American Educational Research Journal
Spring 2000, Vol. 37, No. 1, pp. 123–151

The Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class Size Experiment

Barbara Nye
Tennessee State University
Larry V. Hedges
Spyros Konstantopoulos
The University of Chicago

The effects of class size on academic achievement have been studied for decades. Although the results of small scale randomized experiments and large-scale econometric studies point to positive effects of small classes, some scholars have seen the evidence as ambiguous. This paper reports analyses of a 4-year, large-scale randomized experiment on the effects of class size, project STAR in Tennessee. Although implementation was not perfect, these analyses suggest that shortcomings in implementation probably led to underestimates of the effects of class size. The analyses reported here suggest class size effects that are large enough to be important for educational policy and that are quite consistent across schools. Thus, small classes appear to benefit all kinds of students in all kinds of schools.

BARBARA NYE is the director of the Center of Excellence for Research and Policy on Basic Skills, Tennessee State University, 330 10th Ave. N., Box 141, Nashville, TN 37203-3401. Her specialties include research on class size, teacher education, science education, and policy studies.

LARRY V. HEDGES is the Stella M. Rowley Professor of Education Psychology, Sociology, and the Harris Graduate School of Public Policy Studies at the University of Chicago. His address is Department of Education, University of Chicago, 5835 S. Kimbark Ave., Chicago, IL 60637. His primary research interests are statistical methods for education and cognitive science and policy studies.

SPYROS KONSTANTOPOULOS is a doctoral student at the Department of Education at the University of Chicago. His primary research interests are statistical methods in educational and psychological research.

How to allocate resources to most effectively further the aims of educational systems is one of the enduring questions facing educational researchers and policy makers. Among the most fundamental allocation decisions is the decision of how to assign instructional staff to classes. This decision includes determining class size and whether or not ancillary instructional staff, such as classroom aides, should be employed. Manipulating class size is a policy option that is gaining attention nationwide. About 18 states have adopted policies that reduce class size, with the goal of improving achievement; the reduction of class size is also included in the president's education initiative. In addition, the widespread use of classroom aides indicates a need for more staff.

Although decisions about allocation of resources (particularly class size) are made in every educational system, there is serious debate about whether these allocation decisions even influence academic achievement and other desired outcomes of education. This paper investigates whether assignment to small classes has effects on academic achievement and, if so, how those effects are distributed across schools with different social composition and teacher characteristics.

Related Literature

Using both experimental and nonexperimental methods, the effects of class size have been investigated empirically for decades. Well over 100 experimental and quasi-experimental studies of the effects of class size have been conducted, each involving assignments of students to smaller or larger classes. This body of literature has been reviewed by Glass, Cahen, Smith, and Filby (1982); Glass and Smith (1979); Hedges and Stock (1983); and Mosteller, Light, and Sachs (1996). There is some disagreement about the interpretation of the experimental research (e.g., Educational Research Services, 1980; Slavin, 1984). However, this synthesis of research on the effects of class size suggests positive effects of class size reduction on achievement and affect, with the effects becoming larger as the classes become smaller.

However, these studies suffer from significant limitations. For example, the investigations are typically small scale and short-term. As a result, it is difficult to know whether the effects are a consequence of the special circumstances surrounding the experiment or whether they would have occurred if the smaller classes had been used in a more natural setting. That is, the findings may be internally valid, but it is difficult to know whether they are externally valid and would generalize to other settings.

A second tradition in studying the effects of class size is the econometric work on education production functions (Hanushek, 1986). This research uses the distribution of naturally occurring class sizes, modeling the relationship between class size and an outcome (usually achievement) while controlling for student characteristics such as social class or prior achievement. There have been a considerable number of econometric studies of the effects of class size on achievement (Hanushek, 1989; Hedges, Laine, &

Greenwald, 1994). There is some controversy over the interpretation of the econometric studies. Some researchers (Hanushek, 1989) are persuaded that effects of class size must be small because so few of the studies found statistically significant effects. Others (Greenwald, Hedges, & Laine, 1996) have argued that a better gauge of the size of the effects is the magnitude of the actual regression coefficients obtained in the studies, regardless of their individual statistical significance. Whatever the proper method of summary may be, the econometric studies are probably more externally valid than small-scale experiments because they use data from operating schools.

Econometric studies, however, are limited in their internal validity. That is, it is difficult to know if the relations between class size and achievement (controlling for student background) are causal. In many cases, the student background data are rather limited and may fail to fully account for individual differences among students assigned to classes of different sizes. For example, a plausible hypothesis is that achievement (or expected achievement) causes students to be assigned to classes of different sizes, not the other way around. In other words, students are assigned to smaller classes precisely because their achievement is low (e.g., in compensatory or remedial programs). Thus, the previous research of both types is limited for the purposes of determining whether there are class size effects.

Project STAR

An important source of information on class size that has not been adequately exploited to date is the Tennessee class size experiment or Project STAR (Student-Teacher Achievement Ratio). This randomized experiment was commissioned in 1985 by the Tennessee state legislature and implemented by a consortium of Tennessee universities and the Tennessee State Department of Education. The total cost of the experiment, including the cost of hiring new teachers and classroom aides, was approximately $12 million.

Initially, all Tennessee school districts were asked to participate in Project STAR. About 180 schools in roughly 50 of the 141 school systems in the state expressed interest in participating in the project. Only about 100 schools had enough students in each grade to meet the size criterion for participation (at least 57 students per grade necessary to form one small and two regular-sized classes). This size criterion, which was necessary to permit assignment to class types within schools, excluded very small schools from the study. Ultimately, 79 elementary schools in 42 school districts became sites in the STAR experiment. Districts had to agree to participate for 4 years and to allow site visitations for verification of class sizes, interviewing, and data collection, including additional student testing. They also had to allow random assignment of pupils and teachers to class types from kindergarten through Grade 3.

The state paid for the additional teachers and classroom aides, and only class size conditions changed within the schools. School districts and buildings followed their own policies and curricula. The study design provided that no student would receive any less service than would normally be provided by the state as a consequence of being in Project STAR. Thus, there was no incentive for any student *not* to participate and schools had an incentive (in the form of overall greater resources) to participate.

The experiment randomly assigned kindergarten students into small classes (13–17 students), large classes (22–26 students), or large classes with a full-time classroom aide. Teachers were also randomly assigned to classes of different types. The assignments of students and teachers to class type were maintained through the third grade. Some students entered the study in the first grade and subsequent grades but were randomly assigned to classes at that time.

This data source (called "one of the great experiments in education in U.S. history" by Mosteller et al., 1996, p. 814) substantially mitigates many of the problems of other class size research. Because it is a large-scale, randomized experiment that randomizes both teachers and students into classrooms within each participating school, it has high internal validity. Project STAR involves a range of schools from a rather diverse state. It includes both large, urban and small, rural districts and participants from some of the wealthiest and poorest school districts in the country. Therefore, Project STAR includes essentially the entire range of educational conditions that occur in American education and it is more likely to be generalizable than smaller, more circumscribed studies conducted in only one location. Moreover, it was conducted for 4 years as part of the everyday operation of the schools and, therefore, is likely to avoid the effects associated with new, experimental programs.

The Project STAR database and the long-term follow-up of the students (Nye et al., 1994) are part of a larger program of research on class size conducted by the Center of Excellence for Research in Basic Skills at Tennessee State University.

Previous Analyses of Data From Project STAR

Despite the great importance of the Project STAR data, previous analyses have not been entirely satisfactory. The analyses reported during the original project (Word et al., 1990) were not extensive or sophisticated. The few published analyses of the data (Finn & Achilles, 1990) did not remedy the situation, and the most prominent discussions of the data (Mosteller et al., 1996) relied on the original analyses.

There are three major problems with previous analyses. One is purely technical; the other two are more conceptual. The technical problem is common to many analyses of school and classroom effects that ignore the consequences of clustering of students within schools and classrooms (Bryk & Raudenbush, 1992). The issue is that students in the same school are likely to be more similar in their achievement test scores than students in different schools. Consequently, the sampling uncertainty of treatment effects and other features of the data tends to be underestimated, often making results appear to have greater statistical significance than is justified. The analyses of variance (ANOVA) and multivariate analyses of variance (MANOVA) reported in earlier analyses of the Project STAR data are subject to these deficiencies.

The first conceptual problem concerns the fidelity of the treatment implementation. The most important treatment in the Project STAR study is the manipulation of class size. To meaningfully interpret the results of the experiment, it is important to know whether the class sizes intended by the experimenters were realized in the actual experiment. Were the classes assigned to be small actually in the size range intended? Were the classes intended to be larger actually in the size range intended?

The second conceptual problem concerns the interpretation of results of imperfect randomized experiments. By "imperfect," we mean experiments in which there are students who drop out (attrition) or students who eventually appear in classes of a type other than that to which they were originally assigned (switching between treatment groups).

Given this definition of imperfect experiments, it is apparent that Project STAR (and virtually every other large-scale field experiment) is imperfect (Cook & Campbell, 1979). The fact that an experiment is imperfect does not mean that its results are inevitably compromised. It does mean, however, that the mere fact of randomization alone, which guarantees the internal validity of perfect randomized experiments, cannot be relied upon to assure the validity of imperfect experiments. Thus, it is essential to examine the threats to validity posed by attrition and switching among treatment groups; most previous published analyses of the STAR data ignored these problems.

More recent (to date, unpublished) analyses of data from the STAR experiment (Hanushek, 1999; Krueger, 1997) have attempted to address the question of imperfections in the experiment but not the issue of clustering within schools and classrooms. For example, Krueger (1997) studied the pattern of missing data and inputted values for missing test score data to see if the results of the experiment would change when the inputted values were added. He concluded that the treatment effects were largely unaffected by attrition, but that they seemed to be greatest among students who began the experiment in kindergarten. Hanushek (1999) believes that attrition may have affected estimates of the effects of small classes and raises the issue of whether the randomization may have failed in some way, but notes that the latter is quite difficult to determine from the available data.

Validity of Findings From Project STAR

We begin by addressing the conceptual issues of fidelity of treatment implementation, attrition, and switching among treatments. In a later section, we address the technical issues in new analyses using hierarchical linear models, including whether the effects increase with longer exposure to small classes.

Table 1
Distribution of Actual Class Sizes Among Classes Assigned to Each Type

	Kindergarten			First grade		
	Type of class			Type of class		
Actual class size	Small	Regular	Regular/aide	Small	Regular	Regular/aide
12	8			2		
13	19			14		
14	22			18		
15	23		1	31		
16	31	1		16	1	
17	24	4	1	33	1	
18		1	2	6	2	
19		7	6	3	4	3
20		6	6	1	10	6
21		14	12		18	18
22		20	20		27	15
23		16	21		19	20
24		19	14		16	11
25		6	6		7	9
26		4	3		5	9
27		1	6		2	4
28			1		1	2
29					1	2
30					1	1
Total	127	99	99	124	115	100
Average	14.96	22.16	22.54	15.52	22.47	23.20

(Continued)

Fidelity of Implementation

In the STAR experiment, the primary treatment of interest is the manipulation of class size. The intent of the STAR experiment was to compare the achievement of students in small classes (13–17 students) with that of students in larger classes (22–26 students) and in larger classes with full-time aides. However, as in any field experiment, it is important to determine how well the treatment was implemented, because implementation is never perfect at all sites. Thus, in evaluating the STAR experiment, it is important to determine the actual size of the classes to see if the intent of the experimenters was realized.

The actual sizes of the classes assigned to each type (small, regular, and regular with aide) are given in Table 1. The upper lefthand panel provides the number of pupils in each of the 325 kindergarten classrooms in the study. It shows that the small class treatment was relatively well implemented. All of the 127 classes assigned to the small type had 17 or fewer students and the

Table 1 (Continued)

	Second grade			Third grade		
	Type of class			Type of class		
Actual class size	Small	Regular	Regular/aide	Small	Regular	Regular/aide
11				2		
12	3			2		
13	16			15		
14	27			17		
15	32			31		
16	29	1		31		1
17	19			27		
18	6			10	1	
19	1	3	3	5		4
20		-2	1		9	3
21		7	11		11	12
22		23	21		13	16
23		20	21		10	14
24		22	25		15	14
25		9	15		16	15
26		6	7		5	12
27		4	1		5	8
28		1			2	6
29		2	2		2	2
Total	133	100	107	140	89	107
Average	15.16	23.29	23.32	15.53	23.42	23.77

average number of students in the kindergarten classes assigned to be small was almost exactly 15. The average actual number of students in the classes assigned to be regular was about 22; there was more variation in the actual sizes of classes assigned to be regular (or regular with a full time aide). A total of 33 (33%) of the 99 kindergarten classes assigned to be regular sized and 28 (28%) of the classes assigned to be regular sized with aide were smaller than the minimum intended. A total of 7 of the classrooms assigned to be regular sized were as small as some of those assigned to be small classes. Such an overlap of class sizes would tend to weaken any observed effects of class sizes because some of the classes assigned to be regular sized were actually small.

The upper righthand panel of Table 1 shows that the average actual class sizes in the first grade (15.5 students, 22.5 students, and 23.2 students, respectively) remained consistent with those intended in treatment assignment. However, there is somewhat more overlap between the actual class sizes of different types. For example, 10 (8%) of the 124 classes assigned to be small had more than 17 students, whereas 36 (31%) of the 115 classes assigned to be regular sized and 27 (27%) of the 100 classes assigned to be regular sized with aide actually had fewer than 22 students.

The lower lefthand panel of Table 1 shows that the *average* actual class sizes in the second grade are again consistent with the treatment assignment, with classes assigned to be small actually having an average of 15.2 students, those assigned to be regular sized actually having an average of 23.3 students, and those assigned to be regular sized with a full-time aide actually having an average of 23.3 students. Only 7 (5%) of the 133 classes assigned to be small had more than the nominal maximum of 17 students, and only 28 (13.5%) of the 207 total classes assigned to either of the two regular size categories were smaller than the nominal minimum of 22.

The lower righthand panel of Table 1 shows that the situation is similar in the third grade. The average actual class sizes were quite close to those intended (15.5, 23.4, and 23.8, respectively). Only 15 (11%) of the classes assigned to be small were larger than nominal, whereas 41 (20%) of the 196 classes assigned to be regular sized were smaller than nominal.

Thus, although there was some overlap in the actual sizes of the classes assigned to be large compared with those assigned to be small, nearly all (over 89% in every grade) of the classes assigned to be small had actual sizes within the intended range. A majority (over 69% in every case) of the classes assigned to be larger had actual class sizes within the intended range. It is important to recognize that the overlap in the actual class size distributions would tend to reduce the absolute effect of class size. This is because the classes of different types (particularly the larger classes) tended to be more like the small classes than called for in the experimental design.

Attrition From the Study

In the STAR experiment as in other large field studies, there was attrition. For various reasons, some of the students who began the study did not remain in the schools or they were absent when the achievement tests were given. Attrition can be a source of bias in estimating treatment effects if the students who drop out of one treatment group are systematically different from those in other treatment groups. In particular, if the dropouts from the regular-sized classes had higher average achievement test scores than the dropouts from the small classes, then the attrition would lead to positive bias in the effect of small classes; it would make small classes look more favorable than they actually are. Alternatively, if the dropouts from the smaller classes had higher achievement test scores than those of the regular classes, then attrition would lead to negative bias.

Approximately 20%–30% of the students in the STAR study left the study each year. Table 2 shows the number and percentage of students who left each treatment group during each year of the study, along with the average reading and mathematics achievement test scores for those who dropped out. For example, Table 2 shows that 500 (26.3%) of the 1,900 kindergartners assigned to small classes dropped out before they were tested in the first grade and that the average reading and mathematics test scores of the dropouts from small classes were 431.4 and 473.5, respectively.

Table 2
Student Attrition Across Grades

		Dropouts			
	Total *N*	*N*	Percent	Average math score	Average reading score
Kindergarten to first grade					
Kindergarten (small class)	1,900	500	26.3	473.5	431.4
Kindergarten (regular class)	2,194	668	30.5	467.8	425.3
Kindergarten (regular/aide class)	2,231	642	28.8	469.0	427.3
Totals and averages	6,325	1,810	28.6	469.8	427.7
First to second grade					
First grade (small class)	1,925	443	23.0	517.8	504.7
First grade (regular class)	2,584	732	28.3	504.8	489.2
First grade (regular/aide class)	2,320	605	26.1	512.5	497.6
Totals and averages	6,829	1,780	26.1	510.6	495.9
Second to third grade					
Second grade (small class)	2,016	380	18.9	568.3	570.5
Second grade (regular class)	2,329	525	22.5	563.1	564.4
Second grade (regular/aide class)	2,495	522	20.9	561.5	563.4
Totals and averages	6,840	1,427	20.9	563.8	565.6

Table 2 shows that in each year, the percentage of dropouts was slightly smaller in the small classes than in either of the larger classes. The difference in attrition rates between small and regular-sized classes is statistically significant at the .05 level for all grades. The difference in attrition rate between first and second grade between small and regular with aide classes is also statistically significant.

To determine whether these differences in attrition rates might bias small class effects, we examined the differences in achievement test scores between small and regular classes of students who later dropped out versus those who did not (Table 3). That is, for each year, we examined the achievement test scores in the previous year of the students who dropped out (leavers) and those who did not (stayers). We computed the treatment effect (the difference between achievement test scores in small and regular classes) for the leavers and stayers separately, then tested whether they were different. We found that, for every year and every subject matter, the treatment effect computed for leavers was nearly identical to that computed for the stayers. None of the differences was statistically significant at the .05 level.

Because the treatment effect (in the previous year) for the students who would drop out the next year was not different from that of the students who did not drop out the next year, it seems implausible that attrition substantially biased the treatment effects in the following year.

Table 3
Treatment Effect (Difference Between Small and Regular Classes) for Stayers and All Students (Grades K–3)

		Math			Reading		
Grade	Group	Treatment effect	Gap	*t* test	Treatment effect	Gap	*t* test
K	All students	7.94	−0.12	.946	5.46	0.42	−.017
	Stayers in Grade 1	8.06			5.04		
1	All students	11.36	0.77	.335	12.86	1.24	−.051
	Stayers in Grade 2	10.59			11.62		
2	All students	8.41	0.33	.564	9.22	0.34	.618
	Stayers in Grade 3	8.08			8.88		

Note. We computed the *t* test on the difference between stayers and leavers.

Switching Among Class Types

The experimental design called for students to be randomly assigned to class type in kindergarten or whenever they entered the experiment. In the first grade only, students who were assigned to regular-sized classes or regular-sized classes with aides were randomized again to receive either regular-sized classes or regular-sized classes with aides. Therefore, if all the actual placements had been as assigned in the experiment, all of the students assigned to small classes would have remained in small classes. Fifty percent of those assigned to either type of regular-sized classes at kindergarten would have been reassigned to the other type regular-sized class in first grade. There should have been no reassignments in either second or third grade.

The pattern of actual transitions among assignments is given in Table 4. The top panel shows that only 92.3% of the students assigned to small classes in kindergarten remained in small classes in first grade, not 100% as specified in the experimental design. About 4.3% of the students in the small kindergarten classes were reassigned to regular first-grade classes and about 3.4% of the students were reassigned to regular classes with full-time aides. Only 48.3% of the students in regular-sized kindergarten classes were assigned to regular-sized first-grade classes and only 44.4% of the students in regular-sized classes with full-time aides were assigned to the same type of class in first grade. In both cases, the experimental design called for 50% of the students to remain in the classes of each type. There appears to be somewhat greater transition into small classes than out of them.

Examining the pattern of transitions among class types from first to second grade and from second to third grade (middle and bottom panels in Table 4), more than 95% of students in small classes in either of these 2 years remained in small classes. However, there was somewhat greater transition out of the two types of regular-sized classes.

Table 4

Student Transitions Among Types of Classes Across Grades

From kindergarten to first grade				
		Type of first-grade class		
Type of kindergarten class	*N*	Small class	Regular class	Regular/aide class
Small class	1,400	92.3%	4.3%	3.4%
Regular class	1,526	8.3%	48.3%	43.4%
Regular/aide class	1,589	7.7%	47.9%	44.4%
Total	4,515			
From first to second grade				
		Type of second-grade class		
Type of first-grade class	*N*	Small class	Regular class	Regular/aide class
Small class	1,482	96.8%	1.6%	1.6%
Regular class	1,852	8.2%	80.9%	10.9%
Regular/aide class	1,715	2.3%	6.7%	91.0%
Total	5,049			
From second to third grade				
		Type of third-grade class		
Type of second-grade class	*N*	Small class	Regular class	Regular/aide class
Small class	1,636	95.6%	2.3%	2.1%
Regular class	1,804	9.3%	82.3%	8.4%
Regular/aide class	1,973	2.0%	3.9%	94.1%
Total	5,413			

What are the implications for bias? One might characterize these transitions as "exchanging" of students between assignment types. If the achievement test scores of the students exchanged between assignment types differ substantially, then such exchanges can bias estimates of treatment effects. For example, suppose that the students who were assigned to small classes in kindergarten and were actually in (were exchanged to) regular- sized classes in first grade had lower average achievement test scores than those who were assigned to regular classes in kindergarten and were actually in (were exchanged to) small classes in first grade. Then, because the small classes exchanged relatively lower-achieving students for relatively higher-achieving ones, the small class effect would be overestimated.

Table 5 presents the mean mathematics and reading achievement test

Table 5
Student Transitions From Kindergarten to First Grade

Type of class transition	*N*	Percent	Average math score	Average reading score
From small kindergarten class size				
Small to small	1,292	92.3	497.8	444.6
Small to regular	60	4.3	485.1	432.1
Small to regular/aide	48	3.4	487.0	432.5
Total	1,400			
From regular kindergarten class size				
Regular to small	126	8.3	488.5	438.5
Regular to regular	737	48.3	488.5	438.2
Regular to regular/aide	663	43.4	491.2	439.3
Total	1,526			
From regular/aide kindergarten class size				
Regular/aide to small	122	7.7	481.5	434.9
Regular/aide to regular	761	47.9	490.1	440.7
Regular/aide to regular/aide	706	44.4	486.9	436.8
Total	1,589			

scores of the students making each possible transition between class types from kindergarten to first grade. The table shows that the students who continue in small classes have substantially higher average achievement test scores than those who continue in either type of regular-sized classes. It also shows that the students who exchange from small classes to regular-sized classes have slightly lower achievement test scores in both reading and mathematics than those who make the reverse transition (from regular-sized classes to small classes). The students who make the transition from small classes to regular classes with aides have higher mathematics and lower reading achievement test scores than those making the reverse transition. Finally, the students making transitions between the two types of regular-sized classes have nearly identical average achievement test scores. Given the conflicting directions of these differences, it is difficult to determine whether any bias due to these exchanges would tend to increase or decrease the small class effects at first grade.

Table 6 presents the mean mathematics and reading achievement test scores of the students making each possible transition between class types from first grade to second grade. The table shows that the students who continue in small classes have substantially higher average achievement test scores than those who continue in either type of regular-sized classes. It also shows that the students who change from small classes to regular-sized classes have substantially higher achievement test scores in both reading and mathematics compard with those who make the reverse transition (from regular-sized classes to small classes). The students who make the transition from small classes to regular classes with aides have higher mathematics and

Table 6
Student Transitions From First Grade to Second Grade

Type of class transition	*N*	Percent	Average math score	Average reading score
From small first-grade class size				
Small to small	1,435	96.8	544.8	537.5
Small to regular	23	1.6	546.8	549.0
Small to regular aide	24	1.6	527.5	506.7
Total	1,482			
From regular to first-grade class size				
Regular to small	152	8.2	529.1	523.8
Regular to regular	1,498	80.9	534.3	524.5
Regular to regular/aide	202	10.9	525.0	506.6
Total	1,852			
From regular/aide first-grade class size				
Regular/aide to small	40	2.3	521.7	510.8
Regular/aide to regular	115	6.7	526.3	503.0
Regular/aide to regular/aide	1,560	91.0	536.1	531.1
Total	1,715			

lower reading achievement test scores than those making the reverse transition. Therefore, it would appear that any bias due to these exchanges in the estimate of the small class effect would tend to decrease the small class effect at the second grade.

Table 7 presents the mean mathematics and reading achievement test scores of the students making each possible transition between class types from second grade to third grade. The table shows that the students who continue in small classes have substantially higher average achievement test scores than those who continue in either type of regular-sized classes. It also shows that the mean achievement test scores in both reading and mathematics of the students who exchange from small classes to regular-sized classes are substantially lower than those for students who make the reverse transition (from regular-sized classes to small classes). The students who make the transition from small classes to regular classes with aides also have lower mean achievement test scores in both mathematics and reading than those making the reverse transition. Therefore, it would appear that any bias in the estimate of the small class effect due to these exchanges would tend to increase the small class effect at the third grade.

Therefore, the direction of biases in the small class effect due to shifting between treatment types would seem to vary across grades. The small class effect may be overestimated in first and third grades, but may be underestimated in second grade. However, the total number of students switching between small classes and either type of regular-sized classes is relatively small in proportion to the total sample size. In addition, it is evident that most of the switching was *into*, rather than *out of*, the small classes. This certainly

Table 7
Student Transitions From Second to Third Grade

Type of class transition	*N*	Percent	Average math score	Average reading score
From small second-grade class size				
Small to small	1,564	95.6	590.7	594.7
Small to regular	37	2.3	563.9	571.7
Small to regular/aide	35	2.1	562.7	572.8
Total	1,636			
From regular second-grade class size				
Regular to small	167	9.3	580.9	582.9
Regular to regular	1,485	82.3	583.0	585.4
Regular to regular/aide	152	8.43	562.1	562.7
Total	1,804			
From regular/aide second-grade class size				
Regular/aide to small	40	2.0	576.7	580.6
Regular/aide to regular	76	3.9	552.7	558.0
Regular/aide to regular/aide	1,857	94.1	583.2	587.7
Total	1,973			

suggests that these transitions did not occur at random and were probably the result of a belief on the part of teachers or parents that small classes were more desirable.

However, to further examine the possible effects on conclusions about transitions between small and regular-sized classes, we analyzed the effects of treatment using the treatment originally assigned (even if it was not the treatment actually received). This is the equivalent of the "intention to treat" analysis often used in clinical trials in medicine. The notion is to count, for the purposes of analysis, any individuals randomized to receive a treatment as if they received it. The intention to treat analysis of a randomized experiment will produce a conservative estimate (i.e., an underestimate) of the treatment effect as long as the treatment is not actually harmful. Although some might contend that class size is unrelated to academic achievement, we are not aware of any evidence or argument that small classes *decrease* academic achievement. Consequently, the analysis of the STAR data using treatments "as-assigned" should provide a conservative estimate of the small class effect.

The as-assigned analysis of the STAR data was conducted using exactly the same procedures and hierarchical linear models as the primary analyses using the treatments as received.

Methods

The design of the STAR experiment involves the random assignment of students and teachers to treatments within schools. Conceptually, the study is a series of within-school experiments conducted using the same procedures and outcome variables. Because the variance in student achievement within schools is typically different than the variance between schools, the sampling design involves a clustering or hierarchical structure that should be taken into account in the analysis. One such analysis is the use of hierarchical linear models (Bryk & Raudenbush, 1992). Such models permit the analysis and pooling of school-specific regressions (including treatment effects) in a manner that takes the clustering of the sample by school.

The within-school model used in our primary analysis treated student achievement as a function of student characteristics (gender and social class as operationalized by free lunch eligibility), treatment group assignment, and the interaction of assignment with gender. This specification was chosen because of the widely known relation between socioeconomic status (SES) and school performance (White, 1982), the fact that gender is also related to school performance (Hedges & Nowell, 1995), and the fact that classroom process in small classes may differentially affect the genders (Brophy & Good, 1974). The specific model for achievement test score Y_{ij} of the ith student in the jth school was

$$Y_{ij} = \beta_{0j} + \beta_{1j}FEMALE_{ij} + \beta_{2j}SES_{ij} + B_{3j}SMALL_{ij} + \beta_{4j}AIDE_{ij} + \beta_{5j}FS_{ij} + \epsilon_{ij},$$

where $FEMALE_{ij}$ is a dummy variable for gender; SES_{ij} is a dummy variable for free lunch eligibility; $SMALL_{ij}$ and $AIDE_{ij}$ are indicator variables for small class size and having a full-time classroom aide, respectively; FS_{ij} is the interaction of FEMALE and SMALL; and ϵ_{ij} is a student-specific sampling error.

We modeled variation across schools in each of the school-specific regression coefficients according to the geographic location of the school, the mean level of teacher experience in the school, the SES level of the school (as the proportion of all students receiving free lunches), and the percentage of Black students in the school. This specification was chosen because it was similar to that used in many school effect studies (Lee & Bryk, 1989). The specific Level 2 model for the mth coefficient in the jth school, β_{mj}, was therefore:

$$\beta_{mj} = \gamma_{0m} + \gamma_{1m}INNER_j + \gamma_{2j}RURAL_j + \gamma_{3j}URBAN_j + \gamma_{4j}TEACHERED_j + \gamma_{5j}TEACHEREXP_j + \gamma_{6j}SCHOOLSES_j + \gamma_{7j}MINORITY_j + \eta_{mj},$$

where $INNER_j$, $RURAL_j$, and $URBAN_j$ are indicators of the geographic location of the school; $TEACHERED_j$ is the percentage of the teachers in the school with an advanced degree; $TEACHEREXP_j$ is the average number of years of teacher experience in the school; $SCHOOLSES_j$ is the percentage of students receiving free or reduced price lunches; $MINORITY_j$ is the proportion of students in the school who are Black; and η_{mj} is a Level 2 residual (random effect). Therefore, the object of the statistical analysis is to estimate the eight fixed effects (the γ_{mj}) determining each of the six β_{ij}s and the corresponding between-school variance components (variances of the η_{mj}s).

Although the two-level analysis is an improvement on the analyses that ignore within-school clustering altogether, another type of clustering also exists within schools: clustering by classroom. That is, even within the same school, students in the same classroom may exhibit less achievement variance than students in different classrooms, even if the classrooms are assigned to the same type of treatment. This clustering is associated with the effects of teachers and classrooms independent of any treatment assignments. For example, particular teachers may be more compensatory or more meritocratic in their instructional efforts, leading to either reduction or expansion of the variation in achievement in their classrooms (or they may be more or less successful in their attempts to use the same strategy). These differences will cause classroom variation in achievement net of any differences in treatment assignment.

The appropriate analysis to recognize the between-class clustering would assign a component of variance to differences between classes receiving the same type of treatment. Unfortunately, there are too few classrooms within schools receiving the same treatment to carry out this analysis when all three treatment types are distinguished. However, with the possible exception of the effect in inner city kindergartens, neither the main effect (the intercept) nor any of the interactions with (predictors of) the aide effect are statistically significant. Similarly, other analyses of the STAR data (using the conventional statistical methods that probably overstated the statistical significance of effects) failed to find main effects or interactions with the aide effect. Such analyses led other investigators to conclude that there was no effect of full-time classroom aides and to decide to pool the two types of regular-sized classes in their analyses.

If we also make the assumption that there is no effect from classroom aides, then there are at least two classes of the same type (regular classes) in each school, and it is possible to estimate the between-classroom/within-school variance to obtain a more accurate representation of how the effects of small classes might vary across schools. Such a model would technically be a three-level hierarchical linear model. The Level 1 model would be the same as the two-level model, but we permit the small class effect (or more precisely, the intercept β_{0j} of the Level 1 model) to vary across classes within schools. This approach permits the estimation of the between-school variation of the small class effect net of the usual variation in achievement test scores between classes within schools.

The specific model for achievement test score Y_{ij} of the ith student in the jth class of the kth school (the Level 1 model) was

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}FEMALE_{ijk} + \beta_{2jk}SES_{ijk} + \beta_{3jk}SMALL_{ijk} + \beta_{4jk}FS_{ijk} + \epsilon_{ijk},$$

where $FEMALE_{ijk}$ is a dummy variable for gender; SES_{ijk} is a dummy variable for free lunch eligibility; $SMALL_{ijk}$ is an indicator variable for small class size; FS_{ijk} is the interaction of FEMALE and SMALL; and ϵ_{ijk} is a student-specific sampling error.

The specific model for variation of coefficients between classes within schools (the Level 2 model) was

$$\beta_{0jk} = \pi_{00k} + \xi_{0jk},$$

where π_{00k} is a school-specific intercept and ξ_{0jk} is a classroom-specific random effect. Thus, the variance of the ξ_{0jk} provides a description of the variation of average achievement test scores across classes net of the effects of student gender, SES, and treatment assignment. All other coefficients were constrained to be constant within schools, that is, $\beta_{mjk} = \beta_{mjk}$ for $m > 0$.

We modeled variation across schools of each of the school-specific regression coefficients according to the geographic location of the school, the mean level of teacher experience in the school, the SES level of the school (the proportion of all students receiving free lunches), and the percentage of minority students in the school. The specific Level 3 model for the mth coefficient in the jth class of the kth school β_{mj} was therefore:

$$\pi_{mjk} = \gamma_{0m} + \gamma_{1m}\text{INNER}_k + \gamma_{2m}\text{RURAL}_k + \gamma_{3m}\text{URBAN}_k + \gamma_{4m}\text{TEACHERED}_k + \gamma_{5m}\text{TEACHEREXP}_k + \gamma_{6m}\text{SCHOOLSES}_k + \gamma_{7m}\text{MINORITY}_k + \eta_{mk},$$

where INNER_k, RURAL_k, and URBAN_k are indicators of the geographic location of the school; TEACHERED_k is the percentage of the teachers in the school with an advanced degree; TEACHEREXP_k is the average total number of years of teacher experience in the school; SCHOOLSES_k is the percentage of students receiving free or reduced price lunches; MINORITY_k is the proportion of students in the school who are Black; and η_{mk} is a Level 3 residual (random) effect. Therefore, the object of the statistical analysis is to estimate the eight fixed effects (the γ_{mk}) determining each of the five π_{ij}s (and, therefore, β_{ijk}s), the between-classes/within-schools variance component of the β_{0jk}s, and the corresponding between-school variance components of the η_{mj}s.

In each case, we conducted separate analyses for each of the two dependent variables (SAT mathematics and reading test scores) for each of the four grade levels (kindergarten through third grade); each analysis was repeated eight times.

We have chosen to model treatment effects as categorical (small or regular-sized classes) rather than to estimate an effect of the numerical value of class size. We did so because the randomization was carried out across values of the categorical variable, not the numerical values of class size. Therefore, in principle, randomization guarantees freedom from bias in estimates of the effects associated with the categorical values of the treatment effects; it would not do so for estimates of the effects of numerical values of class size.

Results and Discussion

The results of the each of the different types of analyses are reported in the subsections of this section, along with some interpretive comments.

Table 8

Fixed Effects and Variance Components Estimates: Two-Level Model Using Actual Assignment

	Math				Reading			
	K	1	2	3	K	1	2	3
Intercept (β_{0j})								
Intercept (γ_{00})	−0.016	0.003	0.015	0.008	−0.029	0.019	−0.006	0.008
Inner city school	−0.318	−0.076	−0.089	0.137	−0.103	0.007	−0.190	0.040
Rural school	−0.025	0.189	0.093	0.088	−0.059	0.105	0.044	0.008
Urban school	−0.066	0.112	−0.273	−0.192	0.022	0.092	−0.084	−0.205
Percent of teachers in regular classes having an advanced Ed degree	−0.075	0.200	0.088	0.166*	−0.028	0.078	0.000	0.082
Teachers' total experience in regular classes	0.006	−0.005	−0.001	0.002	0.012	0.003	−0.001	−0.000
Percent of students in regular classes receiving free or reduced lunch	0.264	0.015	0.396	−0.032	0.187	−0.165	0.212	0.085
Percent of Black students in regular classes	−0.074	−0.237	−0.591*	−0.524*	−0.234	−0.384	−0.563*	−0.523*
Residual variance	0.196*	0.145*	0.121*	0.104*	0.182*	0.109	0.067*	0.057*
Female (β_{1j})								
Intercept (γ_{01})	0.150*	0.019	0.014	0.024	0.172*	0.184*	0.214*	0.196*
Inner city school	0.062	−0.089	−0.049	−0.071	−0.027	−0.065	0.023	−0.049
Rural school	0.156	0.040	−0.122	0.058	0.122	0.168	0.021	0.078
Urban school	0.108	0.065	0.030	0.164	0.208	0.224	−0.005	0.211
Percent of teachers having an advanced Ed degree	0.037	0.002	−0.092	−0.191	0.018	−0.082	−0.082	−0.164
Teachers' total experience	−0.004	0.003	0.010	0.004	−0.019	0.001	0.004	0.000
Percent of students receiving free or reduced lunch	−0.368	−0.140	0.274	0.096	−0.295	−0.287	0.146	−0.099
Percent of Black students	0.296	0.238	−0.223	0.077	0.206	0.263	−0.170	0.168
Residual variance	0.004	0.007	0.021	0.033*	0.004	0.009	0.006	0.022

Low SES (β_{2j})								
Intercept (γ_{02})	−0.434*	−0.445*	−0.475*	−0.428*	−0.473*	−0.493*	−0.497*	−0.461*
Inner city school	0.013	0.247	−0.085	−0.058	0.039	−0.032	0.004	−0.177
Rural school	−0.001	0.035	0.046	0.092	0.055	−0.089	0.103	0.056
Urban school	−0.032	−0.044	−0.048	0.371*	0.005	−0.039	0.212	0.156
Percent of teachers having an advanced Ed degree	0.075	−0.122	0.189	−0.105	0.097	0.005	0.117	0.037
Teachers' total experience	−0.011	−0.004	−0.015*	−0.011	0.003	−0.005	−0.008	−0.015
Percent of Black students	0.169	−0.054	0.183	0.278	0.256	0.092	0.342*	0.325
Residual variance	0.022	0.064*	0.011	0.023*	0.021*	0.071	0.020*	0.033*
Small class (β_{3j})								
Intercept (γ_{03})	0.225*	0.303*	0.181*	0.150*	0.231*	0.206*	0.226*	0.175*
Inner city school	0.263	−0.173	−0.170	0.441	−0.018	−0.223	0.128	0.388
Rural school	0.271	−0.092	−0.252	−0.016	0.034	0.153	0.024	−0.011
Urban school	0.133	−0.077	−0.622*	−0.071	0.057	0.104	−0.312	0.043
Percent of teachers in small classes having an advanced Ed degree	−0.041	0.121	−0.034	0.052	−0.092	−0.020	−0.044	0.015
Teachers' total experience in small classes	0.014	−0.005	0.010	0.010	0.006	−0.003	0.013*	0.017*
Percent of students in small classes receiving free or reduced lunch	−0.400	−0.356	0.081	−0.845*	−0.172	−0.372	0.251	−0.419
Percent of Black students in small classes	0.247	0.416	0.011	0.143	0.270	0.479	−0.065	−0.007
Residual variance	0.160*	0.187*	0.209*	0.177*	0.126*	0.101*	0.144*	0.113*
Regular/aide class (β_{4j})								
Intercept (γ_{04})	0.005	0.035	0.034	−0.011	0.035	0.080*	0.063	0.005
Inner city school	0.542*	−0.034	0.196	0.270	0.489*	−0.011	0.296	0.211
Rural school	0.067	0.207	−0.053	0.028	0.013	0.159	0.073	−0.082
Urban school	0.111	0.073	−0.192	−0.048	0.024	0.118	−0.171	−0.286

Table 8 (Continued)

	Math				Reading			
	K	1	2	3	K	1	2	3
Percent of teachers in regular/aide classes having an advanced Ed degree	0.018	−0.126	−0.241*	0.002	0.055	−0.006	−0.122	0.096
Teachers' total experience in regular/aide classes	0.005	−0.001	−0.002	0.007	0.010	0.002	0.004	0.005
Percent of students in regular/aide classes receiving free or reduced lunch	−0.353	0.184	0.183	−0.298	−0.140	0.180	0.115	−0.010
Percent of Black students in regular/aide classes	−0.161	0.156	−0.356	−0.225	−0.104	−0.034	−0.257	−0.246
Residual variance	0.166*	0.110*	0.157*	0.190*	0.109*	0.056*	0.126*	0.110*
Female-Black interaction (β_{5j})								
Intercept (γ_{05})	−0.105*	−0.046	0.063	−0.026	−0.048	0.072	0.006	0.018
Inner city school	0.084	0.008	0.363	−0.343	0.016	0.178	0.376	−0.318
Rural school	−0.084	0.110	0.372*	0.029	0.002	−0.056	0.225	0.043
Urban school	0.069	0.190	0.491	0.065	−0.021	0.009	0.343	0.026
Percent of teachers having an advanced Ed degree	−0.104	−0.180	−0.195	0.088	−0.238	0.057	−0.155	0.075
Teachers' total experience	0.004	−0.001	−0.018	−0.001	0.011	−0.005	−0.010	0.005
Percent of students receiving free or reduced lunch	0.507	0.343	−0.594	0.791*	0.571	0.146	−0.733*	0.631
Percent of Black students	−0.203	−0.165	0.283	−0.133	−0.188	−0.146	0.208	−0.046
Residual variance	0.024	0.004	0.022	0.065*	0.057	0.012	0.012	0.042

Note. SES = socioeconomic status. Boldface headings denote coefficients in the level 1 model.
*$p < .05$.

Analyses of the Effects of Treatments as Received

The most conventional analysis of an experiment is an analysis of the treatments as received by the students. In this analysis, the students who may have been initially assigned to receive a different treatment than the one they actually received are treated as if they were assigned to the treatment they received. Later, we discuss the issue of students who switched from one treatment to another and therefore received treatments other than those to which they were initially assigned. The results of the two-level analyses (where treatment assignment is the treatment actually received) are summarized in Table 8.

These results demonstrate that, across all schools, females have significantly higher achievement test scores in reading (at all grades) and mathematics at kindergarten. Students with low SES have significantly lower achievement test scores in reading and mathematics at all grade levels, although the significant variance component indicates that the size of the SES effect differs across schools. Perhaps more important for this experiment is that the average effect of small classes is significant and positive in both mathematics and reading at every grade level, ranging from 0.15 to 0.30 *SD* units. The average effects of having a full-time classroom aide and the female/small class interaction are generally not significant.

There is evidence that both the mean level of achievement (adjusted for SES) and the effect of SES vary across schools, as evidenced by the statistically significant variance components for these effects. There seems to be relatively little variation across schools in the effects of gender and the interaction of gender and small class size, as evidenced by the fact that this variance component is statistically significant in only one of eight analyses (for mathematics achievement test scores at grade 3). The variation of both the small class effect and the effect of full-time aides also varies across schools; however, variation of these effects is confounded with ordinary between-classroom/within-school variation, so these variance estimates are difficult to interpret (the three-level analysis that follows helps to clarify this issue).

Few of the predictors in the between-school model explain much of the variance in model coefficients. For example, none of the school-level predictors explain as much as 10% of the between-school variation in small class effects in any grade or subject matter. The only potential exception is the fact that full-time instructional aides have a significant effect on both mathematics and reading in inner city schools at kindergarten. This may suggest a positive effect of full-time aides in facilitating communication skills and school readiness among the most disadvantaged kindergartners.

Using related analytic models, other hierarchical linear model analyses were carried out to eliminate all of the predictors that were not statistically significant and to examine the slightly different codings of the variables. The results of these analyses were not qualitatively different from those reported here. In each case, the small class effect had about the same positive impact

Table 9

Fixed Effects and Variance Components Estimates: Three-Level Model Using Actual Assignment

	Math				Reading			
	K	1	2	3	K	1	2	3
Intercept (β_{0j})								
Intercept (γ_{00})	−0.012	0.000	0.007	0.003	−0.021	0.013	−0.011	0.002
Inner city school	−0.404	0.041	0.046	0.153	−0.137	−0.001	−0.146	0.055
Rural school	0.018	0.165	0.097	0.028	0.013	0.089	0.019	0.005
Urban school	−0.006	0.077	−0.237	−0.284	0.091	0.076	−0.101	−0.239
Percent of teachers in regular classes having an advanced Ed degree	−0.053	0.086	−0.133	0.247*	−0.053	0.052	−0.052	0.133
Teachers' total experience in regular classes	0.039*	−0.009	−0.001	0.001	0.043*	0.004	−0.002	0.001
Percent of students in regular classes receiving free or reduced lunch	0.193	−0.227	0.352	−0.016	0.005	−0.255	0.283	0.079
Percent of Black students in regular classes	0.164	−0.229	−0.699*	−0.571*	0.037	−0.344	−0.686*	−0.526*
Residual variance	0.131*	0.105*	0.083*	0.065*	0.125*	0.082*	0.041*	0.033*
Female (β_{1j})								
Intercept (γ_{01})	0.150*	0.016	0.014	0.028	0.172*	0.181*	0.210*	0.200*
Inner city school	0.056	−0.070	−0.052	−0.091	−0.007	−0.070	0.022	−0.085
Rural school	0.155	0.042	−0.131	0.059	0.130	0.164	0.012	0.030
Urban school	0.098	0.077	0.043	0.171	0.218	0.221	−0.017	0.165
Percent of teachers having an advanced Ed degree	0.083	−0.028	−0.113	−0.166	0.046	−0.070	−0.072	−0.155
Teachers' total experience	−0.005	0.003	0.009	0.001	−0.019	0.000	0.002	−0.001
Percent of students receiving free or reduced lunch	−0.368	−0.176	0.312	0.017	−0.322	−0.272	0.155	−0.073
Percent of Black students	0.281	0.245	−0.256	0.150	0.203	0.263	−0.197	0.139
Residual variance	0.003	0.004	0.003	0.027*	0.004	0.004	0.002	0.015*

Low SES (β_{2j})								
Intercept (γ_{02})	−0.440*	−0.447*	−0.462*	−0.427*	−0.476*	−0.488*	−0.492*	−0.462*
Inner city school	0.002	0.251	−0.057	−0.109	−0.028	−0.032	0.048	−0.182
Rural school	−0.001	0.008	0.042	0.138	0.038	−0.106	0.101	0.133
Urban school	−0.009	−0.054	−0.011	0.397*	0.009	−0.054	0.226	0.205
Percent of teachers having an advanced Ed degree	0.030	−0.107	0.158	−0.112	0.104	0.018	0.113	0.036
Teachers' total experience	−0.012	−0.001	−0.011	−0.009	−0.004	−0.005	−0.006	−0.015
Percent of Black students	0.135	−0.103	0.152	0.361*	0.233	0.074	0.305	0.407*
Residual variance	0.003	0.056*	0.007	0.014*	0.021	0.059*	0.011*	0.018*
Small class (β_{3j})								
Intercept (γ_{03})	0.215*	0.281*	0.162*	0.150*	0.202*	0.169*	0.185*	0.173*
Inner city school	0.006	−0.083	−0.195	0.259	−0.249	−0.224	0.008	0.228
Rural school	0.255	−0.214	−0.349*	0.010	−0.015	0.059	−0.095	0.044
Urban school	0.130	−0.096	−0.571*	−0.040	0.046	0.042	−0.311	0.176
Percent of teachers in small classes having an advanced Ed degree	−0.027	0.130	−0.032	0.004	−0.119	−0.023	−0.055	−0.013
Teachers' total experience in small classes	0.016	−0.007	0.012	0.010	0.009	0.002	0.015*	0.016*
Percent of students in small classes receiving free or reduced lunch	−0.343	−0.342	0.181	−0.758*	−0.122	−0.418	0.316	−0.379
Percent of Black students in small classes	0.406	0.250	−0.077	0.342	0.296	0.496	−0.106	0.151
Residual variance	0.002	0.040*	0.072*	0.024*	0.019*	0.020	0.030	0.016
Female-Black interaction (β_{4j})								
Intercept (γ_{04})	−0.105*	−0.034	0.054	−0.028	−0.052	0.078	0.006	0.012
Inner city school	0.135	−0.036	0.321	−0.273	0.004	0.186	0.356	−0.224
Rural school	−0.140	0.088	0.378*	0.036	−0.051	−0.068	0.238	0.138
Urban school	0.060	0.165	0.505*	0.072	−0.054	0.007	0.398	0.132
Percent of teachers having an advanced Ed degree	−0.109	−0.138	−0.088	0.027	−0.235	0.061	−0.101	0.008
Teachers' total experience	−0.007	0.001	−0.018	0.006	0.002	−0.005	−0.012	0.008
Percent of students receiving free or reduced lunch	0.541	0.323	−0.596	0.936*	0.582	0.160	−0.750*	0.512
Percent of Black students	−0.321	−0.137	0.353	−0.286	−0.283	−0.173	0.263	0.022
Residual variance	0.002	0.017	0.023	0.045*	0.026	0.013	0.007	0.022

Note. SES = socioeconomic status. Boldface headings denote coefficients in the level 1 model.
*$p < .05$.

and was statistically significant. There was no consistent relation between the small class effects and any of the school-level predictors.

In order to clarify the variation in small class effects, we analyzed the data using a three-level model, which included a random effect for between-class/within-school differences. The results of this analysis are summarized in Table 9. The overall pattern of results from the three-level analysis is virtually identical to that of the two-level analysis. The only substantial difference is that the between-school variance (component) of the small class effect is negligible and statistically insignificant in six of the eight analyses. This suggests that the small class effect is remarkably consistent across schools.

Analysis of the Effects of Treatments as Assigned

In order to investigate the potential effects of students who switched from the treatment in which they were originally assigned to a different treatment, we conducted an analysis of the effects of treatments as they were initially assigned (regardless of the treatment received). The results of a three-level version of these analyses are summarized in Table 10. Comparing Tables 9 and 10, it is clear that the same pattern of results emerges regardless of whether initial assignment or actual classroom type is used. Specifically, although the average small class effects are slightly smaller in the analysis using initial assignment, the small class effects are all statistically significant and of the same qualitative magnitude as in the analysis by actual classroom type. As in the previous analysis, there is essentially no variation in the small class effect across schools.

Analyses of Effects by Number of Years in Small Classes

In order to address the question of whether the effects of small classes are cumulative, we carried out analyses to determine whether the small class effect is greater for students who had more years in small classes. Table 11 shows the effects of participating in 1–4 years of small classes from the three-level analyses. (Note that it is only possible to have 2 or more years of small classes in Grade one or higher, 3 or more years of small classes in Grade two or higher, and 4 years in small classes in Grade three.) All of the small class effects are statistically significant. At every grade, the effects of small classes are greater for more years spent in small classes.

Conclusions

The analyses reported here support the validity of the conclusion that small classes in the early grades lead to higher academic achievement. The class sizes that were intended in the design of the study were not always attained, usually because the larger classes were smaller than anticipated. Such overlap in the class sizes of nominally larger and small classes would tend to reduce the size of the small class effect. Project STAR was also subject to substantial attrition. The treatment effect in the year previous to dropping out, however, was the same for students who dropped out later and those who did not, suggesting that the observed differences in achievement between small and larger classes were not due to differential attrition. The effect of switching between classes is more difficult to determine qualitatively. An analysis estimating the small class effect using the initial assignment of students resulted in estimates of small class effects that were almost identical to those obtained using the actual class type.

Statistical analyses taking into account the clustering of students within schools and classrooms provide evidence for a positive effect of small classes in mathematics and reading achievement test scores at every grade level from kindergarten through Grade three. These small class effects are remarkably consistent across schools, suggesting that small classes benefit students of all types in all kinds of schools.

Small class effects are also greater for students who have experienced more years in small classes. If the small class effects were a consequence of failure of initial randomization, so that students with higher initial achievement were assigned to small classes, it would be difficult to explain this effect. However, the interpretation of this effect is somewhat ambiguous because students were not randomly assigned to differing numbers of years in small classes.

It is interesting that the small class effects found in this analysis of the STAR data are at least qualitatively consistent with those that would have been predicted from meta-analyses of small-scale experiments on the effects of class size. Applying the results reported for randomized experiments in the Glass and Smith (1979) meta-analysis of research on the relationship between class size and achievement, a reduction of class size from 25 to 15 students would have been expected to yield an increase in achievement of .215 *SD*, in the center of the range of .15–.30 *SD* obtained in the STAR experiment.

The small class effects found in this analysis of the STAR experiment are also roughly consistent with those found in the meta-analysis of the econometric studies carried out by Greenwald et al. (1996). Using the median standardized regression coefficient for post-1970 studies in their Table 5 ($\beta = .047$) and the estimate that a *SD* of class size is about 3 (see their Table 7), reduction of class size from 25 to 15 would have been expected to increase achievement by .157 *SD*, at the lower end of the small class effects obtained in this analysis of the STAR data.

The STAR study provides an important (and perhaps the strongest) piece of the converging evidence about the effectiveness of small classes in promoting achievement. Problems in the implementation of the experiment probably led to underestimates of the effects of small classes. However, the magnitude of effects in this study is quite consistent with that obtained in small-scale randomized experiments (whose generalizability might be questioned) and with the results of econometric studies (whose internal validity

Table 10

Fixed Effects and Variance Components Estimates: Three-Level Model Using Initial Assignment

	Math				Reading			
	K	1	2	3	K	1	2	3
Intercept (β_{0j})								
Intercept (γ_{00})	−0.012	0.007	0.007	0.004	−0.021	0.015	−0.010	−0.000
Inner city school	−0.404	0.054	0.029	0.143	−0.137	−0.011	−0.156	0.034
Rural school	0.018	0.125	0.071	0.025	0.013	0.066	0.002	−0.000
Urban school	−0.006	0.043	−0.288	−0.302	0.091	0.056	−0.136	−0.250
Percent of teachers in regular classes having an advanced Ed degree	−0.053	0.049	−0.132	0.301*	−0.053	0.056	−0.040	0.167
Teachers' total experience in regular classes	0.039*	−0.009	0.004	−0.000	0.043*	0.005	−0.001	−0.000
Percent of students in regular classes receiving free or reduced lunch	0.193	−0.203	0.380	−0.015	0.005	−0.204	0.307	0.096
Percent of Black students in regular classes	0.164	−0.297	−0.719*	−0.566*	0.037	−0.380	−0.705*	−0.531*
Residual variance	0.131*	0.099*	0.075*	0.061*	0.125*	0.078*	0.037*	0.031*
Female (β_{1j})								
Intercept (γ_{01})	0.150*	0.021	0.022	0.047	0.172*	0.184*	0.218*	0.210*
Inner city school	0.056	−0.097	−0.001	−0.142	−0.007	−0.070	0.031	−0.102
Rural school	0.155	0.028	−0.103	−0.013	0.130	0.153	0.023	−0.017
Urban school	0.098	0.040	0.037	0.108	0.218	0.196	−0.037	0.099
Percent of teachers having an advanced Ed degree	0.083	−.054	−0.199	−0.132	0.046	−0.095	−0.137	−0.114
Teachers' total experience	−0.005	0.003	0.010	−0.001	−0.019	0.002	0.003	−0.002
Percent of students receiving free or reduced lunch	−0.368	−0.129	0.256	0.172	−0.322	−0.227	0.127	0.045
Percent of Black students	0.281	0.210	−0.269	0.056	0.203	0.204	−0.205	0.057
Residual variance	0.003	0.003	0.015*	0.028*	0.004	0.008	0.009	0.016*

Low SES (β_{2j})								
Intercept (γ_{02})	−0.440*	−0.452*	−0.462*	−0.427*	−0.476*	−0.492*	−0.492*	−0.462*
Inner city school	0.002	0.231	−0.068	−0.091	−0.028	−0.046	0.036	−0.162
Rural school	−0.001	0.007	0.055	0.133	0.038	−0.107	0.097	0.137
Urban school	−0.009	−0.059	−0.007	0.375*	0.009	−0.064	0.213	0.198
Percent of teachers having an advanced Ed degree	0.030	−0.111	0.169	−0.120	0.104	0.023	0.137	0.033
Teachers' total experience	−0.012	−0.001	−0.012	−0.007	−0.004	−0.006	−0.007	−0.014
Percent of Black students	0.135	−0.100	0.175	0.328*	0.233	0.078	0.310*	0.402*
Residual variance	0.003	0.058*	0.006	0.014*	0.021	0.059*	0.010*	0.015*
Small class (β_{3j})								
Intercept (γ_{03})	0.215*	0.199*	0.121*	0.141*	0.202*	0.123*	0.175*	0.154*
Inner city school	0.006	−0.048	−0.091	0.284	−0.249	−0.202	−0.226	0.262
Rural school	0.255	−0.100	−0.151	−0.110	−0.015	0.088	0.039	−0.013
Urban school	0.130	−0.039	−0.268	0.150	0.046	0.094	−0.111	0.202
Percent of teachers in small classes having an advanced Ed degree	−0.027	0.157	−0.142	−0.053	−0.119	−0.040	−0.246*	−0.069
Teachers' total experience in small classes	0.016	−0.004	0.010	0.012	0.009	0.003	0.008	0.015*
Percent of students in small classes receiving free or reduced lunch	−0.343	−0.019	−0.084	−0.497	−0.122	−0.267	0.110	−0.311
Percent of Black students in small classes	0.406	0.048	0.089	0.070	0.296	0.329	0.199	0.137
Residual variance	0.002	0.031*	0.040*	0.009	0.019	0.008	0.007	0.003
Female-Black interaction (β_{4j})								
Intercept (γ_{04})	−0.105*	−0.053	0.007	−0.107	−0.052	0.072	−0.045	−0.027
Inner city school	0.135	0.090	0.214	−0.165	0.004	0.247	0.440	−0.252
Rural school	−0.140	0.150	0.347*	0.320	−0.051	−0.062	0.255	0.319
Urban school	0.060	0.331	0.588*	0.285	−0.054	0.080	0.549*	0.349
Percent of teachers having an advanced Ed degree	−0.109	0.029	0.231	−0.018	−0.235	0.157	0.166	−0.105
Teachers' total experience	−0.007	0.000	−0.020	0.009	0.002	−0.011	−0.017	0.008
Percent of students receiving free or reduced lunch	0.541	0.183	−0.447	0.451	0.582	0.034	−0.751*	0.175
Percent of Black students	−0.321	−0.059	0.380	0.033	−0.283	−0.083	0.275	0.291
Residual variance	0.002	0.009	0.016	0.047*	0.026	0.030	0.013	0.006

Note. SES = socioeconomic status. Boldface headings denote coefficients in the level 1 model.
*$p < .05$.

Table 11
Average Cumulative Effects of Small Class Assignment Across Grades 1–3

	Grade 1		Grade 2		Grade 3	
	Math	Reading	Math	Reading	Math	Reading
Small class in any grade (β_{3j})						
Intercept (γ_{03})	0.280*	0.150*	0.151*	0.176*	0.151*	0.163*
Small class in two or more grades (β_{3j})						
Intercept (γ_{03})	0.325*	0.245*	0.213*	0.240*	0.192*	0.239*
Small class in three or more grades (β_{3j})						
Intercept (γ_{03})			0.301*	0.336*	0.256*	0.270*
Small class in all grades (β_{3j})						
Intercept (γ_{03})					0.352*	0.386*

*$p < .05$.

might be questioned). Together, all of this evidence points to positive effects of small classes on achievement that are large enough to be educationally significant.

Project STAR has not answered all the important questions about the effects of small classes on achievement test scores and other desirable outcomes of schooling. One important issue that was not addressed in the STAR experiment is the cost effectiveness of small classes. In an environment of scarce resources, cost must be a consideration in deciding on policies. More information on cost effectiveness is clearly needed in conjunction with the effects of small classes and alternative policies to increase academic achievement.

It is not yet clear *how* small classes lead to higher achievement. Understanding the mechanism could lead to more effective ways to implement class size reductions and to improve their effectiveness. Such understanding is obviously desirable. It is also unclear whether the improvements in achievement produced in small classes will persist throughout schooling and, if so, how they will make themselves apparent. For example, will students from small classes simply persist at a consistently higher level of achievement, or will the advantage of small classes lead to changes in growth trajectory over time? Additional analyses of the data from the Lasting Benefits Study (Nye et al., 1994), the long term follow-up to the STAR project, may provide answers to these questions.

References

Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences.* New York: Holt, Rinehart, and Winston.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Cook, T. D., & Campbell D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Educational Research Services. (1980). *Class size research: A critique of recent meta-analysis*. Arlington, VA: Author.

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., Nye, B. A., & Zaharias, J. B. (1992). Carry-over effects of small classes. *Peabody Education Journal, 67*, 75–84.

Glass, G. V, Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School class size: Research and policy*. Beverly Hills: Sage.

Glass, G. V., & Smith, M. E. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1*, 2–16.

Greenwald, R., Hedges, L. V., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 361–396.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *The Journal of Economic Literature, 24*, 1141–1177.

Hanushek, E. A. (1989). The impact of differential school expenditures on school performance. *Educational Researcher, 18*, 45–65.

Hanushek, E. A. (1999). *Some findings from the Tennessee STAR experiment and other investigations of class size reductions*. Rochester, NY: Wallis Institute of Political Economy, Department of Economics, University of Rochester.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*, 5–14.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high scoring individuals. *Science, 269*, 41–45.

Hedges, L. V., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal, 20*, 63–85.

Krueger, A. B. (1997). *Experimental estimates of education production functions*. Unpublished manuscript, Princeton University, NJ.

Lee, V., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62*,172–192.

Mosteller, F., Light, R. J., & Sachs, J. A. (1996). Sustained inquiry in education: Lessons learned from skill grouping and class size. *Harvard Educational Review, 66*, 797–842.

Nye, B. A., Zacharias, J. B., Fulton, B. D., Achilles, C. M., Cain, V. A., & Tollett, D. A. (1994). *The lasting benefits study: A continuing analysis of the effect of small class size in kindergarten through third grade on student achievement test scores in subsequent grade levels. Seventh grade* (Tech. Rep.). Nashville: Tennessee State University, Center of Excellence for Research in Basic Skills.

Slavin, R. (1984). Meta-analysis in education: How has it been used? *Educational Researcher, 13*, 6–15, 24–25.

White, C. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461–481.

Word, E., Johnston, J., Bain, H., 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: Tennessee Department of Education.

Manuscript received April 1998
Revision received February 4, 1999
Accepted August 26, 1999