The Journal of Experimental Education, 2001, 69(3), 245–257

The Long-Term Effects of Small Classes in Early Grades: Lasting Benefits in Mathematics Achievement at Grade 9

BARBARA NYE
Tennessee State University

LARRY V. HEDGES
SPYROS KONSTANTOPOULOS
The University of Chicago

ABSTRACT. Reducing class size to increase academic achievement is a policy option currently of great interest. Although the results of small-scale randomized experiments and some interpretations of large-scale econometric studies point to positive short-term effects of small classes, some scholars view the evidence as ambiguous. Project STAR in Tennessee—a 4-year, large-scale randomized experiment on the effects of class size—provided persuasive evidence that small classes have immediate positive effects on academic achievement. Unlike most other early education interventions, these effects persisted for several years after the children returned to regular-sized classes. The authors of the present article report analyses of a 6-year follow-up of the students in that experiment. Class-size effects persisted for at least 6 years and remained large enough to be important for educational policy. The results suggest that small classes in early grades have lasting benefits and that those benefits are greater for minority students than for White students.

Key words: class size, longitudinal study, long-term effects, randomized experiments

CLASS-SIZE REDUCTION is a policy option that is gaining increasing attention nationwide. Some 18 states have recently adopted policies that reduce class size with the goal of improving achievement. Class-size reduction was also included in President Clinton's education initiative.

Until recently, there was no consensus about whether class-size reduction led

Address correspondence to Larry V. Hedges, Department of Education, The University of Chicago, 5835 S. Kimbark Avenue, Chicago, IL 60637. E-mail: hedge@cicero.src.uchicago.edu

to increases in academic achievement and other desired outcomes of schooling. Even today, there is no universal consensus among scholars on interpreting the evidence on the relationship between class size and academic achievement. However, we argue that important evidence was provided by Project STAR, the Tennessee class-size experiment. This experiment provided rather strong evidence that class-size reduction leads to immediate increases in academic achievement in both reading and mathematics, with some evidence of larger effects for minority students. It is not clear from the STAR findings, however, whether those increases in academic achievement disappear after the students are no longer in small classes or persist.

In the present study, we investigated whether assignment to small classes in the primary grades has lasting effects on mathematics achievement and, if so, whether those effects vary across schools with different characteristics, such as those in the state of Tennessee. Mathematics achievement is often characterized as a gatekeeper for college admission, a critical filter restricting choice of majors, and a significant predictor of overall college success (College Board, 1990). Mathematics achievement has also been found to be an important predictor of other desirable life outcomes, such as earnings (see, e.g., Paglin & Rufolo, 1990).

Research on the Effects of Class Size

Well over 100 experiments and quasi-experimental studies of the effects of class size have been conducted, each involving assignment of students to smaller or larger classes. This literature was reviewed by Glass and Smith (1979); Glass, Cahen, Smith, and Filby (1982); Hedges and Stock (1983); and Mosteller, Light, and Sachs (1996). Although there is some disagreement about the interpretation of the experimental research (e.g., Educational Research Services, 1980; Slavin, 1984), these syntheses of research on the effects of class size suggest positive effects of class-size reduction on achievement and affect, with the effects becoming larger as classes become smaller.

However, most class-size experiments have been small scale and short term. Consequently, it is difficult to know whether the special circumstances surrounding the experiment caused the effects and whether they would have occurred in a more natural setting. That is, the small-scale experiments may have high internal validity, but it is difficult to know whether they will generalize to other settings.

Econometric work on education production functions (e.g., Hanushek, 1986) provides another means of studying the effects of class size. This research relies on the fact that naturally occurring class sizes vary across schools. It models the relationship between class size and an outcome (usually academic achievement) while controlling for student characteristics such as social class or prior achievement. The considerable number of econometric studies of the effects of class size on achievement have been reviewed elsewhere (see, e.g., Greenwald, Hedges, &

Laine, 1996; Hanushek, 1989; Hedges, Laine, & Greenwald, 1994). There is some controversy over the interpretation of the econometric studies. However, although few researchers would question the external validity of econometric studies, randomized experiments are generally thought to have higher internal validity.

An important addition to the research on class size was the Tennessee class-size experiment, Project STAR (Student–Teacher Achievement Ratio). This study (called "one of the great experiments in education in U.S. history" by Mosteller et al., 1996, p. 814) has substantially mitigated many of the problems of other class-size research. It is a large-scale experiment in which both teachers and students were randomized into classrooms within each of a broad range of participating schools. Consequently, it has both high internal validity and considerable external validity.

Initially, all Tennessee school districts were asked to participate in Project STAR, and about 100 schools had enough students in each grade to meet the size criterion for participation (at least 57 students per grade were 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 (of the 141 districts in Tennessee) became sites in the STAR experiment. Districts agreed to participate for 4 years and to allow site visitations for verification of class sizes, interviewing, and data collection, including extra student testing. They also allowed random assignment of pupils and teachers to class types from kindergarten through Grade 3.

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

We based this article on the data collected as part of the Lasting Benefits Study conducted by the Center of Excellence for Research and Policy on Basic Skills at Tennessee State University (Nye et al., 1994). The analyses reported during the original project (Word et al., 1990) and some analyses conducted since that time (e.g., Finn & Achilles, 1990), as well as reports based on those analyses (such as Mosteller, 1995), have important weaknesses. However, both these analyses and more recent (and satisfactory) analyses of data from Project STAR (e.g., Krueger, 1999) support the conclusion that small classes in kindergarten through Grade 3 lead to higher academic achievement at Grade 3. The effects of small classes were found to be remarkably consistent across schools, suggesting that small classes benefit students of all types in all kinds of schools (Nye, Hedges, & Konstantopoulos, 2000).

One particularly interesting aspect of earlier analyses of Project STAR is the

finding that the benefits of small classes were larger for minority students than for White students (e.g., Finn & Achilles, 1990; Krueger, 1999). Although the differential effects were not always statistically significant, class-size effects were larger for minority students than for White students in nearly all of the analyses conducted.

Lasting Benefits of Early Education Interventions

Although a number of early educational interventions have demonstrated immediate benefits, the effects typically disappeared over time so that the achievement of students who received the intervention was no higher than that of those who did not. For example, McKey et al. (1985) found that initial positive effects of Head Start completely faded out 3 years after the intervention. Haskins (1989) reported similar results in a comprehensive review of model preschool and Head Start programs. In an analysis of 300 early intervention programs, White (1985) found a pattern of initial effects that faded out over time, as did Barnett (1992). The Perry Preschool Project appears to be a notable exception to these findings, producing lasting gains in achievement (see, e.g., Barnett, 1985). However, research on many other early interventions suggests that achievement effects disappear by Grade 8 (Barnett, 1992), although Barnett (1995) offered a somewhat more optimistic view of possible lasting effects of early interventions.

There is relatively little evidence about whether small classes in early grades have enduring effects on later academic achievement. However, the Lasting Benefits Study, which tracked students from Project STAR after the conclusion of the experiment, did find enduring benefits through Grade 7 (Nye et al., 1994). Subsequent analysis of that data also revealed that students from small classes in the early grades had higher achievement in reading and mathematics as late as Grade 8 (Nye, Hedges, & Konstantopoulos, 1999).

The Present Article

In this article, we address the question of the long-term effects of small classes by examining the achievement of Project STAR students when they were in Grade 9, 6 years after the experiment had ended. The use of the Project STAR data to examine long-term effects of small classes was complicated by at least two factors.

The first and undoubtedly most important factor is attrition. Even though the students and teachers were initially assigned to classes at random, differences in patterns of attrition between small and regular-sized classes could bias the results of long-term follow-up comparisons of students initially in small and regular-sized classes. Of course, attrition is a problem that vexes not only Project STAR but nearly every longitudinal field study. (The possible effects of attrition within

Project STAR before Grade 3 were considered in a separate article and were found to be negligible, Nye et al., 2000.)

The second complicating factor is that, because new students were enrolled (and randomized) each year, all students did not spend the same amount of time in the experiment. That is, some of the students were part of the experiment for only 1 year (e.g., Grade 3), whereas others were part of the experiment for 4 years (K–3). If the effects are at all cumulative, then one would expect the effects on students with 4 years of small classes to be larger than those on students with only 1 year of small classes.

To address these complicating factors, we carried out several analyses to discern the effects of each of them. We first present an analysis in which the treatment is defined as class-size type (small or regular) at Grade 3. This analysis included students who entered the experiment at any point between kindergarten and Grade 3 and who received at least 1 year of the small (or regular) class size. This type of analysis might be considered typical for a longitudinal experiment.

To examine the effects of a greater or lesser number of years of treatment, we also present the results of analyses in which the treatment was defined as being in small classes for all 4 years (K–3) of the experiment. By comparing the results of this analysis with those of the first analysis, we gained insight about whether treatment effects were larger for students who had received more years of small classes. We used this analysis instead of an analysis by number of years in small classes because the sample sizes for each number of years (i.e., 1, 2, 3, or 4) in small classes were too small for meaningful analyses of class-size effects. Because the pattern of attrition was slightly different in the two analyses, we examined the pattern of differential attrition between treatment groups for each analysis separately.

Method

The design of the STAR experiment involved randomly assigning students and teachers to treatments within schools. The statistical analysis was a $2 \times 2 \times 2 \times 2$ (Treatment: small/regular-sized classes × Gender: male/female × Race: minority/White × Student Socioeconomic Status: free-lunch eligible/not eligible at Grade 3) analysis of variance (ANOVA), with socioeconomic status (SES) as the covariate. Note that, because our analyses revealed no effects of the full-time classroom aides in these data, we have not included terms for the effects of aides in this analysis. Therefore, the regular-sized class group included students in regular-sized classes both with and without a classroom aide. The dependent variable was the comprehensive mathematics achievement tests administered by the state of Tennessee to ninth graders. To express the magnitude of treatment effects only, we standardized the outcome variable so that the overall variance in the entire sample was 1. Thus, the small-class effects estimated are interpretable as the effect of being in a small class, expressed in standard deviation units.

Results

The Effects of Having Small Classes in at Least One Grade From K–3

In the first analysis of the experiment, we defined the treatment group to include all students who participated in the experiment for 1 or more years and were assigned to a small class in Grade 3. The treatment group therefore included students who had received 1 or more years of small classes. The control group consisted of students who had participated in the experiment for 1 or more years and had been assigned to a regular-sized class in Grade 3.

Attrition from the study. As expected in any longitudinal study, some students who began the study did not remain in the schools studied or were absent when the achievement tests were given. Attrition can be a source of bias in estimating treatment effects if students who leave one treatment group before follow-up are systematically different from corresponding students in other treatment groups. By the ninth grade, follow-up data could not be obtained on more than one half of the students who were in the STAR study in Grade 3. Table 1 contains the average SAT mathematics and reading achievement test scores for students who had participated in Grade 3 of the STAR experiment, broken down by whether the ninth-grade test data were available. Those whose test scores were available (the *stayers*) provided the data analyzed in this study. The students whose ninth-grade data were not available were the *leavers* whose characteristics might have biased the results.

For example, 1,181 (60.9%) of the 1,938 third graders assigned to small classes (who had third-grade math test scores) left the study before they were tested in the ninth grade. The average achievement of the students who left small class-

TABLE 1
Third-Grade Achievement Test Scores for Students Present (Stayers) and Not Present (Leavers) at Grade 9: Small Class in Grade 3

	Small class		Regular-sized class		
Grade 9 status	*n*	Average	*n*	Average	Difference
	Math achievement at Grade 3				
Present (stayers)	757	609.54	1,726	607.13	2.41
Not present (leavers)	1,181	631.56	2,418	621.76	9.80
Difference					−7.39*
	Reading achievement at Grade 3				
Present (stayers)	755	608.08	1,703	604.24	3.84
Not present (leavers)	1,160	629.53	2,387	618.92	10.61
Difference					−6.77*

*$p < .001$.

es was higher than that of the students who left regular-sized classes (see Table 1). The difference between the mean achievement of the leavers and that of the stayers who had been in small classes was about the same as the difference between the mean achievement of the leavers and that of the stayers who had been in the regular-sized classes.

The third-grade achievement gap between small and regular-sized classes—the crude treatment effect at third grade—was larger among the leavers than among the stayers. The differences between these gaps were statistically significant for both reading and mathematics achievement. Thus, it appears that the treatment effect at third grade was larger among the leavers than among the stayers. That achievement at third grade was positively correlated with achievement at ninth grade suggests that differential attrition may have led to underestimation of the effect of small classes on achievement. That is, it is not plausible that attrition between third and ninth grades made small classes look more favorable.

Treatment effects. The results of the ANOVA are summarized in Table 2, and the factor level means are given in Table 3. The effect of small classes was statistically significant, $F(1, 3545) = 12.701$, $p < .001$, and positive, with a treatment effect of about .146 standard deviation units. That is about 97% as large as the effect on mathematics achievement obtained at the end of third grade (which was about .150 standard deviations; see Nye et al., 2000).

The Minority × Small Class interaction was also statistically significant, $F(1, 3545) = 9.643$, $p < .01$, indicating that the small-class effect was larger for minority students than for White students. The small-class effect was .186 standard deviations for minority students and only .002 standard deviations for White students. None of the other interactions were statistically significant.

Mathematics achievement was significantly higher for girls than for boys, $F(1, 3545) = 16.499$, $p < .01$. The difference in mathematics achievement favoring

TABLE 2
ANOVA Summary on the Relationship Between Math Achievement and Small Class at Grade 3

Source	*df*	*SS*	*MS*	*F*	*Pr > F*
Female students	1	16.021	16.021	16.499	$p < .0001$
Minority	1	5.800	5.800	5.973	$p < .05$
Low SES at Grade 3	1	31.893	31.893	32.844	$p < .0001$
Small class at Grade 3	1	12.333	12.333	12.701	$p < .001$
Small Class × Female interaction	1	0.567	0.567	0.584	—
Small Class × Minority interaction	1	9.364	9.364	9.643	$p < .01$
Small Class × Low SES at Grade 3 interaction	1	2.992	2.992	3.081	—
Within groups	3,545	3,442.341	0.971		
Total	3,552	3,521.311			

TABLE 3
Ninth-Grade Average Math Test Scores, by Gender, Race, SES, and Grade 3 Class Size

Variable	Small class in Grade 3		Regular-sized class in Grade 3[a]		Overall	
	M	*n*	*M*	*n*	*M*	*n*
Gender						
Male	59.59	379	58.60	1,413	58.79	1,792
Female	62.88	383	60.59	1,387	61.09	1,770
Race						
Minority	60.38	342	57.16	1,316	57.82	1,658
White	61.83	419	61.79	1,476	61.80	1,895
SES						
Low	58.84	438	57.10	987	57.64	1,425
High	64.37	324	60.94	1,813	61.46	2,137
Math achievement	61.19	762	59.59	2,800	59.93	3,562

[a]The referent group consisted of students who were in regular-sized classes in Grade 3 and who were part of Project STAR in previous grades.

girls is surprising. In some earlier grades, the girls in Project STAR had higher mathematics achievement than the boys (although the differences were not significant). Our result may represent a widening of that gap, but the causes and implications are not clear. As expected, the low-SES and minority students had significantly lower achievement than the other students, $F(1, 3545) = 32.844$, $p < .01$, and $F(1, 3545) = 5.973$, $p < .05$, respectively.

We carried out a series of other analyses using related analytic models. For example, we eliminated the interaction effects that were not statistically significant and examined slightly different codings of the variables. None of these analyses suggested results that were qualitatively different from those reported here. In each case, the small-class effect had about the same positive magnitude and was statistically significant.

The Effects of Having Small Classes in All 4 Years From K–3

To get some idea of whether longer exposure to small classes produced greater effects than fewer years of small classes, we conducted an analysis in which the treatment group was defined as those who received 4 years of small classes. The comparison group consisted of all other children in the experiment, regardless of whether they had been assigned to the regular-sized class group or had entered the experiment later than kindergarten, had been assigned to the small class group at some point, and had received up to 3 years of small classes.

Attrition from the study. Because the definition of treatment groups in these analyses was quite different from that in the previous analyses, the pattern of attrition need not be the same. Table 4 contains data on the stayers and the leavers.

The treatment was defined as 4 years of small classes. The results of these analyses were substantially the same as those of the previous analysis of attrition.

In each analysis, the achievement gap between small and regular-sized classes (the crude treatment effect) was considerably larger among the leavers than among the stayers. The fact that the treatment effect was larger among the leavers than among the stayers suggests that attrition may have led to an underestimate of the treatment effect at ninth grade. Hence, it is not plausible that attrition made small classes look more favorable at ninth grade.

Treatment effects. The results of the ANOVA in which the small-class group was taken to be those who had received 4 years of small classes are summarized in Table 5, and the factor level means are given in Table 6. The effect of small classes was statistically significant, $F(1, 3545) = 31.047$, $p < .01$, and positive, with a treatment effect of about .340 standard deviation units. That is considerably larger than the effect estimated in the previous analysis and about 97% as large as the treatment effect in mathematics achievement obtained at the end of third grade for the students who had experienced 4 years of small classes (which was about .352 standard deviations, see Nye et al., 2000).

The Minority × Small Class interaction was also statistically significant, $F(1, 3545) = 4.575$, $p < .05$, indicating that the small-class effect was larger for minority students than for White students. The effect of 4 years of small classes was .207 standard deviations larger for minority students than for White students. None of the other interactions were statistically significant.

As in the previous analysis, girls had significantly higher mathematics achievement than boys, $F(1, 3545) = 7.940$, $p < .01$. Low-SES students had sig-

TABLE 4
Third-Grade Achievement Test Scores for Students Present (Stayers) and Not Present (Leavers) at Grade 9: Small Class in All Grades

Grade 9 status	Small class in all grades		Control group		Difference
	n	Average	*n*	Average	
		Math achievement at Grade 3			
Present (stayers)	305	614.82	2,178	606.89	7.93
Not present (leavers)	509	614.67	3,090	622.23	19.44
Difference					−11.51*
		Reading achievement at Grade 3			
Present (stayers)	304	615.22	2,154	604.03	11.19
Not present (leavers)	502	639.81	3,045	619.52	20.29
Difference					−9.10*

*$p < .001$.

TABLE 5
ANOVA Summary on the Relationship Between Math Achievement and Small Class in All Grades K–3

Source	*df*	*SS*	*MS*	*F*	*Pr > F*
Female students	1	7.685	7.685	7.940	$p < .01$
Minority	1	0.728	0.728	0.752	—
Low SES at Grade 3	1	14.567	14.567	15.051	$p < .0001$
Small class in Grades K–3	1	30.049	30.049	31.047	$p < .0001$
Small Class × Female interaction	1	0.264	0.264	0.273	—
Small Class × Minority interaction	1	4.428	4.428	4.575	$p < .05$
Small Class × Low SES at Grade 3 interaction	1	1.136	1.136	1.174	—
Within groups	3,545	3,431.082	0.968		
Total	3,552	3,489.939			

TABLE 6
Ninth-Grade Average Math Test Scores, by Gender, Race, SES, and Grades K–3 Class Size

Variable	Small class in K–3		Regular-sized class in K–3[a]		Overall	
	M	*n*	*M*	*n*	*M*	*n*
Gender						
Male	63.12	137	58.43	1,655	58.79	1,792
Female	66.76	156	60.54	1,614	61.09	1,770
Race						
Minority	64.60	110	57.34	1,548	57.82	1,658
White	65.33	183	61.42	1,712	61.80	1,895
SES						
Low	62.48	145	57.09	1,280	57.64	1,425
High	67.59	148	61.01	1,989	61.46	2,137
Math achievement	65.06	293	59.47	3,269	59.93	3,562

[a]The referent group consisted of students who had participated in Project STAR and who were not in small classes in all Grades K–3.

nificantly lower achievement than other students, $F(1, 3545) = 15.051$, $p < .01$. However, although minority students had lower achievement than White students, that difference was not statistically significant, $F(1, 3545) = 0.752$, $p > .05$.

It is important to recognize that in this analysis, the students were not randomly assigned to receive 4 (as opposed to 1, 2, or 3) years of small classes. Therefore, the students who received 4 years of small classes, for example, were likely to be different (in ways not entirely accounted for by race, gender, and social class) from those who did not. For example, they may have come from families that were less likely to move. Assuming that students who come from more stable homes may have higher levels of achievement, these students might

have been better prepared to benefit from small classes, which could exaggerate the effects. On the other hand, the comparison group against which the students who had received 4 years of small classes were contrasted included substantial numbers of students (at least $n = 835$, or more than 25% of the comparison group) who had received 1 or more years of small classes, which would tend to reduce the apparent effects of small classes.

Taken together, these findings suggest that, although the small-class effect may have been somewhat smaller at the Grade 9 follow up than at the Grade 3 posttest, the effect remained substantial compared with the small-class effect at Grade 3 (no less than 90% of this effect). That is, the small-class effect on achievement may have diminished somewhat, but it definitely did not fade to statistical or practical insignificance after 6 years.

Discussion

The STAR experiment demonstrated that small classes led to significantly higher achievement for students in reading and mathematics at the conclusion of the experiment. Our analyses of the data collected by the Lasting Benefits Study demonstrate that the positive effects of small classes in early grades resulted in mathematics achievement gains that persisted at least through Grade 9. Moreover, the achievement effects were similar in magnitude to those observed at Grade 3. That is, the effects of small classes in kindergarten through Grade 3 on achievement did not appear to be fading out by Grade 9.

Lasting benefits were found for all kinds of students in all kinds of schools, but the benefits of small classes were significantly larger for minority students. Thus, small classes may be a way to benefit all students while reducing the gap in achievement between White and minority students.

This 6-year follow-up study was also subject to substantial attrition, but the students who dropped out of small classes actually had higher achievement than those who dropped out of regular-sized classes, suggesting that the observed differences in achievement between students who had been in small and regular-sized classes was not due to attrition.

The STAR study provided an important (and perhaps the strongest) piece of converging evidence about the effectiveness of small classes in promoting achievement. The present study provides evidence that these effects do not disappear over time but confer on students from small classes in early grades achievement benefits that last at least until high school. This evidence points to the positive effects of small classes on achievement that are large enough and of sufficient duration to support policies of class-size reduction resulting in small classes (15–17 pupils) in the primary grades.

This study also demonstrates that students who experienced more years of small classes in kindergarten through Grade 3 had higher levels of achievement

6 years later than students who had fewer years of small classes. Although the evidence that the effects of small classes compound is not definitive, it is strongly suggestive. Therefore, it is likely that the lasting benefits increase with the number of years in small classes.

The mechanism by which small classes lead to higher achievement, and differentially higher achievement for minority students, is not clear. There are, however, some obvious hypotheses. First, small classes may permit teachers to more effectively individualize instruction. For example, with small classes teachers could identify and remedy incipient problems among students at risk for low achievement. Second, small classes may make existing instruction more effective. For example, small classes may lead to fewer disruptions or more effective whole-class instruction.

Understanding the mechanism could lead to more effective ways to implement class-size reductions and to improve their effectiveness. Such understanding is obviously desirable. Similarly, the effect of small changes in class size or of altering the teacher–pupil ratio (without necessarily changing class size) is not immediately obvious from this experiment.

This research has not even answered all the important questions about the effects of small classes on achievement and other desirable outcomes of schooling. For example, the effects of small classes on achievement growth are not yet clear. Although it appears that students in small classes experience an increase in achievement that persists over time, the growth trajectories of students from small and regular-sized classes have not been studied in detail. This too would be desirable.

Other questions concern the effects of small classes on the other outcomes of schooling such as student engagement, motivation, and persistence. Of particular concern is whether students who are in small classes in early grades will be more likely to graduate from high school, more likely to go to college, and more likely to successfully participate in the labor force. Researchers may use data that are continuing to be collected by the Lasting Benefits Study to ultimately answer some of these questions.

REFERENCES

Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool Program and its policy implications. *Educational Evaluation and Policy Analysis, 7,* 333–342.

Barnett, W. S. (1992). Benefits of compensatory preschool education. *Journal of Human Resources, 27,* 279–312.

Barnett, W. S. (1995). The long term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5,* 25–250.

The College Board. (1990). *Changing the odds.* New York: Author.

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.

Glass, G. V, & Smith, M. E. (1979). Meta-analysis of research on class size and achievement. *Edu-*

cational Evaluation and Policy Analysis, 1, 2–16.
Glass, G. V, Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School class size: Research and policy.* Beverly Hills, CA: Sage.
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*(4), 45–65.
Haskins, R. (1989). Beyond metaphor: The efficacy of early childhood education. *American Psychologist, 44,* 274–282.
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*(3), 5–14.
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. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114,* 497–531.
McKey, R. H., Condelli, L., Ganson, H., Barret, B. J., McConkey, C., & Plantz, M. C. (1985). *The impact of Head Start on children, families, and communities.* Washington, DC: U.S. Government Printing Office. (DHHS Publication No. OHDS 85–31193)
Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5,* 113–127.
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., Hedges, L. V., & Konstantopoulos, S. (1999). The long term effects of small classes: A five year followup of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 21,* 127–142.
Nye, B. A., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37,* 123–151.
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.
Paglin, M., & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male–female earnings differences. *Journal of Labor Economics, 8,* 123–144.
Slavin, R. (1984). Meta-analysis in education: How has it been used? *Educational Researcher, 13*(8), 6–15, 24–25.
White, K. R. (1985). The efficacy of early intervention. *The Journal of Special Education, 19,* 401–416.
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, TN: Tennessee Department of Education.