

The Enduring Effects of Small Classes

JEREMY D. FINN AND SUSAN B. GERBER

State University of New York at Buffalo

CHARLES M. ACHILLES

Eastern Michigan University

JAYNE BOYD-ZAHARIAS

HEROS, Inc.

The purpose of this investigation was to extend our knowledge of the effects of small classes in the primary grades on pupils' academic achievement. Three questions were addressed that have not been answered in previous research: (1) How large are the effects of small classes relative to the number of years students participate in those classes? (2) How much does any participation in small classes in K-3 affect performance in later grades when all classes are full-size? (3) How much does the duration of participation in small classes in K-3 affect the magnitude of the benefits in later grades (4, 6, and 8)? Rationales for expecting the continuing impacts of small classes were derived in the context of other educational interventions (for example, Head Start, Perry Preschool Project). The questions were answered using data from Tennessee's Project STAR, a statewide controlled experiment in which pupils were assigned at random to small classes, full-size classes, or classes with a full-time teaching assistant. Hierarchical linear models (HLMs) were employed because of the multilevel nature of the data; the magnitude of the small-class effect was expressed on several scales including "months of schooling." The results for question (1) indicate that both the year in which a student first enters a small class and the number of years (s)he participates in a small class are important mediators of the benefits gained. The results for questions (2) and (3) indicate that starting early and continuing in small classes for at least three years are necessary to assure long-term carryover effects. Few immediate effects of participation in a class with a full-time teacher aide, and no long-term benefits, were found. The results are discussed in terms of implications for class-size reduction initiatives and further research questions.

The issue of class size is near and dear to the heart of teachers, parents, and educational policy makers. None would argue that larger classes are better, and most would assert the obvious advantages of small classes over large.

Parents with adequate resources pay for their children to attend private schools, at least in part to reap the benefits of the small-class environment.

Recent decades have produced well over 100 empirical studies of class size. Because the studies used nonexperimental procedures, and because many involved small samples or were of short duration, few definitive conclusions could be drawn. Tentative conclusions were summarized in several widely read reviews, specifically the Glass-Smith meta-analysis (1978), and reviews by Educational Research Service (Robinson and Wittebols, 1986; Robinson, 1990) and Slavin (1989). The reviews converged on four major propositions: (1) "Reduced class size can be expected to produce increased academic achievement" (Glass and Smith, 1978, p. 4), although the effects of even substantial reductions are small (Slavin, 1989). (2) "The major benefits from reduced class size are obtained as the size is reduced below 20 pupils" (Glass and Smith, 1978, p. v). (3) Small classes are most beneficial in reading and mathematics in the early primary grades (Robinson, 1990). (4) "The research rather consistently finds that students who are economically disadvantaged or from some ethnic minorities perform better academically in smaller classes" (Robinson, 1990, p. 85).

In 1985, the Tennessee legislature funded an experiment—Project STAR (Student-Teacher Achievement Ratio)—to provide more definitive answers.¹ Project STAR was a controlled scientific experiment that built on the principles identified in prior research. Students entering kindergarten were assigned at random to a small class (S, 13–17 students), a regular class (R, 22–26 students), or a regular class with a full-time teacher aide (RA) within each participating school. The within-school randomization controlled for a host of between-school differences, including differences in the populations served, differences in per-pupil expenditures and instructional resources, and differences in the composition of the school staff. Teachers were assigned to the classrooms at random. The class arrangement was maintained throughout the day, all year long. There was no intervention other than class size and teacher aides.

Children were kept in the same experimental conditions (S, R, or RA) for up to four years, through grade 3. A new teacher was assigned to the class each year. Over 6,000 students in 329 classrooms in 79 schools in 46 districts participated in STAR in the first year, and almost 12,000 students participated in the course of the four-year study. An array of outcome measures was administered at the end of each school year, including both norm-referenced and criterion-referenced achievement tests in reading, mathematics, and other school subjects.

All pupils returned to full-size classes in grade 4. Fortunately, through funding for the Lasting Benefits Study (LBS) by the Tennessee State Department of Education and recent work by HEROS, Inc. (Pate-Bain et al., 1997), STAR participants have been followed through their high school years and

beyond. In addition to annual tests of academic achievement, student behavior was assessed in grades 4 and 8, attitudes toward school were assessed in grade 8, and school experiences were recorded throughout the grades.

Project STAR has already made major contributions to research on educational processes. It provided the most definitive answers to date about the effects of attending a small class in the primary grades. Many districts and states, the U.S. Department of Education, and several other countries have used STAR findings to guide class-size reduction initiatives. Further, STAR demonstrated that high-quality experimentation in education is feasible. It motivated decision makers to investigate ways to strengthen educational interventions by conducting further experiments on policy issues.² Finally, STAR produced a remarkable database now being used by research teams to answer further questions related to class size, questions about child development generally, and a host of other questions to guide educational policy.

The present investigation addressed three primary questions—questions not answered in previous work. First, we reexamined academic achievement for grades K through 3 to determine the extent to which the effects of small classes or teacher aides are related to the *number of years* that students participate in these settings. At the same time, we examined the impact of small classes on in-grade retentions during these years. The remaining questions pertained to long-term carryover effects of small classes: How much does any participation in small classes in K–3 affect performance in later grades, when children return to full-size classes? And how much does the *duration* of participation in small classes in K–3 determine the magnitude of continued benefits (in grades 4, 6, and 8)? Because samples and methods, not to mention the state of prior research, differ somewhat for short-term and long-term outcomes, the analyses are described in separate sections of this paper.

PART 1: IMMEDIATE EFFECTS OF SMALL CLASSES (GRADES K–3)

The original analyses of STAR data (Word et al., 1990; Finn and Achilles, 1990) consisted of cross-sectional analyses of achievement at the end of each year of experimentation, using nested ANOVA and MANOVA models. These analyses showed that students in small classes had superior academic achievement to students in regular-size classes in every school subject in every grade (K–3); high statistical significance was found for every test, subtest, and multivariate set of tests, including both the norm-referenced and criterion-referenced batteries.³ Further, in each grade, there were some significant interactions with urbanicity and/or race: Minority students and students attending inner-city schools reaped the greatest benefits of attend-

ing small classes. No differences were found between regular-size classes and classes with a full-time teacher aide.

Several factors suggest that the effects of small classes originally reported were understated. Because of student mobility and in-grade retentions, the STAR sample became successively more complex in grades 1, 2, and 3 (see STAR sample below). To the extent permitted by STAR data, both of these issues were addressed in the present study. Despite the complexities, all prior analyses of the K–3 results found significant benefits of small classes in all four years.

Concomitant studies of teaching practices and children's behavior in small classes revealed a set of mechanisms that help explain these differences. Students who attended small classes displayed improved learning behavior, increased engagement in school, and decreased disruptive or withdrawn behavior compared to their counterparts in regular-size classes (Finn, 1998; Finn, Fulton, Zaharias, and Nye, 1989). Teachers in STAR small classes spent increased time in direct instruction, and less time on managerial/organizational tasks (Evertson and Folger, 1989), a finding replicated in the year-long observational study, *Success Starts Small* (Achilles, Kiser-Kling, Owen, and Aust, 1994).

The present investigation reexamined the cross-sectional data, adding three features. First, we partitioned the sample more finely to examine the number of years a pupil participated in a small, regular, or teacher-aide class; this analysis also allowed us to control for student mobility during the experimental years. Second, we used hierarchical linear modeling (HLM) procedures to determine if the original findings replicate using an alternative statistical approach. Third, we focused on the question "How large is the effect?" by estimating strength-of-effect measures in terms of scale scores and in terms of months of schooling, a scale familiar to most educators.

METHODS

The STAR Sample

The selection of schools resulted in over 6,300 kindergarten students in 325 classrooms in 79 schools in 46 districts participating in the first year of the study (see Word et al., 1990; Finn and Achilles, 1990). Students and teachers were assigned at random to a small class (S), regular-size class (R), or regular-size class with a full-time teacher aide (RA) within each school. Larger schools had more than one of each class type so that all kindergarten pupils were allocated. With few exceptions, students were kept in the same class grouping throughout the years they participated in the experiment (for up to four years).

Kindergarten was not mandatory in Tennessee in 1985, so the number of students in the grade-1 sample was larger. Most (at least 90%) had attended some form of kindergarten, but not necessarily in a STAR school. Students entering STAR schools in grade 1, and others in grades 2 and 3, were also assigned to the three class types at random. Sample sizes and the composition of the original STAR sample are given in Table 1. Of the students identified as minority, 98.7% were African American.

Several factors caused the composition of the classes to become more complex in each successive grade. At the end of kindergarten, approximately one-half of regular-class students were assigned at random to teacher-aide classrooms and approximately one-half of teacher-aide students were assigned to regular classrooms.⁴ No students were purposely moved into or out of small classes, and no further reassignments were made after this point.

Second, migration of students into and out of STAR schools—a fact of life in both regular and experimental school programs—added to the complexity. Most grade-1 students had attended kindergarten but some had not. A small number of students, by changing schools, moved into a STAR classroom of a different type. The mixture of students in some classrooms became more complex in grades 2 and 3, with some students having attended the same type of class for zero, one, or two previous years. Krueger (1999) performed a careful analysis of student migration during the four-year experiment and concluded that it did not bias the class-size effect. Approximately 55% of kindergarten and grade-1 entrants participated in the experiment for three or more consecutive years.

Finally, in-grade retentions contributed to the complexity of the sample. Students retained in grade during the study were lost from STAR for sub-

Table 1. Sample sizes for experimental years (K–3)

	Grade			
	K	1	2	3
Number of schools	79	76	75	75
Number of classes	325	339	340	338
Small	127	124	133	140
Regular	99	115	100	90
Aide	99	100	107	108
Median class size				
Small	15	15	15	16
Regular	22	22	23	24
Aide	23	23	23	24
Number of students	6325	6829	6840	6802
Percent minority	32.8	33.1	34.6	33.4
Percent receiving free lunch	48.3	50.2	48.8	48.4

sequent years. While the number of retained students was not very great in total, fewer were retained in small classes compared to regular classes (Harvey, 1993). Interviews with teachers indicated that teachers in small classes had more confidence that students could be passed to the next grade and would receive the additional academic support they required. This left fewer spaces in small classes to accommodate retainees from the previous year. Project STAR did not collect the data needed to estimate the impact of student retentions fully; in the present study, we used the student's age to create a proxy for retention.

Sample sizes for our analyses, after eliminating students missing critical information, ranged from 5,394 to 5,910 for particular test batteries in particular grades.

Measures

The reading and mathematics scales of the Stanford Achievement Tests (SATs; Psychological Corporation, 1983) were administered in the spring of each year. This study used the scale scores created for SAT through item response theory (IRT) that can be compared across grades. For strength-of-effect measures, we also used the grade-equivalent (GE) scale. A grade equivalent of 2.4, for example, means that the pupil is performing like a typical student in the 4th month (December) of grade 2. Thus, if the student is actually in the 4th month of grade 1, he/she is performing quite well. If the student who took the test is actually in the 10th month (June) of grade 2, he/she is performing like students who have had 6 months less schooling. Grade equivalents can be obtained directly from the tables given in the publisher's manuals, or by fitting a curve of mean or median scale scores to the year and month of schooling in which the test was taken (see, for example, Shulz and Nicewander, 1997). The use of GEs has been subject to some debate, focused largely on the interpretation of individual students' scores (see Burket, 1984; Hoover, 1984; Peterson, Kolen, and Hoover, 1989; Yen, 1986). It is clear that GEs are not appropriate for estimating "rate of growth" because the scale is tied to the month/year metric; average growth for a cohort is always about 1.0 GE per school year. However, they are a useful way to compare the means of several groups at a particular grade level, and can be interpreted in terms familiar to educators—months of schooling.

Beginning in first grade, the Basic Skills First (BSF) tests, a set of curriculum-referenced tests developed by the State of Tennessee, were also administered to each student. These were constructed from well-specified lists of objectives in reading and mathematics at each grade level. A student was considered to have mastered an objective if (s)he answered 75% of the items correctly. The numbers of objectives in the tests were as follows—

grade 1: 8 reading and 11 mathematics; grade 2: 12 reading and 15 mathematics; grade 3: 10 reading and 15 mathematics. The present study used the number of objectives mastered in each subject in analyzing the BSF results.

Each pupil was coded according to the number of years spent in a particular class type ("duration"). For example, in grade 1, students in small, regular, or teacher-aide classes were coded to indicate whether this was their first or second year in that class type. In grade 2, duration was coded one, two, or three. Students with particular combined experiences were eliminated from the analysis. This includes students who, for whatever reason, had participated in both a small and teacher-aide classroom and students who had been in a small or teacher-aide classroom in previous years but were now in a regular class. As a result, the numbers of students omitted from the grade 1, 2, and 3 cross-sectional analyses were 230, 477, and 662, respectively.⁵

In each grade, the student's age was used to indicate the likelihood that (s)he had been retained in a previous grade. Students who were 14 months or more above the grade-appropriate age were classified as "over age." For example, Tennessee students were admitted to first grade if they reached their sixth birthday (72 months) by September 1. Any first grade student 86 months or older on that date was placed in the over-age category. The limited retention data gathered in STAR indicated that most over-age students had been retained in a previous grade, a finding confirmed by Harvey (1993) from a sample of actual student records. The student's participation in the free lunch program was used as an indicator of socioeconomic status (SES).

Analyses

A three-level HLM analysis was performed for each of the five achievement tests separately at each grade level. Level-1 (student) variables were gender (male–female), race (white–minority), the numerical duration index, the interaction of race with duration, and the test score. Level-2 (classroom) variables included two class-type contrasts: small classes compared to regular classes and teacher-aide classes compared to regular classes. Level-3 variables were two dummy codes to compare three school locations: suburban and rural schools compared to inner-city schools, respectively. The interactions of class type with race, duration, and urbanicity were included in the model as well as the interactions of race with duration, urbanicity with duration, and the three-way race-by-duration-by-class type and urbanicity-by-duration-by-class type interactions.⁶ All tests were conducted at the .01 significance level. When the three-way interactions and some two-way interactions were found to be nonsignificant, the models were reestimated retaining only the significant effects.

Several follow-up analyses were performed. First, SES was added to the final model and all analyses were rerun to see if there were any noteworthy changes in the results; the SES contrast was coded as “high–low” (full-price lunch–free lunch). Second, specific contrasts were tested in the grade 1, 2, and 3 data not represented by individual regression coefficients. In grade 1, we compared the performance of students who had been in small classes (and teacher aide classes) for one year or two consecutive years with the performance of grade-1 regular-class students. In grade 2, we compared the performance of students who had been in small classes (and teacher-aide classes) for one year, two years, or three years, respectively, with the performance of students who had been in regular classes; and so on.⁷ One, two, and three years in a small class or teacher-aide class were viewed as different treatment conditions to be compared to the “norm” (students in regular classes). As a final set of analyses, the over-age variable was added to the reduced model to see if achievement was affected above and beyond the effects of different retention rates.

Several aspects of the analysis addressed the issue of student mobility. The duration variable, obtained for both regular and experimental students, provided some statistical control because students with shorter durations in STAR were likely to have transferred from another school. Further, the analysis included three indicators linked closely with mobility: student race, SES, and school urbanicity.

Effect-size measures were derived from the final models as follows. Predicted means (“fitted means”) for small and regular classes were obtained from the reduced HLM model. For both SAT and BSF tests, the mean difference was divided by the standard deviation of all students in regular classes in that grade. Next, the predicted means on the SAT tests were converted to grade equivalents (GEs) using a curve derived from the tables published by the Psychological Corporation. These were subtracted to estimate the difference in months of schooling.

IMMEDIATE EFFECTS: RESULTS

The final HLM regression coefficients for grades K–3, including the control for SES, are given in Appendix Table A-1. Differences on background characteristics were consistent with prior research on the performance of elementary-grade pupils. White students performed significantly better than minorities on all achievement tests in all grades, and higher-SES students performed better than lower-SES students. In fact, the SES effect was substantially greater than the racial difference in most instances. Girls performed significantly better than boys on all verbal measures (Total Reading, Word Study Skills, BSF Reading) in each grade. Although girls performed better than did boys in mathematics in kindergarten, the differences were

not significant in grade 1 and were significant only on the BSF Mathematics test in grades 2 and 3. All comparisons of school locations were statistically significant. Students in suburban and rural settings performed better than did students in inner-city schools on every test in each grade.

Overall Differences among Class Types

Overall differences among class types (labeled “Small” in Table A-1) were consistent. In every grade (K–3), students attending small classes performed better academically—on all achievement tests—compared to students in regular-size classes.⁸ In grade 3, when tests were administered in other subject areas, the same results were found on the Language, Science, and Social Science tests as well. No overall differences were found between teacher-aide classes and regular classes (labeled “Aide” in Table A-1) at the .01 level of significance.

The HLM analyses also revealed several significant or marginally significant interactions with class type. Interactions in grade 1 and 2 involved the duration variable discussed below. In grade 3, the urbanicity-by-small interaction was significant for Word Study Skills. The means⁹ indicate that this occurred because inner-city students benefitted more from small classes than did students in other locations, achieving a small-class advantage nearly twice as large as that of students in suburban and rural schools.

How large were the benefits of small classes? To answer this question, effect sizes were estimated for the small class–regular class difference (Table 2). The magnitude of the overall small class advantage was about 0.2σ on each test in kindergarten, increasing to about 0.3σ in grade 1 and then declining slightly in grades 2 and 3. The decreased effect sizes in grades 2 and 3 reflect in part that these samples include pupils who attended small classes for different numbers of years—from 1 to 3 years in grade 2 and from 1 to 4 years in grade 3.

The grade-equivalent results for the SAT tests estimate the benefits of attending a small class in “months of schooling.” At the end of kindergarten, students in small classes were about one-half month ahead of their peers in reading and about 1.6 months ahead in mathematics. Each of these increased in grade 1. Grade-1 students who attended small classes were about 1.3 months ahead of regular-class students in reading and about 2.8 months ahead in mathematics. In grade 2, small-class students were between 3.5 and 4.8 months ahead of students in regular classes. In grade 3, the advantage was still greater in verbal areas but appeared to drop somewhat in mathematics.¹⁰ In general, the advantage of small classes was both statistically significant in every grade and, in months of schooling, increased from year to year.

Table 2. Effect sizes for small classes

Grade	SAT tests ^a			BSF tests ^a		SAT grade equivalents ^b		
	Total Reading	Word Study Skills	Total Mathematics	Reading	Mathematics	Total Reading	Word Study Skills	Total Mathematics
Kindergarten	0.21	0.20	0.19	—	—	0.5	0.5	1.6
Grade 1	0.30	0.29	0.31	0.31	0.21	1.3	0.9	2.8
Grade 2	0.26	0.25	0.25	0.21	0.18	3.9	4.8	3.5
Grade 3	0.22	0.21	0.15	0.14	0.11	4.5	5.5	2.6

^aDifference between mean performance of small-class students and regular-class students, divided by standard deviation of all regular-class students.

^bDifference between small-class students and regular-class students in months of schooling.

Differences According to the Number of Years in Small Classes

Students in grades 1, 2, and 3 were classified according to the number of years they had attended a particular class type. The small class–regular class comparisons are displayed in Figure 1; detailed HLM results for the analysis of number of years are given in Table A-1. In Figure 1, the vertical axis corresponds to the effect size, that is, the number of standard deviations separating the mean performance of students in small classes and the mean of students in full-size classes. The dots are placed at a height corresponding to this axis. For the SAT tests, each dot corresponding to a statistically significant difference is also labeled with the effect size in terms of months of schooling (GE).

The upper-left portion of Figure 1 shows results for the Reading scale of the Stanford Achievement Tests. In kindergarten, the advantage of attending a small class was about 0.21σ or one-half month on the grade equivalent scale; this difference is significant at $p < .001$. In grade 1, students attending small classes for the first time performed better in reading than students in regular classes. The effect size was 0.16σ , corresponding to an advantage of about 6/10 of a month; the difference was significant at $p < .01$. Grade-1 students who had attended small classes for two years (since kindergarten) had a greater advantage over regular-class students. The effect size was 0.40σ , corresponding to an advantage of about 1.9 months of schooling; this difference was significant at $p < .001$.

In grade 2, the reading effect sizes for students who attended small classes for one, two, and three years were 0.12σ , 0.24σ , and 0.36σ , or advantages of approximately 1.8 months, 3.7 months, and 5.7 months, respectively. These differences were significant at the .05 and .001 levels. In grade 3, the advantage was not statistically significant for students who attended small classes for the first time, but was statistically significant for those who had attended for one or more previous years. The advantages of attending a small class for two years (since grade 2), for three years (since grade 1), and for four years (since kindergarten) were approximately 3.3 months, 5.2 months, and 7.1 months, respectively.

The same pattern was obtained for every achievement test (including Word Study Skills; see Table A-1): In each grade, there was a clear benefit of attending small classes for additional years. Grade-1 students who were in small classes for the first time scored between 0.6 and 2.0 months ahead of regular-class students on the SAT tests; those who attended small classes for two years scored between 1.9 and 3.4 months ahead of regular-class students. In grade 2, the advantage of having spent three years in a small class was substantially greater than two years, which was, in turn, greater than the advantage of being in a small class for the first time. In grade 3, the advantage of having been in a small class for four years reached 7.1

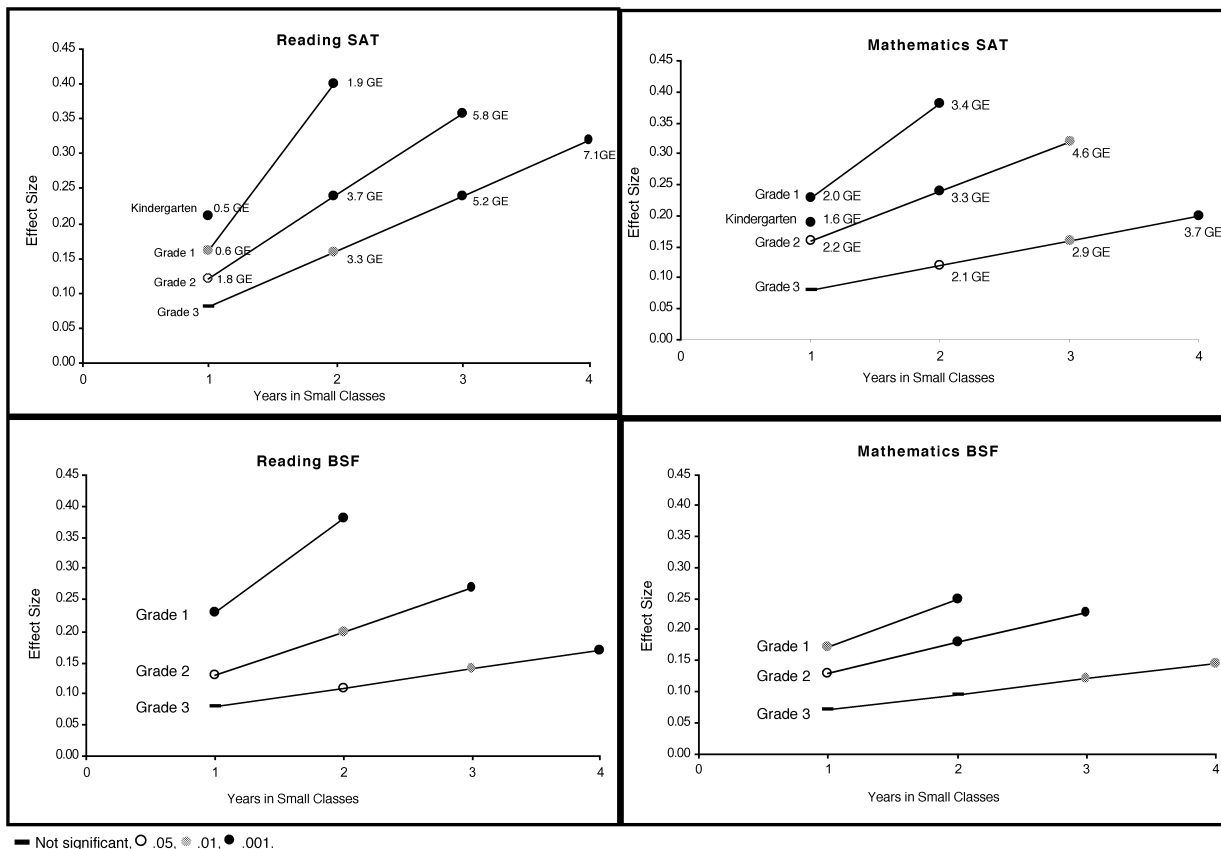


Figure 1: Small-class effects by number of years of participation (duration)

months and 8.2 months in the verbal areas and 3.7 months in mathematics. Effect sizes for the BSF tests also increased monotonically as students spent additional years in a small class.¹¹

The effect sizes shown in Figure 1 and the significance levels in Table A-1 demonstrate two principles: The first is the importance of continuous attendance in small classes. It is clear from the figure that the effects increase with each additional year in a small class. In addition, in later grades, the effects were only statistically significant for those students who had continued small-class participation. In grade 2, none of the one-year comparisons was statistically significant at $p < .01$ (although several were “marginally” significant). In contrast, students who attended small classes for two years (grades 1 and 2) or for three years (grades K, 1, and 2) performed significantly better than did students in regular classes. Likewise, grade-3 students attending small classes for the first time performed better, but not significantly so, compared with students in regular classes. Those who attended small classes for two years (grades 2 and 3) performed significantly better than regular-class students on two of the tests (SAT Reading and Word Study Skills). Students who attended small classes for three or four years performed significantly better than regular class students on all achievement tests.

The second principle is the importance of starting small classes in the earliest grades. Note, for example, that one year in a small class was significantly related to students’ performance in kindergarten and grade 1, less significant in grade 2, and not significant in grade 3. Likewise, two years in a small class had significant impacts on performance in grade 1 and grade 2 but produced mixed results in grade 3. Thus, the earlier small classes are introduced, the greater the potential for a strong impact on academic achievement.

Other findings. Several specific interactions with duration were also statistically significant; we examined the interaction means for each. In grade 1, the duration-by-race-by-small interaction occurred because minority students who attended small classes for two years scored substantially closer to white students who attended small classes for two years on the BSF reading test. This reduced achievement gap was not found for any other combination of race, duration, or class type. In grade 2, the duration-by-urbanicity-by-small interaction occurred because inner-city students who attended small classes for three years scored closer to suburban and rural students who attended small classes for three years on the BSF reading test. Again the achievement gap was reduced for students in inner-city schools.

The overall comparison between teacher-aide classes and regular classes was not significant for any test in any grade. However, several specific

contrasts were statistically significant. In grade 1, students who had teacher aides for two years outperformed their regular-class counterparts on the verbal tests (SAT Reading and Word Study Skills; BSF Reading). In grade 2, students who had teacher aides for three years outperformed their regular-class counterparts on the SAT verbal tests. In grade 3, no teacher-aide effects attained the .01 significance level. We view these results as worthy of note and further exploration. It is possible, for example, that aides are most useful in the reading areas during the years when reading is emphasized (particularly in grade 1).

In-Grade Retentions

Project STAR did not collect extensive information about student retentions; we used the “over-age” variable as a proxy. In kindergarten, 3.2% of students in small classes were over-age in comparison to 2.6% of students in regular and teacher-aide classes. Although the random assignment to the three conditions was carried out and monitored carefully, slightly more retainees or late entrants were assigned to small classes. In grade 1, 7.5% of small-class students were over-age in comparison to 13.1% of regular students and 13.7% of teacher-aide students. These pupils may have been retained either in kindergarten or in grade 1 the previous year. In grade 2, the percentages were 12.4% in small classes compared to 19.2% and 20.1% in regular and teacher-aide classes. In grade 3, the percentages were 16.9% in small classes compared to 21.1% and 21.5% in regular and teacher-aide classes.

Because fewer students were retained in small classes than in regular or aide classes, there was less space for retained students and fewer retainees were assigned to small classes each year. Differential retentions are both an outcome of small class participation and a factor that may confound the analysis of academic achievement. Thus, we repeated the final analyses of achievement (including SES) with over-age as an additional control variable.

The difference in performance between over-age and non-over-age students was highly significant ($p < .001$) on every test in every grade, with over-age students having substantially poorer achievement scores. The tests of significance among class types, however, showed no consistent or noteworthy change from the uncontrolled analysis. Some p -values were unchanged—for example, all tests of significance for small classes in grades K and 1. Others increased—for example, several tests of particular contrasts in grades 2 and 3. And others decreased slightly—for example, the three- and four-year comparisons for the two mathematics tests in grade 3.

Likewise, the teacher aide findings were not changed in the reanalysis, although several contrasts showed less statistical significance. In grade 1, the two-year contrast for BSF Reading did not reach the .01 significance

level. In grade 2, all of the three-year contrasts for SAT tests became non-significant; and in grade 3, the largest four-year contrast (SAT Reading) was reduced substantially.

In general, when we asked whether small classes had an impact on academic achievement above and beyond the effects of in-grade retention, the answer was “yes.” Controlling for retentions, a number of comparisons reached higher levels of significance. In addition to confirming the pattern of overall effectiveness of small classes, the analysis suggested that even one year in a small class in grade 2 or 3 is worthwhile for many students. On the other hand, the apparent effectiveness of teacher aides for students in aide classes for three or four years may be a function of differential retentions rather than a true benefit to pupil performance.

CONCLUSIONS

The reexamination of data from Tennessee’s Project STAR for grades K to 3 substantiate three conclusions. First, on average, students in small classes perform better than do students in regular classes or regular classes with teacher aides in each grade on all tests of academic performance. The “typical” magnitude of the small-class advantage was approximately 0.2σ on every test for the total sample of pupils in each grade. This finding parallels earlier work using other statistical approaches (for example, Finn and Achilles, 1990), although we did not find as many significant interactions with race/ethnicity or urbanicity as found previously. Those interactions that were significant or approached statistical significance were all in the direction of greater small-class advantages for students who usually have lower performance, that is, minorities or students attending inner-city schools.

Second, both the year in which a student first enters a small class and the number of years (s)he participates in a small class are important mediators of the benefits gained. For students who began small classes in kindergarten or grade 1, there was a significant advantage in both reading and mathematics. For those who attended small classes in *both* kindergarten and grade 1, the small class advantage was greater still, putting these students from approximately 1.3 months to 3.4 months ahead of their counterparts who attended full-size classes.

In grades 2 and 3, the benefits of attending a small class were not statistically significant unless students had attended for at least two years. Again, in each grade, the benefits of additional years in a small class were greater. In grade 2, for example, the achievement advantage for students who had participated in small classes for one, two, and three years was 0.12σ , 0.24σ , and 0.36σ , respectively, in reading and 0.16σ , 0.24σ , and 0.32σ , respectively, in mathematics. In terms of months of schooling, these advantages placed the students from 1.8 months, to 3.7 months, to 5.8

months ahead of their peers in reading, and from 2.2 months, to 3.3 months, to 4.6 months ahead in mathematics. The results were not diminished when statistical control was introduced for the number of over-age students enrolled in each class.

Third, in general, we found few if any academic benefits associated with a full-time teacher aide. Several statistically significant differences were found for grade-1 students who had attended teacher-aide classes for two years, and several were found for grade-2 students who had attended for three years. The significance levels were reduced, in some cases to nonsignificance, when statistical control for student retentions was added to the analysis. More research on the specific duties of teacher aides and how they may interact with retained or low-performing students may be useful.

PART 2: ENDURING EFFECTS OF SMALL CLASSES (GRADES 4–8)

Project STAR ended after grade 3, and all students returned to regular classes in grade 4. Achievement and behavior data continued to be collected, allowing us to ask how strong the benefits of small classes would be after pupils return to full-size classes. We began our study of carryover effects with a question of theory: Under what conditions would we *expect* the benefits to endure into later grades?

Other early-childhood programs have had mixed results. It is not uncommon that immediate cognitive benefits, reflected in tests of academic achievement, tend to decrease over time (White, 1985–86). There is some evidence of lasting benefits for non-achievement outcomes, however (Barnett, 1995; Lazar and Darlington, 1982). Both the Perry Preschool Project (Berrueta-Clement et al., 1984) and most Head Start programs exhibit this pattern.¹² In both of these programs, experimental and comparison-group students were indistinguishable on tests of academic achievement three years after students left the program. However, there were indications that program students were less likely to be placed in special education, less likely to be retained in grade, and more likely to graduate than their comparison-group counterparts (Berrueta-Clement et al., 1984; McKey et al., 1985).

The Chicago Parent Child Centers (CPC) program documented continuing effects on academic achievement (Reynolds, 1997). Designed to aid low-income students, especially those not served by Head Start, the programs have components for preschool, kindergarten, and grades 1 through 3. Thus, children can participate for as many as six years. Preschool programs meet for one-half day, but other portions are full-day programs. Class activities, focusing on the development of language and reading skills, take place in small classes. Each classroom also has a full-time teacher aide, and

parental involvement is extensive. An evaluation of continuing effects showed that CPC students outperformed non-program students in reading and mathematics in grades 3 and 5, and in mathematics in grade 8; effect sizes ranged from $.17\sigma$ to $.34\sigma$ (Reynolds, 1997). Like those in the Perry Preschool Project and Head Start, students attending the CPC were less likely to be retained in grade.

What features of educational programs are likely to foster long-term benefits? Summarizing the evaluations of 36 early childhood programs, Barnett (1995) concluded that to have any long-term effects at all, school-age services “must actually change the learning environment in some significant ways” (p. 43). Ramey and Ramey (1998) forwarded a conceptual framework for early interventions and identified six principles of efficacy. The principles rest on the concept that “fragmented, weak efforts in early intervention are not likely to succeed, whereas intensive, high-quality, ecologically pervasive interventions can and do” (p. 109). The primary principles are: (1) developmental timing, that is, start early and continue; (2) program intensity, that is, the importance of many hours per day, days per week, and weeks per year of the intervention; and (3) direct provision of learning experiences, rather than relying on intermediary sources; parent training alone, for example, is not likely to have an enduring impact on children’s learning.

The principle of developmental timing is consistent with all that is known about the importance of intervening when the rate of growth is greatest, and the increasing difficulty of effecting change as children grow older (for example, Bloom, 1964). Studies of program duration support the principle of program intensity. In one, Head Start–only students were compared with others who remained for an additional four years in the continuation, Project Follow Through (Abelson, Zigler, and DeBlais, 1974). Students in the extended program outperformed their counterparts on several aspects of the Peabody achievement tests, leading the authors to conclude that “gains accruing from compensatory education programs are commensurate with duration” (p. 770). An evaluation of CPC programs (Reynolds, 1997) examined both starting age and duration. Students who participated in the CPC for one or two years were indistinguishable from the control group on grade-8 achievement tests, while those who participated for five or six years were superior in both mathematics and reading. The advantage of duration remained significant controlling for age of entry, while age of entry was not significant controlling for duration. Duration in the intense CPC program was more important than age of entry.

The Perry Preschool Project and most Head Start programs are of limited intensity (see Zigler and Styfco, 1994). Although beginning at an early age (Perry at age 3, Head Start at 4), the Perry program lasted for two years,

ending when pupils entered regular schools at age 5. Likewise, most Head Start programs begin when the child is 4 and end when regular school begins in kindergarten. Neither program engages pupils for the full day. The Perry intervention involved about $2\frac{1}{2}$ hours of school time daily; typical Head Start programs involve about $3\frac{1}{2}$ hours of class time, four or five days per week. Both extend the hours of engagement somewhat through home visits, although there is much variability in this practice among Head Start sites (Zigler and Styfco, 1994).

When students leave programs such as Perry, CPC, or Head Start, they are likely to enter half-day kindergartens with programs targeted to average pupils. The advantages of full-day kindergartens—especially those with a heavy academic emphasis—are well documented (Achilles, Nye, and Bain, 1994–1995; Cryan *et al.*, 1992; Finn, n.d.; Karweit, 1989; Naron, 1981). Yet by 1997, only 10 states¹³ required all districts to offer full-day programs (Council of Chief State School Officers, 1998). The remaining states required half-day programs (21), required half-day or full-day programs at the discretion of individual districts (10), or had *no* statewide policy about the provision of kindergarten (10).

Although directed to all students—not just children of poverty—Tennessee’s Project STAR also started early. Beginning with full-day kindergartens, STAR was a high-intensity intervention, affecting children for the entire school day every day of the school year, for up to four consecutive years. STAR impacted the learning setting directly and thus influenced all student-teacher interactions taking place in that setting. We would expect that (1) there would be positive impacts of small classes on academic achievement in the later grades, and (2) the long-term impact would be related to the duration of a student’s participation in small classes in the early grades.

METHODS

Samples and Measures for Enduring Effects

Beginning in the fall of 1989, when STAR pupils were in grade 4, Tennessee initiated the Tennessee Comprehensive Assessment Program (TCAP). Every pupil was tested on the TCAP battery, which included norm-referenced tests from the Comprehensive Tests of Basic Skills (CTBS/McGraw Hill, 1989) and BSF tests for each grade in reading and mathematics. In the first year of testing, some districts declined to participate, resulting in a smaller follow-up sample for grade 4. However, most or all districts participated in the TCAP program by the time STAR pupils reached grades 6 and 8.

The present study used the IRT scale scores from the CTBS Total Mathematics, Total Reading, Science, and Social Science scales, and the number

of objectives mastered on the BSF reading and mathematics tests as outcome measures. The total numbers of objectives on the test were 8 and 8 in grade 4, 7 and 9 in grade 6, and 10 and 8 in grade 8 for reading and mathematics, respectively.

The full sample for the study of enduring effects included students who had attended STAR regular classes or else small or teacher-aide classes for one or more years. A few students who attended both a small and teacher-aide class at some point (for example, by changing schools) were eliminated. The grade-4 sample comprised 4,015 students in 61 schools providing data on the CTBS and 4,045 students providing data on the BSF tests (see Table 3). While these schools had participated in STAR the previous year, only 40% of inner-city schools provided scores for the grade-4 analysis. The grade-6 and grade-8 samples were larger and more dispersed as students entered middle- and high-school grades. The original proportions of inner-city, suburban, and rural schools were reestablished. The grade-6 sample comprised 6,100 students in 518 schools who provided CTBS results and 2,737 students in 213 schools who provided BSF scores.¹⁴ The grade-8 sample comprised 5,835 students in 489 schools who provided CTBS results and 5,217 students with BSF scores. Although the number of schools is large, most pupils were concentrated in a subset of them. For example, in

Table 3. Sample sizes for post-experimental years

	Grade		
	4	6 ^a	8
Number of schools	61	518	489
Number of students (CTBS tests/BSF tests)	4015/4045	6100/2737	5835/5217
Percent minority	19.8	32.1	32.4
Percent receiving free lunch	39.3	47.1	47.0
Number of schools in duration sample	61	464	405
Number of students in duration sample (CTBS tests)			
Small classes, by duration			
1 year	72	277	263
2 years	66	174	182
3 years	181	250	223
4 years	637	693	657
Regular classes	506	923	899
Teacher aide classes	1371	1989	1940

^aBSF data only available for 2737 students in 213 schools in Grade 6 in full sample. For duration analysis, Grade-6 BSF data only available for 2017 students in 158 schools.

sixth grade, 229 schools had five or more STAR pupils; these schools accounted for 90.3% of our grade-6 sample. In eighth grade, 183 schools had five or more STAR pupils, accounting for 90.9% of our sample.

To study duration (Research Question 3), a sub-sample of students was selected who entered Project STAR in kindergarten or grade 1 and who participated in the same class type (small, regular, or teacher aide) for one year or for two, three, or four consecutive years. Kindergarten and grade-1 entrants constituted the original experimental plan and had the opportunity to participate for up to four years. Students who entered STAR classes in grade 2 or 3 and those with other patterns of small-class participation were eliminated from this analysis;¹⁵ the resulting samples were smaller, due primarily to the elimination of students who entered STAR in grade 2 or 3 (see Table 3).¹⁶

Analyses

The full-sample analysis examined overall differences among students who had been in small, regular, or teacher-aide classes in K–3. This consisted of testing a series of HLM models for two levels of data, students and schools. Student-level variables included gender (male–female), race (white–minority), two class-type contrasts (small–regular; aide–regular), and the test score. School-level variables included two contrasts among three school locations: suburban and rural schools compared to inner-city schools, respectively. The interactions of class type with race and with urbanicity were included in the model as well. When interactions were found to be non-significant, the “reduced” main-effects model was rerun and SES (full-price lunch–free lunch) was added to see if any noteworthy changes in the results occurred. Effect sizes were estimated in standard deviation units and in grade equivalents, using the final models in the same manner as in the K–3 analysis.

To study the effects of duration, we classified each pupil according to the grade of entry into STAR (K or 1) and the extent of continuous participation. Students were identified who had one year of participation in a small or teacher-aide class (K or 1), two years in sequence (K–1 or 1–2), three years in sequence (K–1–2 or 1–2–3), or four years. Students who never participated in either experimental condition were classified as regular-class pupils and were also classified by the number of years they had been in Project STAR.

The study of duration was complicated by the fact that students who left STAR before reaching third grade, due to their family moving or to being retained in grade, were not a representative cross-section of all pupils. For example, students who move from school to school are generally from

lower-income families; that, and changing schools itself, may result in reduced academic achievement. In all, 32.8% of students who entered STAR kindergarten classes were minority and 48.3% participated in the free-lunch program. Of those who left STAR classes before reaching third grade, 39.9% were minority and 58.6% were receiving free lunches. Similarly, 39.3% of students who entered STAR classes in grade 1 were minority and 61.2% were receiving free lunches. Of those who left STAR classes before third grade, 43.6% were minority and 63.8% were receiving free lunches. Statistical control for school urbanicity and student race, SES, and duration of participation in STAR helped tighten the internal validity of the conclusions.

Preliminary analyses indicated a confounding of grade of entry with the effects of duration, for several reasons. Students who entered in grade 1 could not participate in the experiment for more than three years. Also, when students were cross-classified by starting grade, duration, and class type, the grade-4 sample had a number of very small cell *ns*. Every analysis, regardless of test, grade, or other variables in our models, indicated that both short-term and long-term effects were mediated by the number of years a student participated in a small class. For these reasons, buttressed by the CPC findings that duration was more important than starting age (Reynolds, 1997), we completed our analyses by focusing on duration of small-class participation.

The primary analyses consisted of testing HLM models with two levels of data, students and schools. The only school-level variable was urbanicity. Student-level variables included the outcome measure plus gender (male–female), race (white–minority), SES (full-price lunch–free lunch), class type (small, regular, or aide), and the numerical duration index (1, 2, 3, or 4). The full model also included interactions with class type and duration (class type-by-duration; class type-by-race; class type-by-urbanicity; duration-by-race; duration-by-urbanicity; duration-by-class type-by-race; duration-by-class type-by-urbanicity). In each grade, contrasts were tested that were not represented by the individual regression coefficients. Specifically, we compared the performance of students who had been in small classes (and teacher-aide classes) for one year, two years, three years, and four years, respectively, with the performance of students who had been in regular classes.¹⁷ Again, one, two, three, and four years in a small class or teacher-aide class were viewed as different treatment conditions to be compared to the “norm” (regular classes). Students with shorter durations in Project STAR may have transferred to other schools and thus comprised a selectively mobile group. The inclusion of race, SES, urbanicity, and duration in Project STAR in the statistical models provided as much control for mobility as possible from our data set.

As in the other analyses, when some or all interactions were found to be nonsignificant, they were eliminated and effect size measures were derived from the reduced models. Predicted means were expressed as effect sizes using the standard deviations of all regular-class students, and these were converted to grade-equivalents using curves derived from the tables published by CTBS/McGraw Hill.

ENDURING EFFECTS: RESULTS

How Much Does Any Participation in Small Classes in K–3 Affect Performance in Later Grades?

The reduced-model HLM coefficients for grades 4, 6, and 8, including the control for SES, are given in Appendix Table A-2. Results for student and school demographics are consistent with the K–3 analysis. Race and SES differences were highly statistically significant on virtually all achievement tests in all three grades. Girls maintained their superiority over boys in reading and mathematics, with very few exceptions, through grade 8, and performed significantly better than did boys in social sciences in grade 8. Boys exhibited superior performance in science in grades 4 and 8.¹⁸ Students in suburban and rural schools performed significantly better than did their inner-city counterparts on every test in all three grades.

Overall differences among class types. Grade-4 students who had attended small classes for one or more years in K–3 scored significantly better on all tests of academic achievement than did students who attended full-size classes; all differences were significant at $p < .01$ or below despite the smaller grade-4 sample size. In grade 6, all differences were significant in favor of small classes except for science. In grade 8, differences favoring small classes were significant at $p < .01$ or below on four of the six achievement tests, and approached significance ($p < .05$) on the other two. No significant effects were found in grade 4, 6, or 8 for students who attended teacher-aide classes.

The small class–regular class differences are given in Table 4. In grade 4, the small-class advantage ranged from 0.11σ to 0.15σ on the CTBS tests and 0.14σ and 0.16σ on the BSF tests. In grade 6, apart from science, the small-class advantages ranged from 0.10σ to 0.20σ , being generally about the same or slightly less than the values for grade 4. In grade 8, the small-class effect sizes ranged from 0.08σ to 0.14σ . In general, effect sizes remained in about the same range from grade 4 to grade 8.

In terms of months of schooling (grade equivalents), the advantage of having attended a small class for one or more years ranged from 2.4 months

Table 4. Overall small-class effect sizes for grades 4, 6, 8

Grade	CTBS tests ^a				BSF tests		CTBS grade equivalents ^b			
	Total Reading	Total Mathematics	Science	Social Science	Reading	Mathematics	Total Reading	Total Mathematics	Science	Social Science
Grade 4	0.15	0.13	0.11	0.12	0.14	0.16	3.9	2.4	3.1	4.6
Grade 6	0.10	0.10	0.02 ^c	0.12	0.20	0.13	3.0	3.1	1.0	5.1
Grade 8	0.11	0.08 ^d	0.09	0.08 ^d	0.14	0.09	4.1	3.4	4.3	4.8

^aDifference between mean performance of small-class students and regular-class students, divided by standard deviation of all regular-class students.

^bDifference between small-class students and regular-class students in months of schooling.

^cNot significant.

^dMarginally significant ($p < .05$).

to 4.6 months in grade 4, from 3.0 to 5.1 months in grade 6 (apart from science), and from 3.4 months to 4.8 months in grade 8. In general, students who attended small classes during the K–3 period were at least 2.5 months ahead in all school subjects, and as much as five months ahead in some, compared to their counterparts who attended regular classes.

Attending a small class in grades K–3 is associated with enduring academic benefits in all school subjects in grades 4, 6, and 8. It is likely that the impact endures beyond grade 8; unfortunately, we did not have achievement test data for higher grades. The question “How large is the effect?” is answered partially from these results. The effect size measures represent *lower limits* on expected outcomes because the sample included students who entered STAR in different grades and participated in small classes for as little as one year and others who participated for two, three, or four years. Despite the diverse nature of the sample, virtually all results were statistically significant and the total lasting impact of small classes, having affected all school subjects for five years or more, was substantial.

How Much Does the Duration of Participation in Small Classes in K–3 Affect Benefits in Later Grades?

The duration sub-sample consisted of those participants who entered Project STAR in kindergarten or grade 1 and who could be classified as having one year of participation or two, three, or four contiguous years. Although the duration sub-sample was about 30% smaller than the full samples in grades 4, 6, and 8, tests of significance for background characteristics (race/ethnicity, SES, gender, and urbanicity) produced results similar to those for the full samples; they are not tabled again. And like the full-sample analysis, the duration analysis produced no teacher aide–regular class contrast that was significant at $p < .01$ in any grade.

Differences according to the number of years in small classes. Figure 2 is a display of the differences in performance between students who attended small classes and those who attended full-size classes; the organization of Figure 2 is the same as that of Figure 1.

The upper-left portion of Figure 2 shows results for the Total Reading scale of the CTBS. In grade 4, the difference between the average performance of students who had been in small classes for one year and students who had been in regular classes was about 0.04σ (not statistically significant). The effect size for students who had been in small classes for two years was approximately 0.12σ (not statistically significant). The effect size for three years in a small class was approximately 0.20σ . This difference corresponds to an advantage of approximately 4.6 months of schooling

(GE) (statistically significant at the .01 level). Four years in a small class produced an advantage of 0.28σ or approximately 6.6 months of schooling (significant at the .001 level).

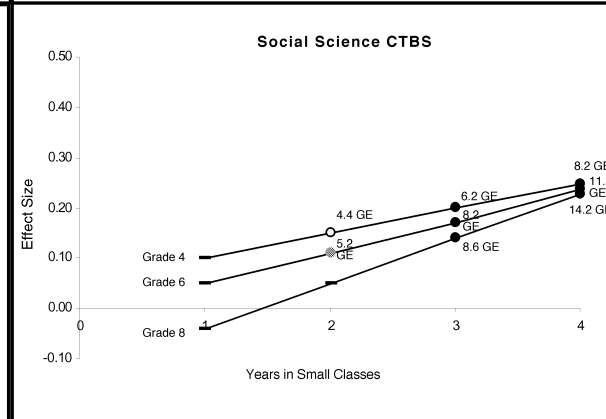
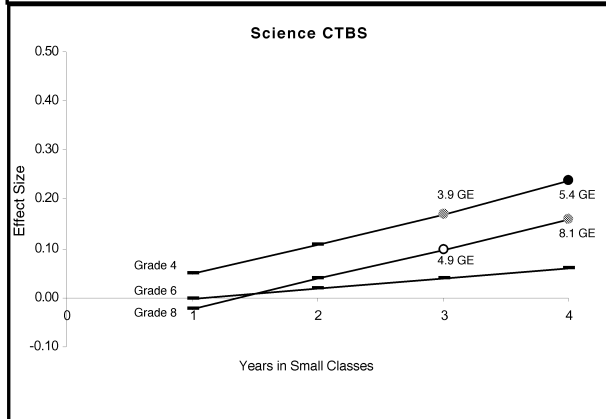
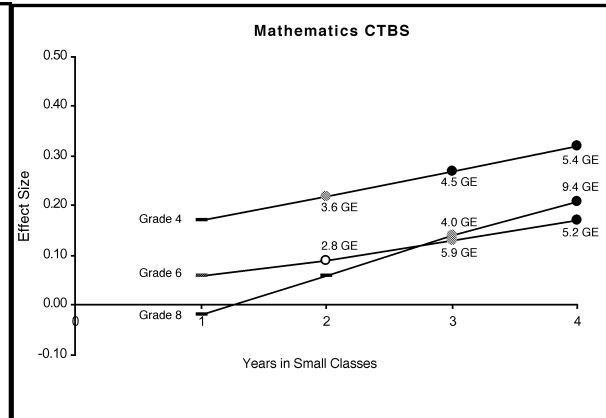
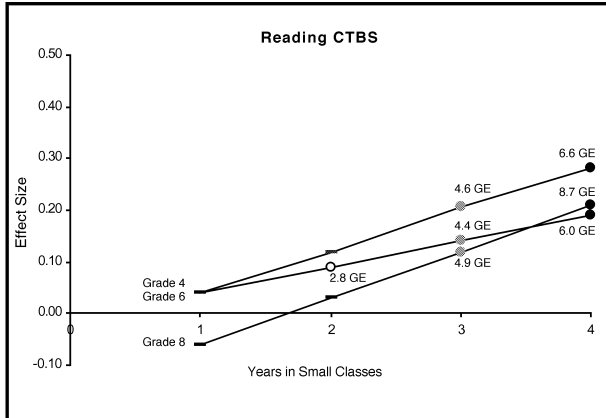
The patterns in grade 6 and grade 8 are similar. Effect sizes for one year in a small class are small and nonsignificant. Effect sizes for two years in a small class are larger but either marginally significant or nonsignificant. Effect sizes for three years in a small class are 0.14σ in grade 6 and 0.12σ in grade 8, corresponding to a 4.4-month advantage and a 4.9-month advantage in reading, respectively; both are statistically significant at the .01 level. The greatest effects were found for four years in a small class, which peaks at approximately a 6.0-month advantage in grade 6 and an 8.7-month advantage in grade 8.

The same pattern was obtained for every CTBS achievement scale and the criterion-referenced BSF tests. Specifically, students who had attended small classes for one year performed no better in grade 4, 6, or 8, on average, than students who had attended regular-size classes. Although there were significant “immediate” benefits of participating in a small class in kindergarten or grade 1 alone, the benefits did not endure through the later grades when all students returned to full-size classes. The one-year effect sizes are small and some are even slightly negative; none is significant at $p < .01$.

Two years in a small class (K–1–2 or 1–2) yielded consistently larger effect sizes than did one year, but only attained statistical significance for some tests in some grades (CTBS Mathematics in grade 4 and CTBS Social Science and BSF Reading in grade 6). Although there may have been some carryover effects of two years in a small class, they were minimal by the time two, four, or six additional years had gone by. Effect sizes for two years in a small class are all larger than for one year in a small class, however; the median effect size is $.10\sigma$ considering grades 4, 6, and 8 together.

Three years in a small class (K–1–2 or 1–2–3) had important carryover effects in later grades. In grade 4, differences favoring small classes were statistically significant on all CTBS and BSF scales. In both grades 6 and 8, differences favored small classes on all tests except CTBS Science and BSF Mathematics. Every effect size for three years in a small class is larger than the respective effect size for two years in a small class. The median effect size is 0.21σ in grade 4, 0.15σ in grade 6, and 0.13σ in grade 8, considering all six tests. On the grade-equivalent scale, the median advantage for students who spent three years in a small class is about 4.5 months in grade 4, about 4.2 months in grade 6, and about 5.4 months of schooling—over one-half of a school year—in grade 8.

The carryover effects of four years in a small class (K–1–2–3) were highly statistically significant on all tests in all later grades, with the single exception of grade 6 science. Every effect size for four years in a small class was



— Not significant, ○ .05, ● .01, ● .001.

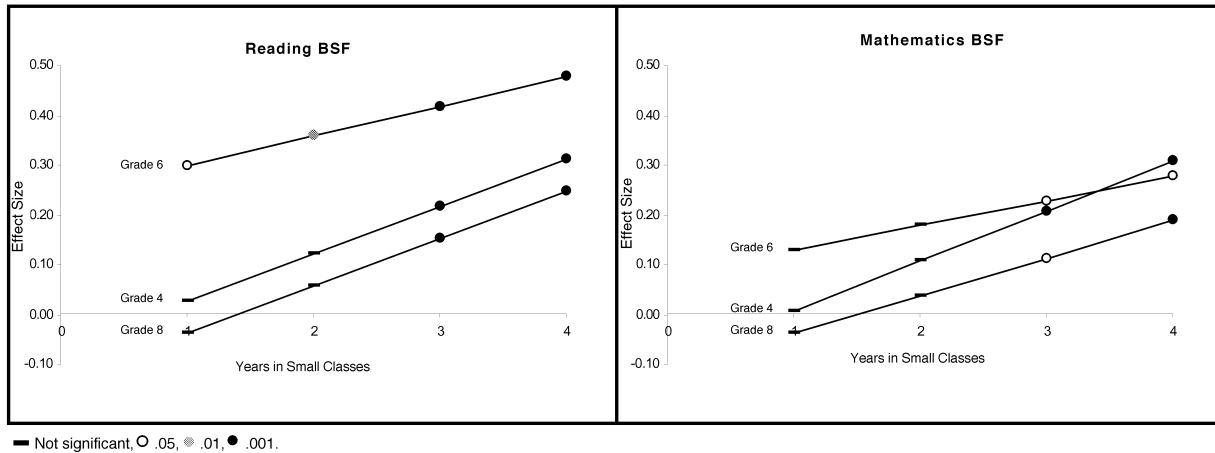


Figure 2: Small-class carryover effects by number of years of participation

larger than the respective effect of three years in a small class. The median effect size was 0.29σ in grade 4, 0.21σ in grade 6, and 0.21σ in grade 8. On the grade-equivalent scale, the median advantage for students who spent four years in a small class was about 5.4 months in grade 4, 5.6 months in grade 6, and 9.0 months in grade 8. In grade 8, students are almost a full school year ahead of their classmates who attended larger classes in K–3.

In contrast to the lower limits estimated from the entire grades 4, 6, and 8 samples, these effect sizes represent *best estimates* of impact on students who attend small classes for one, two, three, or four consecutive years. While one year in a small class did not have a consistently significant impact on later school performance, two years in a small class had somewhat more, and three and four years showed lasting benefits that are statistically and educationally meaningful. Improvements in test scores remained significant through grade 8—fully five years after the small classes were disbanded. Few educational interventions have demonstrated this degree of longevity.

SUMMARY AND CONCLUSIONS

Prior to Project STAR, research on class size was limited largely to nonexperimental studies, most of which were based on small samples or of short duration. To the credit of Tennessee legislators and with the cooperation of schools across the state, a remarkable four-year experiment was designed and executed (see Ritter and Boruch, 1999, for a discussion of the political origins of STAR). Because of its design and magnitude, the results of Project STAR eclipsed most of the work that preceded it.

The original STAR reports demonstrated that small classes in the primary grades have a positive impact on academic achievement in all subject areas in kindergarten through third grade (for example, Word et al., 1990; Finn and Achilles, 1990). Perhaps of equal importance, the project yielded an excellent database for addressing a plethora of additional questions. The present study examined one set of questions in depth: To what extent are the impacts related to the number of years a student attends a small class? We examined the effects on academic achievement both while students attended small classes and in later grades, after they returned to full-size classes.

STAR was a randomized experiment that manipulated and controlled only class size (small or regular) and one pupil-teacher ratio variable (regular and regular with an aide). Natural variations in schooling made STAR classes more heterogeneous each year after kindergarten. Student mobility into and out of STAR schools and in-grade retentions resulted in each class being comprised of students who had participated in a small, regular, or

teacher-aide classroom for different numbers of years. Although these processes occur commonly in elementary and secondary schools, they have the potential to bias statistical outcomes. Students who change schools more frequently are often from lower socioeconomic strata and suffer academically. Fewer students were retained in STAR small classes, leaving fewer “slots” for the previous year’s retainees.

To study duration in small classes, we focused on just those students who spent one year, or else two, three, or four consecutive years, in a small, regular, or teacher-aide class. In examining the data during the experimental years (K–3), we used student age as a proxy for in-grade retention and asked whether benefits of small classes remained even after retentions were controlled statistically. For both the experimental years and post-experimental years, we addressed student mobility as best the data would allow by controlling for factors related to mobility, that is, student race, SES, and school urbanicity.

The immediate benefits of small classes were clear and consistent. Students who attended small classes performed significantly better on all achievement measures in all grades than did students in full-size classes with or without teacher aides. The number of years a student spent in a small class was also a determinant of the benefits realized.¹⁹ In each grade, having participated in a small class for more years produced greater gains than did fewer years. In grades 2 and 3, in fact, the advantages were only statistically significant when a student had been in small classes for two or more years. Those students were generally 3 to 8 months ahead of their peers who attended large classes, depending on the specific test and grade level.

Starting early is also important. One year of participation in a small class had a strong effect on pupil performance in kindergarten and grade 1, but less impact in grades 2 and 3. Two years in a small class had a profound impact on achievement in grade 2 but less effect in grade 3. In general, the later the starting point, the more years of participation are required to improve students’ learning. Starting point and duration cannot be separated clearly in the STAR data, however. Attending small classes for two or more years by grade 2, for example, implies that the student began in a small class by kindergarten or first grade. Thus we are left with the conclusion that entering a small class in the early grades (K or 1) and continuing for two or more years has a meaningful, significant impact on students’ academic achievement.

Writing about non-school settings, social psychologists have documented that individual participation is inversely related to group size; that is, the more people present, the less “involved” each individual tends to be (for example, Darley and Latane, 1968; Levine and Moreland, 1998). STAR-related research suggests that this principle may help explain the benefits

of small classes as well. Pupils in small classes were found to invest more effort toward learning activities, take greater initiative in the classroom, and display less disruptive or inattentive/withdrawn behavior than their counterparts in full-size classes (Finn, 1998; Finn et al., 1989). It appears that the very feature of “smallness” sets the stage for increased student engagement in learning. At the same time, when the class is smaller, teachers are compelled to attend to all students regardless of their abilities, motivation, or classroom demeanor.

In addition to immediate impact, attending small classes also had long-term benefits. In general, students who attended small classes in K–3 performed better academically on all subjects in grades 4, 6, and 8 than their peers who attended full-size classes (see also Nye, Hedges, and Konstantopoulos, 1999). Our expanded analysis of starting point and duration in a small class revealed other considerations. One year in a small class in kindergarten or grade 1 was not sufficient to produce long-term effects, even through grade 4. Two years was somewhat better. Carryover effects were consistently significant only for students who had attended small classes for three to four years. Four years in a small class put students nearly a whole school year ahead of their counterparts who had attended larger classes in K–3. In STAR, four years in a small class was also confounded with starting age, since four-year participants in STAR must have entered small, regular, or teacher-aide classes in kindergarten.

We conclude that entering a small class in kindergarten or grade 1 and remaining in that setting for at least three years produces, on average, significant and noteworthy improvements in academic achievement at least through grade 8 in all school subjects. These findings are consistent with the criteria for lasting effects outlined by Ramey and Ramey (1998): The students started early, continued through a number of years, and their exposure to the program was maximal. Our findings underscore the importance of continued participation in small classes. They refute some detractors of small class policies who argue that one year, at most, produces all the benefits that can be realized from small classes (e.g., Hanushek, 1998). One year (in kindergarten or grade 1) produces early gains but does not produce lasting effects.

This investigation, for the most part, confirmed that attending a class with a full-time teaching assistant had little impact on academic performance in grades K–3 and no significant effects on performance in later grades. The benefits of paraprofessionals may lie in other domains, for example, helping a small number of students who have learning or behavior problems, performing tasks so that a teacher has more time to devote to instruction, or even serving as a culturally relevant link to parents and the community; these possibilities require further study. However, it is clear that adding a teaching assistant to a full-size class does not affect academic

achievement in general and is not an effective alternative to reducing class size if increased learning is the goal.

The results for teacher-aide classes also address a distinction that has sometimes been muddled. In discussing class size research, some authors have failed to distinguish between *pupil-teacher ratios* and the number of pupils in a particular classroom (for example, Hanushek, 1998; Hedges, Laine, and Greenwald, 1994; Wenglinsky, 1997). Pupil-teacher ratios, usually computed for entire school districts or states, include special education and Title I teachers with small classes, subject matter specialists with no full-time classes of their own, librarians, teaching assistants, and others. Aide classes in this study actually had smaller *pupil-teacher ratios* than did small classes: about 24-to-2 or 12-to-1 for aide classes and about 15-to-1 for small classes. Our study shows that small classes have academic benefits not obtained by reducing the pupil-teacher ratio—even in a single classroom.

Many questions remain. Some can be addressed through further analysis of the STAR and follow-up database. For example, Glass and Smith (1978) proposed that a class size “threshold” (fewer than 20 pupils) is required to obtain academic benefits. Natural variation in class sizes within the ranges defined as small or regular in Project STAR could be analyzed to answer the question “How small must a class be to produce the benefits that have been observed?” Also, the database could be used to examine school-to-school variability in the magnitude of the small-class effect. The benefits of small classes were greater in some schools, less in others, and even nonexistent in a few (see Krueger, 1999). Random regression models can be used to ask how much of this variability is due to such factors as school size, the composition of the student body, or the preparation of teachers.

Other questions require new research. For example, like most prior research on class size, Project STAR focused on the primary grades; less is known about the possible benefits of small classes in the middle grades or in high school, although Boozer and Rouse (1995) demonstrated some small-class effects in grades 7 and 8. Even more important, however, is the question of teaching to a smaller group: What can teachers do to take advantage of the opportunities a small class provides, to maximize the impact on pupil learning and behavior? Answers to this question are virtually nonexistent at present. However, well over half of the states and many districts across the country are currently undertaking class-size reduction initiatives; they span the primary, middle, and secondary grades. We urge educators and researchers to make use of these *natural laboratories* to study questions such as these. There is still much to be learned.

APPENDICES

Table A-1. HLM reduced-model results for experimental years (regression coefficients; *p*-values)

Grade/Effect	SAT tests			BSF tests	
	Total Reading	Word Study Skills	Total Mathematics	Reading	Mathematics
<i>Kindergarten</i>					
Gender	-4.9	-5.0	-5.9		
Race	7.1	6.7	17.2		
SES	14.7	15.8	19.8		
Class type					
Small	6.4	7.2	9.1		N/A
Aide	1.1	1.3	0.2		
Urbanicity					
Suburban	13.5	16.9	20.6		
Rural	10.5	11.7	17.7		
<i>Grade 1</i>					
Gender	-10.1	-7.9	-0.2	-0.4	-0.0
Race	7.4	5.4	16.8	0.3	0.6
SES	26.9	24.5	17.6	0.7	0.6
Duration	9.3	7.6	4.8	0.3	0.1
Class type					
Small	16.0	15.8	13.0	0.6	0.5
1 year—regular	8.8	10.7	9.6	0.4	0.4
2 years—regular	21.6	19.8	15.8	0.7	0.5
Aide—Regular	5.6	5.5	2.2	0.2	0.1
1 year—regular	-2.0	0.1	-1.5	0.0	0.0
2 years—regular	11.5	9.7	5.0	0.3	0.2

Urbanicity					
Suburban	36.2	30.0	22.5	0.7	0.9
Rural	41.9	30.7	29.0	0.7	0.9
Interactions ^a				b	

Grade 2

Gender	-8.4	-5.4	-0.2	-0.6	-0.3
Race	8.2	6.1	13.8	0.5	0.4
SES	21.4	22.7	18.6	1.0	0.8
Duration	5.2	6.7	2.4	0.2	0.1
Class type					
Small	11.5	12.5	10.9	0.6	0.5
1 year—regular	5.5	4.1	6.9	0.4	0.3
2 years—regular	10.9	11.6	10.5	0.6	0.4
3 years—regular	16.2	19.2	14.1	0.8	0.5
Aide	3.5	3.7	2.2	0.1	0.1
1 year—regular	-2.7	-5.0	-2.1	-0.1	0.0
2 years—regular	2.8	2.8	1.7	0.1	0.1
3 years—regular	8.3	10.5	5.6	0.3	0.2
Urbanicity					
Suburban	28.5	28.3	18.8	1.5	0.6
Rural	36.7	33.6	30.9	1.7	1.0
Interactions ^a				c	

Grade 3

Gender	-7.8	-8.1	-0.8	-0.5	-0.2
Race	8.0	6.1	8.8	0.4	0.6
SES	15.6	15.3	14.9	0.9	1.1
Duration	3.0	3.4	2.5	0.1	0.1
Class type					
Small	8.1	9.2	6.0	0.3	0.4
1 year—regular	2.9	4.2	3.1	0.2	0.2
2 years—regular	6.0	7.2	4.8	0.3	0.3

continued

Table A-1. continued

Grade/Effect	SAT tests			BSF tests	
	Total Reading	Word Study Skills	Total Mathematics	Reading	Mathematics
3 years—regular	9.2	10.2	6.6	0.4	0.4
4 years—regular	12.4	13.2	8.4	0.4	0.5
Aide	0.7	1.0	-0.7	-0.0	-0.1
1 year—regular	-4.6	-3.8	-3.4	-0.1	-0.2
2 years—regular	-1.4	-0.9	-1.8	-0.1	-0.2
3 years—regular	1.8	2.0	-0.2	0.0	-0.1
4 years—regular	5.1	4.9	1.4	0.1	-0.0
Urbanicity					
Suburban	19.9	23.9	17.2	0.8	1.2
Rural	24.7	28.5	22.9	1.1	1.8
Interactions ^a		^d			

□ NS □ $p < .05$ □ $p < .01$ □ $p < .001$

^aDuration-by-class type interaction also included in all models; required to estimate the contrasts of 1, 2, 3, and 4 years in a small-class setting or teacher aide setting with regular-class students. Not tabled since it does not directly test the interaction of other contrasts in the table.

^bMarginal duration-by-race-by-small interaction ($p < .02$).

^cSignificant race-by-aide and duration-by-urbanicity-by-small interactions ($p < .01$).

^dMarginal urbanicity-by-small interaction ($p < .02$).

Table A-2. HLM results for post-experimental years, full sample (regression coefficients; *p*-values)

Grade/Effect	CTBS tests			BSF tests		
	Total Reading	Total Mathematics	Science	Social Science	Reading	Mathematics
Grade 4						
Gender	-6.8	-8.4	5.1	-1.0	-0.6	-0.2
Race	12.5	7.5	12.1	7.7	0.7	0.7
SES	23.3	19.6	21.6	27.3	1.1	1.0
Class type						
Small—Regular	7.6	5.6	6.0	7.1	0.3	0.4
Aide—Regular	2.0	1.0	1.2	2.2	0.1	0.1
Urbanicity						
Suburban	32.7	17.7	38.2	33.3	1.7	1.3
Rural	34.0	22.6	42.4	37.9	1.7	1.5
Grade 6						
Gender	-0.4	-6.4	1.0	-1.2	-0.6	-0.3
Race	13.4	8.6	18.4	5.9	0.7	0.5
SES	19.9	20.8	21.9	19.8	1.2	1.1
Class type						
Small—Regular	4.2	4.6	1.3	5.1	0.5	0.3
Aide—Regular	-0.7	1.5	-2.1	1.9	0.2	0.1
Urbanicity						
Suburban	19.3	15.0	29.7	13.4	1.4	0.8
Rural	17.8	14.9	31.2	12.7	1.6	1.1
Grade 8						
Gender	-4.7	-5.6	8.5	-3.8	-0.7	-0.2
Race	16.0	12.0	15.8	12.9	0.9	0.7
SES	20.8	21.3	17.1	18.6	1.2	1.2
Class type						
Small—Regular	4.8	3.9	4.1	3.8	0.3	0.3
Aide—Regular	0.5	0.1	0.9	0.9	0.1	0.0
Urbanicity						
Suburban	25.7	21.7	26.7	27.7	1.5	1.6
Rural	27.7	24.3	28.0	28.7	1.6	1.7

□ NS □ $p < .05$ □ $p < .01$ □ $p < .001$

Portions of this paper were presented at the conference on the Economics of School Reform at the Hebrew University of Jerusalem, May 23–26, 1999. This work was supported by a grant from the Spencer Foundation entitled “A Study of Class Size and At-Risk Students.” We are grateful to HEROS, Inc., for assistance in obtaining the data used in this investigation, and to Steven Raudenbush and John Willett for advice on the statistical methods.

Notes

1 Project STAR was directed by Ms. Elizabeth Word of the Tennessee Department of Education and conducted by a consortium of researchers from four Tennessee universities. The Principal Investigators were C. M. Achilles (University of Tennessee), H. P. Bain (Tennes-

see State University), J. Folger (Vanderbilt University), J. Johnston (University of Memphis, for three years), and Fred Bellott (University of Memphis, one year). Jeremy Finn was external evaluator for the duration of the project.

2 Several national conferences have been held on this topic in the United States, using STAR as the model.

3 The same findings were confirmed by Goldstein and Blatchford (1998) and by Krueger (1999) using other statistical approaches.

4 This was done to assuage the concerns of some parents whose children had neither small classes nor a teacher aide. By the end of four years, this reassignment was judged to be relatively unimportant because no differences were found between regular and teacher-aide classes.

5 These were already subtracted from the sample sizes reported above (that is, 5,394 to 5,910).

6 Student variables were centered at the classroom means, class variables at the school means, and school variables were grand-mean centered. All terms in the model were treated as fixed except for the student-, classroom-, and school-level intercepts, which were treated as random.

7 These were obtained by evaluating algebraically the “expected values” of the final models for particular values of the independent variables (for example, one year of small classes; two years of small classes; regular classes) and subtracting (see Bryk and Raudenbush, 1992). This process yields a set of coefficients that multiply the regression weights to estimate each contrast of interest. The coefficients can be entered back into the HLM program—the easiest way to compute estimates and tests of the desired contrasts.

8 The small-regular contrast for BSF mathematics in grade 3 was not significant at the .01 level; the *p*-value was .017.

9 Not tabled.

10 This may represent a real decline in the advantage of small classes in mathematics, because the effect size also decreases in grade 3. Although the effect is still substantial, this apparent inconsistency should be examined through further research including an analysis of the mathematics curriculum in grades 1 through 3.

11 The criterion-referenced BSF tests do not lend themselves to grade-equivalent scores.

12 Note, however, that Head Start programs are varied and some have different patterns of outcomes from others.

13 Including the District of Columbia.

14 Due to technical errors, portions of the grade-6 BSF data were lost when the original files were created and are irretrievable. The students for whom data are available, however, comprise a substantial group who attended all three class types in all three school locations (inner-city, suburban, rural). We judged that including these data in the analyses would be worthwhile.

15 Grade-2 and grade-3 entrants also presented problems because of the absence of information about their prior schooling and class placements.

16 Grade-2 and grade-3 entrants comprised about 29% of the overall sample. Samples for regular classes were reduced at a higher rate due to the reassignment of students between teacher-aide and regular classes from kindergarten to grade 1; we classified pupils who participated in teacher-aide classes in any grade as “teacher-aide” students, to keep the definitions consistent.

17 The procedures used were the same as in the K-3 analysis.

18 We examined the data to try to explain the anomalous lack of gender differences on the grade-6 CTBS tests but could find nothing out of the ordinary in the distributions.

19 The median number of students in STAR “regular” classes was 22-24, fewer than the number in many classrooms in schools today. Krueger (1999) found a statistically significant negative correlation between actual class size and achievement among the STAR regular

classes (the control group). It is possible that reducing class sizes in many of today's schools would have larger effects than those found in STAR.

References

- Abelson, W. D., Zigler, E., & DeBlaise, L. L. (1974). Effects of a four-year Follow Through program on economically disadvantaged children. *Journal of Educational Psychology*, *66*, 756–771.
- Achilles, C. M., Kiser-Kling, K., Owen, J., & Aust, A. (1994). *Success starts small: Life in a small class. Small Grant-Field-Based Research Final Report*. Greensboro, NC: University of North Carolina.
- Achilles, C. M., Nye, B. A., & Bain, H. P. (1994–95). The test-score “value” of kindergarten for pupils in three class conditions, at grades 1, 2, and 3. *National Forum of Educational Administration and Supervision Journal*, *12*, 3–15.
- Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, *5* (3), 25–50.
- Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19*. Ypsilanti, MI: High Scope Press.
- Bloom, B. S. (1964). *Stability and change in human characteristics*. New York: John Wiley & Sons.
- Boozer, M., & Rouse, C. (1995, June). *Intraschool variation in class size: Patterns and implication* (Working paper no. 344). Washington, DC: National Bureau of Economic Research, Industrial Relations Section. ERIC document no. ED 385–935.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, NJ: Sage Publications.
- Burket, G. R. (1984). Response to Hoover. *Educational Measurement: Issues & Practice*, *3* (4), 15–16.
- Council of Chief State School Officers. (1996, October). *Key state education policies on K–12 education*. Washington, DC: Author.
- CTBS/McGraw Hill. (1989). *CTBS: Comprehensive tests of basic skills*. Monterey, CA: Author.
- Cryan, J. R., Sheehan, R., Wiechel, J., & Bandy-Hedden, I. G. (1992). Success outcomes of full-day kindergarten: More positive behavior and increased achievement in the years after. *Early Childhood Research Quarterly*, *7*, 187–203.
- Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *10*, 202–214.
- Evertson, C. M., & Folger, J. K., (1989, March). *Small class, large class: What do teachers do differently?* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Finn, J. D. (n.d.). *Full-day kindergarten: Answers with questions*. Philadelphia: Laboratory for Student Success, Temple University Center for Research in Human Development and Education.
- Finn, J. D. (1998). *Class size and students at risk: What is known? What is next?* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Institute on the Education of At-Risk Students.
- Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, *27*, 557–577.
- Finn, J. D., Fulton, B. D., Zaharias, J., & Nye, B. A. (1989). Carry-over effects of small classes. *Peabody Journal of Education*, *67*, 75–84.
- Glass, G. V., & Smith, M. L. (1978). *Meta-analysis of research of the relationship of class size and achievement*. San Francisco: Far West Laboratory for Educational Research and Development.
- Goldstein, H., & Blatchford, P. (1998). Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal*, *24*, 255–268.

- Hanushek, E. A. (1998). *The evidence on class size*. Rochester, NY: University of Rochester, W. Allen Wallis Institute of Political Economy.
- Harvey, B. (1993, December). *An analysis of grade retention for pupils in K-3*. Unpublished doctoral dissertation. University of North Carolina, Greensboro.
- Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Money does matter somewhere: A reply to Hanushek. *Educational Researcher*, 23 (4), 9-10.
- Hoover, H. D. (1984). The most appropriate scores for measuring educational development in the elementary school: GE's. *Educational Measurement: Issues & Practice*, 3 (4), 8-14.
- Karweit, N. (1989). Effective kindergarten programs and practice for students at risk. In R. E. Slavin, N. L. Karweit, & N. A. Madden (Eds.), *Effective programs for students at risk* (pp. 103-142). Boston: Allyn & Bacon.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114, 497-532.
- Lazar, I., & Darlington, R. (1982). *Lasting effects of early education: A report from the consortium for longitudinal studies*. Chicago: University of Chicago Press for the Society for Research in Child Development.
- Levine, J. M., & Moreland, R. L. (1998). Small groups. In D. T. Gilbert, S. T. Fiske, and G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 415-469). New York: McGraw Hill.
- McKey, R. H., Condelli, L., Granson, H., Barnett, B., McConkey, C., & Plantz, M. (1985, June). *The impact of Head Start on children, families, and communities*. Washington, DC: Head Start Bureau, U.S. Department of Health and Human Services. Available through the Government Printing Office.
- Naron, N. K. (1981). The need for full-day kindergarten. *Educational Leadership*, 38, 306-309.
- Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 21, 127-142.
- Pate-Bain, H., Boyd-Zaharias, J., Cain, V. A., Word, E. R., & Binkley, M. E. (1997, September). *The student/teacher achievement ratio project: STAR follow-up studies 1996-1997*. Lebanon, TN: HEROS, Inc.
- Peterson, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221-264). New York: Macmillan.
- Psychological Corporation, Harcourt Brace Jovanovich. (1983). *Stanford Achievement Test* (7th ed.) San Diego: Author.
- Ramey, C. T., & Ramey, S. (1998). Early intervention and early experience. *American Psychologist*, 53, 109-120.
- Reynolds, A. J. (1997). *The Chicago child-parent centers: A longitudinal study of extended early childhood intervention* (Discussion paper no. 1126-97). Madison, WI: Institute for Research on Poverty.
- Ritter, G. & Boruch, R. (1999). The political and institutional origins of the Tennessee experiment. *Educational Evaluation and Policy Analysis*, 21, 111-125.
- Robinson, G. (1990). Synthesis of research on the effects of class size. *Educational Leadership*, 47 (7), 80-90.
- Robinson, G. E., & Wittebols, J. H. (1986). *Class size research: A related cluster analysis for decision making*. Arlington, VA: Educational Research Service.
- Schulz, E. M., & Nicewander, W. A. (1997). Grade equivalent and IRT representations of growth. *Journal of Educational Measurement*, 34, 315-331.
- Slavin, R. E. (1989). Achievement effects of substantial reductions in class size. In R. E. Slavin (Ed.), *School and classroom organization* (pp. 247-257). Hillsdale, NJ: Erlbaum.
- Wenglinsky, H. H. (1997). How money matters: The effect of school district spending on academic achievement. *Sociology of Education*, 70, 221-237.
- White, K. R. (1985-86). Efficacy of early intervention. *The Journal of Special Education*, 19, 400-416.

- Word, E., Johnson, J., Bain, H. P., Fulton, D. B., Boyd-Zaharias, J., Lintz, M. N., Achilles, C. M., Folger, J., & Breda, C. (1990). *Student/teacher achievement ratio (STAR): Tennessee's K-3 class-size study*. Nashville, TN: Tennessee State Department of Education.
- Yen, W. M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement*, 23, 299-325.
- Zigler, E., & Styfco, S. J. (1994). Is the Perry Preschool better than Head Start? Yes and no. *Early Childhood Research Quarterly*, 9, 269-287.

JEREMY D. FINN is Professor of Education at State University of New York at Buffalo. His research interests include students and schools at risk, student resilience, educational equity, classroom organization, and statistical methods. He has been conducting research on class size and teacher aides since 1985, when he began as external evaluator for Tennessee's Project STAR.

SUSAN B. GERBER is Associate Director of the Center for the Study of Technology in Education and Adjunct Assistant Professor in the Graduate School of Education at State University of New York at Buffalo. Her research interests include policy- and instruction-related approaches to providing all students with essential learning opportunities.

CHARLES M. ACHILLES is Professor of Education Leadership at Eastern Michigan University in Ypsilanti and author of *Let's Put Kids First, Finally: Getting Class Size Right* (Corwin Press, 1999). He was a principal investigator of Project STAR and has conducted class size research continuously since 1984.

JANE BOYD-ZAHARIAS is the Director of Health & Education Research Operative Services, Incorporated (HEROS, Inc.), a not-for-profit research agency in Tennessee. Prior to joining HEROS, she served as Director of Class Size Studies at Tennessee State University. Boyd-Zaharias is a co-principal investigator of STAR follow-up studies and has authored more than 30 publications and presentations related to class size research.