 shows that of the thirty SAGE schools, 20 reported an increase in participation in lighted schoolhouse activities, while 3 reported a decrease in participation. Change related to the remaining 7 schools could not be determined due to incomplete or inaccurate information.

References

Bryk, Anthony S. and Stephen W. Raudenbusch. 1992. *Hierarchical Linear Models* (Newbury Park, CA: Sage).

Finn, Jermy D. and Charles M. Achilles. 1990. "Answers and Questions About Class Size: A Statewide Experiment," *American Educational Research Journal*, Fall, Vol. 27 (3): 557-577.

Linn, Robert L. 1983. "Pearson Selection Formulas: Implications for Studies of Predictive Bias and Estimates of Educational Effects in Selected Samples," *Journal of Educational Measurement*, Spring, Vol. 20 (1): 1-15.