Do Minorities Experience Larger Lasting Benefits From Small Classes?

BARBARA NYE
Tennessee State University

LARRY V. HEDGES
The University of Chicago

SPYROS KONSTANTOPOULOS
Northwestern University

ABSTRACT Recent research from randomized experiments on class size points to positive effects of small classes that persist for several years, but the evidence about the social distribution of effects is less clear. Some scholars have contended that the immediate effects of small classes are larger for minorities and for disadvantaged persons (e.g., J. D. Finn & C. M. Achilles, 1990). Those claims have led to policies of class size reduction specifically to reduce inequality in educational outcomes. The authors used data from a 5-year follow-up to Project STAR to investigate whether differential effects of small classes on achievement for minority students persist. A repeated measures analysis suggested that there was a statistically significant, positive differential lasting benefit of 4 years for minorities enrolled in small classes in reading, and a negative differential lasting benefit for girls enrolled in small classes in mathematics over 5 years following the experiment. Thus, it appears that the lasting benefits of 4 years of small classes may reduce racial and ethnic inequality in reading and gender inequality in mathematics.

Key words: minority students, Project STAR, reading and mathematics, small class size

The problem of how one should allocate resources to increase achievement for all students and to close achievement gaps between advantaged and disadvantaged students is a central issue facing education researchers and policymakers. For example, the No Child Left Behind Act (the most important federal legislation on elementary and secondary education) requires that schools make adequate yearly progress in closing achievement gaps, or face sanctions for failing to do so. Several kinds of strategies might be used to address achievement gaps between groups in society.

One approach is that educators provide different or additional instruction or services targeted specifically at the disadvantaged group. Various forms of compensatory education are an example of that approach (e.g., pull-out programs, peer tutoring). The disadvantage of the approach is that the resources allocated specifically for addressing achievement gaps do not directly benefit the advantaged group. An alternative approach is for educators to provide an intervention that would directly benefit all students by increasing their achievement while also closing the minority-majority or advantaged-disadvantaged achievement gaps.

There is some evidence that reducing class size might be such an intervention. Finn and Achilles (1990) analyzed data from the first 2 years of Project STAR (Student Teacher Achievement Ratio; the Tennessee class-size experiment) and found evidence that the positive effects of small classes were larger for minority students. However, those differential effects (Class Size × Minority interactions) were statistically significant only for reading achievement. A more recent analysis of 4 years of Project STAR suggested differential effects for minority students (Kreuger, 1999), but other analyses of the same data have found less persuasive evidence for differential effects (Nye, Hedges, & Konstantopoulos, 2000a). Similar findings of somewhat larger long-term class-size effects for minorities were obtained in the initial reports of the Lasting Benefits Study (the follow-up study of students who had been in Project STAR), but their statistical significance was not evaluated (Nye et al., 1994). Those analyses were widely reported, however, and were interpreted by many educational researchers, policymakers, and practitioners as evidence that the effects of small classes were more positive for minorities and disadvantaged persons than for all students and that these effects persisted over time.

Class size is an appealing policy variable because it is under administrative control. It is not an instructional variable in the sense that knowing class size gives little insight about the instructional processes that may be unfolding in the classroom. However, unlike reform strategies that involve changing instruction in the classroom, class-size reduction is relatively straightforward for teachers to implement given adequate resources. Moreover, unlike strategies that involve changes in instruction, administrators can determine whether class-size reduction has been implemented as planned.

Address correspondence to Larry V. Hedges, Department of Education, University of Chicago, 5835 South Kimbark Avenue, Chicago, IL 60637. (E-mail: lhedges@uchicago.edu)

Manipulating class size is a policy option that is gaining increasing attention throughout the nation. Eighteen states have recently adopted policies that reduce class sizes in order to improve achievement; reduction of class size is included in the recent education initiative signed by President George W. Bush. Some states and localities (such as Madison, Wisconsin) have explicitly linked policies of reducing class size to efforts to reduce achievement gaps between minority and majority students.

To know whether the policy of reducing class size will achieve the intended goal, one must know whether the differential class size-effects for minorities will persist over time or diminish (as have the effects on many other early interventions). In this article, our purpose was to address that question by using data from Project STAR.

Related Literature

There are three traditions of research on the effects of class size on academic achievement. All of those traditions frame the problem as one of estimating the effect of an education policy variable (class size) on a student outcome (academic achievement). The intervening variables in the mechanism by which the treatment produces its effect (presumably by changing instruction) are of little interest. The research question is how much achievement may change if class size is manipulated by a certain amount.

One tradition of research on class size is represented by over 100 quasi-experimental studies of the effects of class size. That work has been reviewed by Glass and Smith (1979), Glass, Cahen, Smith, and Filby (1982), and Hedges and Stock (1983). Their consensus is that the bulk of the studies suggests that class-size reduction improves achievement, although others disagree about the interpretation of the research (e.g., Educational Research Services, 1980; Slavin, 1984). However, there are only 14 randomized experiments among the collection of 100 studies, the majority of which involve only small samples of students. As a result, those small experiments do not individually or collectively provide persuasive external validity ("generalizability").

A second tradition in the study of class size effects is the econometric work on education production functions. That research reported data from surveys or administrative records from large collections of classrooms. The researchers used statistical (multiple regression) models that expressed student achievement as being a function of class size while statistically controlling for student characteristics such as social class, prior academic achievement, race, and gender. If the variables that are used as statistical controls (e.g., social class, prior achievement, race, and gender) adequately control for all other sources of variation in achievement except class size, then these studies provide estimates of the relation between class size and achievement.

Econometric studies that examined the effects of class size on achievement have been reviewed elsewhere (see, e.g., Greenwald, Hedges, & Laine, 1996; Hanushek, 1986, 1989; Hedges, Laine, & Greenwald, 1994). Some controversy exists over the interpretation of the econometric studies, but the consensus is that they are probably less internally valid than experiments for several reasons.

First, the student background data used for statistical controls are often limited. The best studies use data on individual students and include measures of prior achievement, socioeconomic status (SES), and other characteristics. Many studies have much weaker control variables, for example, no pretest scores and only free or reduced-price lunch eligibility. Sometimes the control variables are not available for each individual student but obtainable only at the level of an aggregate such as schools or school district. Such weak control variables may fail to fully account for individual differences between students assigned to classes of different sizes. In such cases, the association between class size and achievement may be confounded with other uncontrolled differences between achievement in classes of different sizes.

For example, a plausible rival hypothesis is that achievement (or expected achievement) causes students to be assigned to classes of different sizes, not vice versa. That is, students may be assigned to smaller classes precisely because their achievement is low (for example in compensatory or remedial programs). In such a case, failure to explicitly control for each student's prior achievement could result in uncontrolled bias.

Second, in econometric studies, researchers often use poor data on the independent variable of class size. For example, researchers may use the ratio of number of students in a school to the number of teachers in that school as a measure of class size for all classes in the school. Such a measure is problematic for two reasons. One reason is that classes in a school may not all be the same size. The other reason is that not all teachers in a school may be assigned to classrooms because there may be teachers with specialized functions who have no classroom assignment. Thus, pupil:teacher ratio may underestimate class size. Poor measurement of the independent variables in regression models can bias estimates of effects (although when there is more than one independent variable the direction of the bias is not obvious without additional information; see, e.g., Cook & Campbell, 1978).

Thus, the previous research of both types (econometric studies and small-scale experiments) is limited for the purposes of determining whether there are main effects of class size on achievement. Previous research is even more limited for evaluating the differential effects of small classes on minorities or the disadvantaged. Relatively few of the small-scale experimental studies and none of the econometric studies have evaluated differential effects for minorities. The few evaluations of differential effects were subject to the same limitations (and generally even lower statistical power) as have plagued the evaluation of main effects (see Cronbach & Snow, 1977).

Third, class-size research involves large-scale experimental studies. There is only one example of this research tradition—the Tennessee class size experiment. Large-scale field experimentation does not have the weaknesses of either small-scale experiments or econometric studies, as discussed in the Project STAR section of this study.

Lasting Benefits of Early Education Interventions

Several early educational interventions have demonstrated immediate effects on student achievement, but the effects typically disappeared over time so that the achievement of students who received the intervention was no higher than those who did not. For example, McKey et al. (1985) found that immediate positive effects of Head Start completely disappeared by 3 years after the intervention. In a comprehensive review of model preschool and Head Start programs, Haskins (1989) reported similar results. White's (1985) analysis of 300 early intervention programs also found a pattern of initial effects that diminished over time, as did Barnett's (1992) investigation. A notable exception to those findings is the Perry Preschool Project, which produced lasting gains in achievement (see e.g., Barnett, 1985). A somewhat more optimistic view of possible lasting effects of early interventions was offered by Barnett (1995).

Project STAR

An important source of information on class size is Project STAR. The randomized experiment, which was commissioned in 1985 by the Tennessee state legislature and implemented by a consortium of Tennessee universities and the Tennessee State Department of Education, is described in detail by Word et al. (1990).

Ultimately, 79 elementary schools in 42 school districts became sites in the STAR experiment. The school districts had to agree to participate for 4 years and to allow site visitations for verification of class sizes, interviewing, and data collection, including extra student testing. School districts 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 schools. Educators in the school districts followed their own policies, curricula, and so forth. Because no student received any less service than would normally have been provided by the state as a consequence of being in the STAR project, there was no incentive for any student or school not to participate.

The experiment randomly assigned kindergarten students into small classes (13 to 17 students), regular classes (22 to 26 students), or regular classes with a full-time classroom aide. Teachers also were assigned randomly 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 assigned randomly to classes at that time.

Project STAR, called "one of the great experiments in education in U.S. history" by Mosteller, Light, and Sachs (1996, p. 814), substantially mitigates many of the problems of other class-size research. Because it is a randomized experiment that randomizes teachers and students into classrooms within each participating school, it has high internal validity. Project STAR involves a broad range of schools from throughout a diverse state. It includes large urban districts and small rural ones, and a range of wealth ranging from some of the wealthiest school districts in the country to some of the poorest. Therefore, Project STAR includes essentially the entire range of education 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 normal operation of the schools that participated and, therefore, is likely to avoid the effects associated with experimental programs.

Previous Analyses of Data From Project STAR

Previous analyses of Project STAR have established the overall effects of small classes on achievement of all students during the experiment (Finn & Achilles, 1990; Krueger, 1999; Nye, Hedges, & Konstantopoulos, 2000b; Word et al., 1990). Other studies examined the differential effect of small classes on minority students during the period of the experiment itself, but not the long-term differential effects for minorities (Krueger, 1999; Nye et al., 1994; Nye, Hedges, & Konstantopoulos, 2000a). Yet, other analyses have concentrated on the long-term effects of reduced class size in the early grades on the achievement of all students (Nye, Hedges, & Konstantopoulos, 1999). The data from Project STAR has not been exploited to date for the purpose of studying the persistence of differential effects of small classes on the achievement of minority students. In this article, we addressed the question of the differential long-term benefits of small classes for minority students by examining the achievement of Project STAR students during the 5 years after the experiment ended, when these students were in Grades 4 to 8.

Method

The design of the Project STAR experiment involves random assignment of students and teachers to treatments within schools. The study is conceptually a series of within-school experiments in which participants use 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 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 (e.g., treatment effects) in a manner that takes the clustering of the sample by school.

In previous analyses of data from Project STAR, researchers found larger immediate effects of small classes for minority students. The mechanism leading to that differential effect is unclear. Perhaps because teachers have more time to allocate to each student in small classes, they can more effectively identify and correct learning difficulties that plague minorities and other groups that have lower average achievement. Also possible is that small classes are subject to less disruption and minority students are differentially affected by disruptions in the learning environment. Whatever the mechanism might be, the fact that larger immediate effects of small classes have been found for minorities raises the question of whether the differential effects persist over time.

Similarly, in their analyses of immediate effects of Project STAR, Nye, Hedges, and Konstantopoulos (2000b) reported some differential effects by gender. Again, the mechanism by which those effects were produced is not clear. However, the fact that larger immediate effects of small classes have been found suggests that these differential effects for gender might persist.

One could carry out separate analyses at each grade level. However, because the data included repeated measures for the same students over time, it is possible to carry out a repeated measures analysis of the data. Because such an analysis uses the data from all five grades in the same analysis, it should produce more powerful tests of the Minority × Small Class interaction than analyses at each grade level separately (at least if the interaction is the same across grades).

Although the analysis of pooled data is probably more powerful than analyses of each grade separately, it also has some disadvantages. The pooled analysis requires pooling data on different tests into the same analysis. If the raw test scores are pooled, there will be differences in both the mean and variability of test scores across grades. The differences in variance across grades are particularly problematic for models that assume a common residual variance. Of course, such pooling makes sense only if the tests are strictly equated. Previous analyses of these data led us to question the adequacy of equating over the 4-year period.

We addressed the problem of unequal variances by standardizing test scores within grades before combining the data. That response may have solved the problem of unequal residual variances but did not entirely solve the problem of making data from different grades comparable (because the true variability of achievement may have changed across grades). Standardizing the test scores made the main effects of grade difficult to interpret. In particular, the grade effects should not have been interpreted as estimates of growth.

Another limitation of our analysis of pooled data was that it assumed that effects, such as the Minority × Small Class interactions, were the same across grades (i.e., the Minority × Small Class × Grade interactions were negligible). Although we investigated the data empirically and found that the interactions were statistically insignificant, the power to test higher order interactions was even smaller than that for testing lower order interactions, so the tests probably had relatively little power.

The specific model for the analyses reported here is one in which the achievement test score Y_{ijk} of the ith student in the *j*th school at the *k*th grade is

$$Y_{ijk} = \beta_{0ij} + \beta_{0ij}\mathrm{GRADE}_{ijk} + \in_{ij}, \tag{1}$$

where β_{0ij} is a student-specific intercept, β_{1ij} is a student-specific grade effect, and $\in_{ij}$, is a student and grade level-specific residual.

The within-school model used in our analysis treated student achievement as a function of (a) student characteristics (gender and minority-group status or social class), (b) class size, (c) interaction of assignment with minority-group status and gender, and (d) three-way interaction of class size, gender, and minority-group status. The specific model for variation of coefficients between students within schools (Level 2 model) is

$$\beta_{0ij} = \pi_{00j} + \pi_{01j}\,\mathrm{Female}_{ij} + \pi_{02j}\mathrm{Minority}_{ij} + \pi_{03j}\,\mathrm{SES}_{ij} + \pi_{04j}\,\mathrm{Small}_{ij} \tag{2}$$

$$+ \pi_{05j}\,\mathrm{Small} \times \mathrm{Female}_{ij} + \pi_{06j}\,\mathrm{Small} \times \mathrm{Minority}_{ij} + \pi_{07j}\,\mathrm{Female} \times \mathrm{Minority}_{ij} + \pi_{08j}\,\mathrm{Small} \times \mathrm{Female} \times \mathrm{Minority}_{ij} + x_{0ij}, \tag{3}$$

where female_{ij} is a dummy variable for gender, $\mathrm{minority}_{ij}$ is a dummy variable for minority-group membership, SES_{ij} is a dummy variable for free-lunch eligibility, smallij is an indicator variable for small-class size, Small × Minorityij is the Small by Minority interaction, Small × Femaleij is the Small by Female interaction, Minority × Femaleij is the Minority by Female interaction, and Small × Female × $\mathrm{Minority}_{ij}$ is the three-way Small by Female by Minority interaction, and $\xi 0_{ij}$ is a student-specific residual.

In either case, the model for the grade effect at Level 2 is

$$\beta_{1ij} = \pi_{10j} + x_{1ij}, \tag{4}$$

where $\pi 10_j$ is a school-specific intercept and x_{1ij} is a student-specific residual.

We modeled all coefficients except the intercept as random whenever the variation across schools was statistically significant. The intercept was always treated as random. The specific Level 3 model for the *m*th coefficient in the model for the intercept and grade-level slope, respectively, in *j*th school (π_{0mj} and π_{1mj}) is, therefore,

$$\pi_{0mj} = g_{0m0} + h_{0mj}, \tag{5}$$

and

$$\pi_{10j} = g_{100} + h_{10j}, \tag{6}$$

where γ_{0m0} and γ_{100} are fixed effects and η_{0mj} and η_{10j} are Level 3 residuals (random effects). Therefore, the object of the statistical analysis is to estimate the 10 fixed effects (intercepts γ_{000}, γ_{010}, . . . , γ_{080}, and γ_{100}) determining each of the 10 π_{ij} effects and the corresponding between-school variance components (variances of the η_{0mj} and η_{10j}). We conducted separate analyses for each of the two dependent variables (California Tests of Basic Skills) mathematics and reading test scores).

Results and Discussion

We first considered the pattern of differential effects of small classes on minority students. Table 1 shows the small-class effects that remained for White students and minority students in reading and mathematics achievement at Grades 4 through 8 (1 to 5 years after the experiment ended). Each of the small-class effects was presented as the difference between the mean achievement in small classes in Grades K to 3 and that in all remaining combinations of classes (e.g., 1, 2, or 3 years in small classes) divided by the overall standard deviation of test scores at that grade in that subject matter. Thus, each of the small-class effects was an effect size expressed in standard deviation units.

Comparing the effect-size estimates and ignoring statistical significance, the table shows that the small-class effect for minorities was larger than for White students in every case except mathematics at Grade 5. In the case of mathematics achievement, the small-class effect for minorities was only slightly larger in Grades 4 and 7, but was approximately 30% larger in Grades 6 and 8. Grade 5 was an anomalous case in which the effect size in mathematics (unlike in reading) for White students was somewhat larger than for minority students. In the case of reading achievement, the small-class effect for minorities was consistently much larger than for White students in all grades.

The small-class effects for boys and girls separately in each grade are shown in Table 2. Comparing those effect size estimates and ignoring statistical significance, the table shows that the small class effect for boys in mathematics was much larger than for girls at every grade level. The pattern in reading was comparable. As in mathematics, the small-class effect in reading was larger for boys in all grades. Thus, it appeared that boys had greater lasting benefits from small classes in mathematics and reading than did their female counterparts.

The differential small class effects are disaggregated by gender and race in Table 3. Comparing those effect size estimates and again ignoring statistical significance, the table shows that the small-class effect for minority girls was

TABLE 1. Small-Class Advantage, by Race

	Mathematics (ES)		Reading (ES)	
Grade level	White	Minority	White	Minority
4	0.25	0.27	0.25	0.33
5	0.26	0.24	0.26	0.31
6	0.21	0.28	0.20	0.37
7	0.31	0.32	0.24	0.36
8	0.25	0.34	0.23	0.32

Note. Raw mean differences were divided by the standard deviation of the particular score across all students in all classes. ES = effect size.

TABLE 2. Small-Class Advantage, by Gender

	Mathematics (ES)		Reading (ES)	
Grade level	Girls	Boys	Girls	Boys
4	0.19	0.35	0.24	0.35
5	0.12	0.42	0.20	0.37
6	0.09	0.44	0.20	0.41
7	0.24	0.55	0.26	0.47
8	0.23	0.46	0.27	0.40

Note. Raw mean differences were divided by the standard deviation of the particular score across all students in all classes. ES = effect size.

TABLE 3. Small-Class Advantage, by Gender Within Race

	Mathematics (ES)				Reading (ES)			
	Boys		Girls		Boys		Girls	
Grade level	White	Minority	White	Minority	White	Minority	White	Minority
4	0.35	0.21	0.14	0.33	0.31	0.30	0.18	0.36
5	0.44	0.27	0.10	0.20	0.36	0.33	0.16	0.29
6	0.39	0.44	0.04	0.13	0.29	0.51	0.11	0.23
7	0.51	0.40	0.14	0.25	0.36	0.40	0.13	0.32
8	0.39	0.39	0.12	0.30	0.30	0.39	0.17	0.25

Note. Raw mean differences were divided by the standard deviation of the particular score across all students in all classes. ES = effect size.

consistently larger than for White girls in mathematics and reading achievement. In contrast, the small class-effect for minority boys was less consistent. Thus, it appeared that minority girls experienced greater lasting benefits of small classes than did White girls.

Although Tables 1, 2, and 3 suggested a pattern of larger estimated lasting effects of small classes in reading (but not in mathematics) for minorities at every grade level, they did not provide the evidence to evaluate the statistical significance of these differences. Therefore, we used hierarchical linear model analyses, whose results are summarized in Table 4. Overall, girls had significantly higher achievement than did boys, with a covariate adjusted advantage of 0.205 standard deviation in mathematics and 0.156 standard deviation in reading. Minority and low-SES students had significantly lower achievement in reading and mathematics. The Female × Minority interaction was not statistically significant for either reading or mathematics achievement, suggesting that minority girls benefited no more or less from small classes than did minority boys.

Perhaps more important for this experiment, the average lasting effect of small classes was positive and statistically significant. The average lasting effect of small classes was about 0.357 standard deviation in mathematics and 0.271 standard deviation in reading achievement. The overall Minority × Small Class interaction was small and statistically insignificant for the outcome of mathematics achievement but positive and statistically significant for the outcome of reading achievement. That finding suggests that this analysis provides no evidence of lasting differential benefits of small classes for minorities in mathematics achievement, but it does suggest differential lasting benefit in reading. On average across the 5 years, minority students experienced a differential lasting benefit of about 0.182 standard deviation in reading—about 67% larger than for nonminority students.

The Female × Small-Class interaction was not statistically significant for reading but was negative and statistically significant for mathematics achievement. That is, girls benefited less than 0.227 standard deviation from small classes in mathematics than did boys. That finding raises the question of whether girls are better off (in terms of mathematics achievement) in small classes or in regular-sized classes.

By combining the main effect of small classes and the Female × Small-Class interaction in mathematics, we determined that the net effect of small classes on mathematics achievement was still positive for girls (e.g., –0.227 + 0.357 = 0.130). Of course, the net effect of small classes on mathematics achievement for boys was even larger (e.g., 0.0 + 0.357 = 0.357).

TABLE 4. Hierarchical Linear Modeling 3-Level Analysis of Differential Minority Effects for Grades 4 Through 8

Variable	Mathematics	Reading
Intercept (β_{0j})		
Intercept (γ_{00})	–0.060*	–0.060*
Residual variance	0.100*	0.061*
Female (β_{1j})		
Intercept (γ_{01})	0.205*	0.156*
Minority (β_{2j})		
Intercept (γ_{02})	–0.264*	–0.429*
Low SES in Grade 3 (β_{3j})		
Intercept (γ_{03})	–0.341*	–0.401*
Small class in Grades K–3 (β_{4j})		
Intercept (γ_{04})	0.357*	0.271*
Female–small-class Interaction (β_{5j})		
Intercept (γ_{05})	–0.227*	–0.089
Minority–small-class Interaction (β_{6j})		
Intercept (γ_{06})	0.014	0.182*
Female–minority (β_{7j})		
Intercept (γ_{07})	–0.035	0.008
Female–minority–small-class interaction (β_{8j})		
Intercept (γ_{08})	0.160	–0.024
Grade (β_{9j})		
Intercept (γ_{09})	–0.010	–0.001
Residual variance	0.012*	0.003*

Note. SES = socioeconomic status.
$^*p < .05$.

Conclusions

The analyses reported here support the validity of the conclusion that small classes in Grades K–3 lead to higher academic achievement over the subsequent 5-year period. Repeated measures analyses that examined the students who received 4 years of small classes (kindergarten through Grade 3) revealed larger lasting benefits from small classes (as expected) and a pattern of positive differential lasting benefits for minority students in reading (but not mathematics). The repeated measures analyses pooling data across grade levels revealed a statistically significant pooled differential lasting benefit on reading achievement for minorities of 4 years of small classes over the 5-year period subsequent to the experiment. Thus, there is evidence that 4 years of small classes produced lasting differential benefits for minority students in reading.

There is some evidence that in mathematics achievement girls experienced smaller lasting benefits from small classes than did boys. However, because the main effect of being female was positive in these data, the negative differential effect for girls lead to a smaller gender difference among students who had 4 years of small classes than among those in regular-sized classes. Thus, evidence indicated that small classes could help reduce overall racial and ethnic inequality in reading achievement and reduce gender inequality in mathematics achievement.

It is remarkable that the intervention of reducing class size in the early grades, unlike many other early interventions, has effects that endure for at least 5 years after the intervention. Moreover, the analyses reported here provide at least some support for the idea that small classes are an intervention that can simultaneously raise achievement for all students as well as reduce inequality.

Perhaps most disappointing about these data is that they shed little light on the mechanism by which class-size reduction affects achievement. Because little data were collected about classroom processes or instruction, this experiment cannot indicate how instruction might have changed in the small classes to increase achievement or why there were differential effects for minority students. We hope that future research will include the collection of data to elucidate a mechanism by which policy variables such as class size produce effects.

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.

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

Cook, T. D., & Campbell, D. T. (1978). *Quasi-experimentation*. Chicago: Rand McNally.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington.

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

Kreuger, A. B. (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*. (DHHS Publication No. OHDS 85-31193) Washington, DC: U.S. Government Printing Office.

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 follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis 21*, 127–142.

Nye, B. A., Hedges, L. V., & Konstantopoulos, S. (2000a). Do minorities and the disadvantaged benefit more from small classes? Evidence from the Tennessee class size experiment. *American Journal of Education, 109*, 1–26.

Nye, B. A., Hedges, L. V., & Konstantopoulos, S. (2000b). 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* (Technical Report). 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*(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., et al. (1990). *Student/teacher achievement ratio (STAR): Tennessee's K-3 class-size study*. Nashville: Tennessee Department of Education.