Educational Evaluation and Policy Analysis
Summer 1999, Vol. 21, No. 2, pp. 97–109

Tennessee's Class Size Study: Findings, Implications, Misconceptions

Jeremy D. Finn
State University of New York at Buffalo

Charles M. Achilles
Eastern Michigan University

After years of debate, speculation, and research, Tennessee's Project STAR produced clear answers to the question, "Do small classes result in improved academic achievement in the elementary grades?" This article describes the features that made STAR unique and summarizes the findings with regard to pupil performance and behavior. New analyses show the magnitudes of the "small-class advantage" during and after the 4-year experimental period. The positive findings of STAR have been greeted with enthusiasm by the education community and are providing impetus for class size reduction (CSR) efforts in many districts. At the same time, some detractors continue to oppose the idea. Although they usually do not take issue with the strength of the STAR design, they disagree that the findings warrant CSR initiatives in most cases. This article examines those arguments critically. Finally, recommendations are offered for policymakers, education practitioners, and researchers for using the information learned to date about the relationship of class size with students' academic achievement.

The issue of class size has been debated by educators for centuries. In fact, one analysis traces writing on the topic to the Babylonian Talmud, in which the maximum size of bible classes was specified as 25 pupils (Angrist & Lavy, 1996). In recent decades, well over 100 empirical studies of class size have been completed. Because the studies employed nonexperimental designs, and because many involved small samples or were of short duration, few definitive conclusions could be drawn. Tentative conclusions were summarized in several widely read reviews, specifically the Glass-Smith meta-analysis (1978) and reviews by the Educational Research Service (Robinson, 1990; Robinson & Wittebols, 1986) and Slavin (1989). The reviews converged on four major propositions. First, "reduced class size can be expected to produce increased academic achievement" (Glass & Smith, 1978, p. 4), although the effects of even substantial reductions are small (Slavin, 1989). Second, "the major benefits from reduced class size are obtained as the size is reduced below 20 pupils" (Glass & Smith, 1978, p. v). Third, small classes are most beneficial in reading and mathematics in the early primary grades (Robinson, 1990). Fourth, "the research rather consistently finds that students who are economically disadvantaged or from some ethnic minorities perform better academically in smaller classes" (Robinson, 1990, p. 85).

In 1985, the Tennessee legislature funded an experiment, Project STAR (Student/Teacher Achievement Ratio), to provide more definitive answers[1]. For several reasons, Project STAR came to eclipse all of the research that preceded it. First, it was a controlled scientific experiment; students entering kindergarten were assigned at random to a small class (13–17 students), a regular class (22–26 students), or a regular class with a full-time teacher aide within each participating school. The within-school randomization controlled for a host of between-school differences, including differences in the populations served, differences in per-pupil expenditures and instructional resources, and

differences in the composition of the school staff. To the extent possible with empirical data, it permitted causal conclusions to be drawn about the outcomes. Teachers were assigned to the classrooms at random. The class arrangement was maintained throughout the day and throughout the school year. There was no intervention other than class size and teacher aides.

Second, the study was extensive. More than 6,000 students in 329 classrooms (representing 79 schools and 46 districts) participated in the first year, and almost 12,000 students were involved in the course of the 4-year intervention. It also had ample duration. Children assigned to one of the three class types were kept in the same experimental condition for 4 years, through Grade 3.[2] A new teacher was assigned to the class each year. All pupils returned to regular classes in Grade 4 when the experiment ended. However, researchers were able to follow the participants through the ensuing grades. To date, follow-up data have been analyzed through Grade 7. Analyses of STAR data are continuing.

Third, researchers collected an array of outcome measures at the most appropriate levels, namely, individual pupils, their teachers, and their schools. Both norm-referenced and criterion-referenced achievement tests were administered at the end of each school year. The Stanford Achievement Test (SAT) battery was administered annually in Grades K–3, and the Comprehensive Tests of Basic Skills (CTBS) were administered in subsequent grades; the state's Basic Skills First (BSF) curriculum-referenced tests in mathematics and reading were administered in Grades 1–3. Learning behaviors were assessed in Grades 4 and 8 and school experiences (e.g., school changes, in-grade retentions) were recorded each year. Teachers and aides completed questionnaires and time logs to document their perceptions and experiences.

Project STAR built on the principles identified in prior research. The intervention began in the primary grades. The study involved a real reduction in class size, from a median enrollment of 24 pupils to a median of 15. The study's design permitted an analysis of the effects on groups of students by race, gender, and socioeconomic status. The teacher aide condition allowed researchers to determine whether reducing the pupil-teacher ratio in a classroom would produce similar effects to reducing the actual class size.

The objectives of this article are (a) to summarize the findings of Project STAR, with particular attention to achievement and behavioral outcomes; (b) to present results of new analyses of the magnitudes of effects produced by STAR small classes; and (c) to discuss the implications of STAR findings for educators and policymakers, clarifying several misinterpretations of the findings expressed by some researchers.

Project STAR: The Findings

Details of the STAR procedures and results to date have been provided in a number of publications, including Achilles, Finn, and Bain (1997); Finn (1998); Finn and Achilles (1990); Mosteller (1995); and Word et al. (1990).[3] For quantitative outcomes, statistical procedures were appropriate to the complex experimental design, specifically, analysis of variance and multivariate analysis of variance models for schools nested within settings (inner city, urban, suburban, rural), schools crossed with classroom arrangements (small, regular, aide), and students nested within classes.[4]

The study yielded an array of benefits of small classes, including improved teaching conditions, improved student performance during and after the experimental years, improved student learning behaviors, fewer classroom disruptions and discipline problems, and fewer student retentions. Among the results obtained with respect to pupils' academic achievement and classroom behavior were the following:

1. Statistically significant differences were found among the three class types on all achievement measures and in all subject areas, in every year of the experiment (K–3). On average, students in small classes evidenced superior academic performance to those in the other conditions.

2. The effects were always attributable to the difference between the average performance of small classes and that of the other class types. No significant differences were found between teacher aide and regular classes in any year of the study.

3. There was no interaction with gender; that is, the small-class advantage was found for boys and girls alike.

4. In each grade, there was some significant interaction with race/ethnicity or with school location. The benefits were substantially greater for minority students or students attending inner-city schools in each year of the study.

5. The small-class advantage was also statistically significant for all school subjects in every

subsequent year (Grade 4 and beyond). Analyses to date have confirmed this result through Grade 7.

6. Students who had been in small classes exhibited superior engagement behaviors in Grade 4 (i.e., more effort spent on learning activities, more initiative taking, and less disruptive or inattentive-withdrawn behavior). Further analyses indicate that the behavioral benefits of small classes may persist and result in reduced in-grade retentions and less need for disciplinary measures.

The basic STAR results for academic achievement have been confirmed by independent analysts using other statistical approaches (e.g., Goldstein & Blatchford, 1998; Krueger, in press). The outcomes themselves have been replicated in several other settings, most notably Tennessee's Project Challenge (Achilles, Nye, & Zaharias, 1995), Wisconsin's Student Achievement Guarantee in Education (SAGE) program (Maier, Molnar, Percy, Smith, & Zahorik, 1997; Molnar, Smith, & Zahorik, 1998), and the Burke County, North Carolina, program (Achilles, Egelson, & Harman, 1995). Research teams are continuing to analyze the STAR database to answer further policy questions.

How Large Were the Effects?

The initial analyses of STAR data focused on item-response-theory (IRT) scale scores produced by the test publishers.[5] Effect sizes for the STAR reading and mathematics tests, taken from Finn (1998), are given in Table 1. Each effect size is the difference between the mean of small classes and the mean of the two other class types, divided by the standard deviation of students in regular classes; separate standard deviations were used for White and minority effect sizes. The particular con-

TABLE 1
Small Class Effect Sizes, Grades K–3

		Grade level			
Scale	Group	K (N = 5,738)	1 (N = 6,572)	2 (N = 5,148)[a]	3 (N = 4,744)[a]
Word Study Skills	White	0.15	0.16	0.11	
	Minority	0.17	0.32	0.34	
	All	**0.15**	**0.22**	**0.20**	
Reading	White	0.15	0.16	0.11	0.16[b]
	Minority	0.15	0.35	0.26	0.35[b]
	All	**0.18**	**0.22**	**0.19**	**0.25[b]**
Total Reading	White		0.17	0.13	0.17
	Minority		0.37	0.33	0.40
	All	**0.18**	**0.24**	**0.23**	**0.26**
Basic Skills First (BSF)–Reading	White		4.8%	1.6%	4.0%
	Minority		17.3%	12.7%	9.3%
	All		**9.6%**	**6.9%**	**7.2%**
Total Mathematics	White	0.17	0.22	0.12	0.16
	Minority	0.08	0.31	0.35	0.30
	All	**0.15**	**0.27**	**0.20**	**0.23**
Basic Skills First (BSF)–Mathematics	White		3.1%	1.2%	4.4%
	Minority		7.0%	9.9%	8.3%
	All		**5.9%**	**4.7%**	**6.7%**

Note. The values for BSF Reading and BSF Mathematics represent differences in the percentage passing (no standard deviation). All other values are mean differences: Small – (Regular + Aide)/2, divided by the standard deviation of the scale. Standard deviations were computed for all students in regular classes and all White and minority students in regular classes separately.

[a] Excluding pupils whose teachers received STAR training.

[b] Total Language scale in Grade 3 (not reading).

trast was chosen to maximize precision after it was discovered that there were no significant differences between regular and teacher aide classes. Effect sizes for the criterion-referenced tests are differences in the percentages of students passing the test (i.e., attaining mastery).

For all students combined, the small-class advantage in kindergarten was approximately 0.15σ to 0.18σ. The small-class advantage in first grade was approximately 0.22σ to 0.27σ. The small-class advantages in Grades 2 and 3 ranged from 0.19σ to 0.26σ.

The small-class advantage for White students was smaller than the overall effect size but statistically significant. The advantage for minority students—most of whom were African American—was larger. In most comparisons, the benefit for minority students was about *two to three times as large* as that for Whites. On the criterion-referenced tests, the small-class advantage for minority students was even more pronounced than on the norm-referenced tests: a 17% advantage in Grade 1 reading and a 7% to 10% advantage in mathematics. The impact of small classes on minority and White students reduced the achievement gap on every test (not to the detriment of either group). For example, the difference in mastery rates between Whites and minorities in Grade 1 reading was "reduced from 14.3% in regular classes to 4.1% in small classes" (Finn & Achilles, 1990, p. 568).

The effect sizes in Table 1 probably underestimate the true differences. As a result of student mobility, approximately 5% to 10% of small classes in Grades 1, 2, and 3 "drifted" above the range defined as small. Similar numbers of regular and regular with aide classes drifted downward, into the range defined as small; effect sizes would undoubtedly be larger if the out-of-range classes were omitted from the analysis. Furthermore, the comparison of small classes with the average of regular and teacher aide classes was sometimes smaller than the contrast of small with regular classes only. We are currently updating this work as well as looking for ways to portray the total impact of an intervention that affects many outcomes over many grades.

Carryover Effects

All children returned to regular-sized classes in Grade 4, and researchers in the Lasting Benefits Study continued to follow a significant portion of these pupils. Achievement scores were available through the Tennessee Comprehensive Assessment Program for Grades 4 through 8. The effect sizes in Table 2 compare students who had been in small classes with students who had been in regular classes during the preceding years. These results are drawn from other reports, particularly Finn, Fulton, Zaharias, and Nye (1989; Grade 4) and Nye et al. (1992, 1993, 1994; Grades 5–7).[6]

The findings are clear and consistent: The advantage of having been in a small class was statistically significant in every subject through Grade 7 (at least). In general, the small-class advantage carried through subsequent years, although the effect sizes were slightly diminished. As in earlier grades, no significant differences were found for students who attended classes with teacher aides.

TABLE 2
Lasting Benefits Effect Sizes, Grades 4–7 (Small Minus Regular)

	Grade level			
Scale	4 (N = 4,230)	5 (N = 4,649)	6 (N = 4,333)	7 (N = 4,944)
Total Reading	0.13	0.22	0.21	0.15
Total Language	0.13	0.18	0.14	0.15
Total Mathematics	0.12	0.18	0.16	0.14
Science	0.12	0.17	0.15	0.14
Social Science	0.11	0.17	0.15	0.10
Study Skills	0.14	0.18	0.16	0.16
Curriculum-based tests: Domains mastered				
Language Arts	0.11	0.34	0.26	0.08
Mathematics	0.16	0.28	0.17	0.08

Note. Grade 4 effect sizes were computed with standard deviations from regular classes only. Other grades used common within-cell standard deviations.

To date, the follow-up data have not been examined for differential effects by race/ethnicity or socioeconomic status.

Another Look at the Academic Gains

The effect sizes in Tables 1 and 2 are relatively stable across the grades. For example, the small-class advantages in total reading for all students were 0.24σ, 0.22σ, and 0.26σ in Grades 1, 2, and 3, respectively. In Grades 4 through 7, after students had returned to regular classes, the effects ranged from 0.13σ to 0.22σ.

The stability of effect sizes is partially a spurious result of the test publishers' IRT scaling procedures. This approach produces scores that have the same standard deviation at all grade levels. Thus, the scale scores of the SAT and CTBS batteries are developmental only to the extent that the means are allowed to increase from grade to grade. If a completely developmental scale were to be used, it would be clear that students also become more heterogeneous in actual skill levels as they progress through the grades. For example, the range of reading skills for most first-grade pupils is from "none" to about the level of a beginning third grader. The range of reading skills of ninth-grade pupils is much wider, from very little (perhaps a Grade 3 or Grade 4 level) to quite sophisticated (perhaps Grade 12 or beyond).

Grade equivalents (GEs) offer one way to view the effects in developmental terms. A GE of 3.4 for a student on test *X*, for example, means that the pupil is performing like a typical student in the 4th month (December) of Grade 3. Thus, if the student is actually in the 4th month of Grade 2, he or she is performing quite well—at the level of students with a full year (10 months) of additional schooling. If the student who took test *X* is actually in the 10th month (June) of Grade 3, he or she is performing in a manner similar to students who have had 6 fewer months of schooling. GEs can be obtained directly from tables given in test publishers' manuals or by fitting a curve of mean or median scale scores to the year and month of schooling in which the test was taken (e.g., see Shulz & Nicewander, 1997).

The use of GEs has been subject to some debate, focused mostly on the interpretation of individual students' scores (see Burket, 1984; Hoover, 1984; Peterson, Kolen, & Hoover, 1989; Yen, 1986). For example, a GE of 5.0 for a third-grade pupil does not mean that the pupil is capable of doing fifth-grade work; it means only that the score was at the median of fifth-grade pupils on this particular form of the test (not a fifth-grade form). And GEs are not appropriate for estimating "rate of growth," since the scale is tied to the month/year metric; average growth for a cohort of pupils on the GE scale is always about 1.0 GE per school year. However, GEs are a useful way to compare the means of several groups at a particular grade level. They are based on the distribution of actual or estimated performance of pupils at a particular grade level in the population, and mean differences can be interpreted in terms familiar to educators (months of schooling).

Table 3 presents effect sizes that have been reestimated in GEs. Each value is the estimated difference between the performance of students in small classes and the performance of students in regular classes on the GE scale. These values were obtained via table lookup. The mean performance of small classes and the mean performance of regular classes were each converted to the GE scale, and subtracted. In brief, Table 3 shows the following:

1. At the end of kindergarten, small-class students are about 1 month ahead of regular-class students in all subjects (actually about 0.7 to 0.9 months).
2. At the end of first grade, small-class students are about 2 months ahead of regular-class students in all subjects.
3. At the end of fifth grade, small-class students are about half a school year (5 months) ahead of regular-class students in all subjects. The effect continued despite their return to full-size classes in Grade 4.

That is, the advantage of small classes continues throughout the school years and generally increases from grade to grade. Our current work will continue to refine and extend these results.

Student Engagement in Learning

In the Grade 4 follow-up study, behavior data were collected in addition to achievement scores. Grade 4 teachers rated each pupil who had been in STAR on the 28-item Student Participation Questionnaire (Finn, Folger, & Cox, 1991). This instrument assesses specific learning behaviors ("engagement behaviors") judged by educators to be important in the classroom. The instrument yields reliable, valid measures of the effort students allot to learning, initiative taking in the classroom, and nonparticipatory behavior (disruptive or inatten-

TABLE 3
Small-Class Advantage in Months of Schooling
(Average GE of Small Classes Minus Average GE of Regular Classes)

During Project STAR (Stanford Achievement Tests)				
	Grade level			
Test	K	1	2	3
Word Study Skills	0.8	1.6	3.5	
Reading	0.7	1.5	2.0	0.9
Total Reading	0.8	1.7	2.7	5.4
Total Mathematics	0.9	2.7	2.1	3.1

Following years (Comprehensive Tests of Basic Skills)				
	Grade level			
Test	4	5	6	7
Total Reading	2.4	4.8		5.8
Total Language	3.0	4.9	6.9	
Total Mathematics	2.0	4.0	4.8	3.6
Science	2.5	5.0	6.4	8.1
Study Skills	3.0	4.1	5.3	7.2

tive-withdrawn behavior). Even after small classes had been disbanded, students who had been in these classes were rated as superior on all three scales; effect sizes were 0.12σ, 0.14σ, and 0.11σ for effort, initiative-taking, and nonparticipatory behavior, respectively, a year after pupils returned to regular classes (Finn, Fulton, Zaharias, & Nye, 1989).

Improvements in behavior are consistent with the finding that proportionally fewer students in small classes in kindergarten and Grade 1 were retained in grade (Harvey, 1993). Other research has demonstrated that disciplinary referrals are reduced in small classes (Achilles et al., 1994; Kiser-Kling, 1995).

These results are noteworthy not only because they demonstrate a carryover effect but because they describe a mechanism by which small classes may have affected pupil performance. Child development specialists have documented that behavior patterns established in the early grades tend to persevere throughout the years. If this is the case, then small-class participation in the primary grades is likely to affect a host of cognitive, affective, and behavioral outcomes in later grades.

What "Explains" the Small-Class Advantage?

Despite dozens of earlier studies, the classroom processes that distinguish small from large classes have proven elusive. For example, a well-designed study of process was conducted in Toronto, Canada (Shapson, Wright, Eason, & Fitzgerald, 1980). Teachers and students in Grade 4 classes were assigned to one of four class sizes: 16, 23, 30, or 37. In addition to achievement measures, observers recorded teacher-pupil interactions, pupil participation, pupil satisfaction, method of instruction, subject emphasis, physical conditions, use of instructional aids, classroom atmosphere, and quality of classroom activities. Additional questionnaires were administered to participating teachers and pupils.

Even with the plethora of measures, most of the findings were negative. Teachers generally had more positive attitudes in the smaller classes and were pleased with the ease of managing a small-class setting. However:

> The observation of classroom process variables revealed very few effects of class size. Class size did not affect the amount of time teachers spent talking about course content or classroom routines. Nor did it affect the choice of audience for teachers' verbal interactions. That is, . . . teachers did not alter the proportion of their time spent interacting with the whole class, with groups, or with individual pupils. (Shapson et al., 1980, pp. 149–150)

No statistical differences were found for most teacher activities, subject emphasis, classroom atmosphere, or the quality measures.

Other research, including STAR and affiliated studies, places these somewhat surprising findings in context. In general, teachers of small classes do not, de facto, alter their primary teaching strategies. Small classes are academically superior not because they encourage new approaches to instruction but because teachers can engage in more (perhaps even *enough*) of the basic strategies they have been using all along. More profound changes occur in students' behavior. The small-class setting promotes students' participation in learning, including students who would be unwilling to participate if they were part of a larger class.

On the teacher side of the equation, an Australian study (Bourke, 1986) identified instructional factors related to class size. Classes in the study ranged from 12 to 33 students. Significant correlates of class size included use of whole-class teaching (negative), amount of noise tolerated (positive), nonacademic management (positive), teacher probes after a question (negative), direct teacher interaction with students (negative), and amount of homework assigned and graded (negative). In North Carolina's Success Starts Small (Achilles, Kiser-Kling, Owen, & Aust, 1994), trained observers assessed more than 7,100 "communication events" in small and regular-sized classes. Events were classified as personal, institutional, or task oriented. The study found a greater percentage of on-task events and a smaller percentage of institutional events (e.g., discipline or organizational) in small classes relative to regular-sized classes. The results suggest that change in teaching behavior is a matter of degree: Smaller classes allow less time to be spent on classroom management and more time to be spent on instruction.

On the student side of the equation, the findings about increased pupil engagement tell an important story. Engagement behaviors are essential to school success: They are strongly correlated with pupil performance, and they explain why some students at risk succeed academically in spite of the obstacles they face; also, disengagement is found more commonly among minority or low-income students attending inner-city schools (Finn & Cox, 1992; Finn, Pannozzo, & Voelkl, 1995; Finn & Rock, 1997). If student engagement is increased in small classes, the effects are likely to be seen in both the short and the long run.

Observations of mathematics and reading lessons in 52 of STAR's Grade 2 classrooms (Evertson & Folger, 1989) confirm the behavior ratings. In mathematics, students in small classes initiated more contacts with the teacher for purposes of clarification, giving answers to questions that were open to the whole class and contacting the teacher privately for help. In reading, more students were on task, fewer students were off task, and students spent less time waiting for the next assignment.

The evidence indicates that the key to the benefits of small classes is increased student engagement in learning. In a small class, every student is on the firing line. It is difficult or impossible to withdraw from teaching-learning interactions in a small-class setting. Social psychologists have long recognized the negative relationship between group size and the participation of individuals—the principle underlying concepts such as "social loafing" and "diffusion of responsibility" (Darley & Latane, 1968; Levine & Moreland, 1998). Previous classroom research has documented the tendency of some students to retreat from active participation in class and the profound effects on academic achievement (e.g., Finn, Pannozzo, & Voelkl, 1995; Kashti, Arieli, & Harel, 1984; Veldman & Worsham, 1983). One report described an unwritten contract between students and teachers in which some students "agree" not to engage in behavior that will call attention in their direction; the contract might be paraphrased "Don't bother me and I won't bother you."

When class sizes are reduced, the pressure is increased for each student to participate in learning, and every student becomes more salient to the teacher. As a result, there is more instructional contact, and student learning behaviors are improved. Further research is needed to corroborate these conclusions. However, it is clear that the advantages are unique to the small-class setting; the feature of "smallness" makes them feasible. The same benefits were not found for teacher aide classes, which involved an increased number of adults in the classroom but not a reduced number of pupils.

Community Response to the Class Size Study

In a recent essay, Robinson (1998) argued that educational research is often dismissed because it ignores the practices and constraints educators and policymakers take to be important. Such was not the case with Tennessee's Project STAR. Class size has always been a central concern of teachers, policymakers, and parents. The design and execution of STAR met with high praise from the research community (e.g., Mosteller, 1995; Orlich,

1991), and STAR has been cited as a model for continuing experimentation on education issues (e.g., Grissmer & Flanagan, 1998; Jencks & Phillips, 1998).

It may be the combined impact of a high-profile question, an excellent research design, positive outcomes, and an appropriate political climate that caused STAR to become the impetus for class size initiatives in the United States and abroad. To date, at least 30 states have undertaken class size reduction efforts in the primary grades. The most extensive is in California, where more than $3 billion has been spent to hire additional teachers and find the required classroom space (see McRobbie, Finn, & Harman, 1998). Some states and districts contain the costs of reducing class sizes by targeting resources to urban schools where they are likely to have the greatest impact. Others have implemented small classes with little or no change in per-pupil expenditure by redeploying existing resources; examples are schools in Boston, the Downtown School in Winston-Salem, North Carolina, and schools in Rockingham, Guilford, and Burke counties in North Carolina (see Achilles, Sharp, & Nye, 1998; McRobbie, Finn, & Harman, 1998; Miles, 1995).

At the same time, some detractors attempt to dismiss the finding that smaller classes are academically beneficial and the implication that small classes should be implemented for 3 to 4 years. Although these researchers accept the validity of STAR analyses, arguments have been forwarded that (a) the findings are inconsistent with other research on the topic and (b) the results imply that small classes should be implemented for only 1 year (e.g., only in kindergarten or Grade 1). When scrutinized carefully, each of these contentions is shown to be incorrect.

Does Other Research on Class Size Show That Small Classes Are Not Beneficial?

In a recent monograph published by the Wallis Institute of Political Economy, economist Eric Hanushek (1998) concluded: "We have extensive experience with class size reduction and it has NOT worked" (p. ii), and "extensive econometric investigation [*sic*] show NO relationship between class size and student performance" (p. iii). These conclusions mirror earlier statements by the same author (e.g., Hanushek, 1996, 1997). However, even a cursory review of the research behind the conclusions reveals that they are based not on studies of class size but on studies of a different construct, pupil-teacher ratio. On the surface, the two methods of counting pupils appear deceptively similar, but they differ in significant ways, particularly in their relationships to students' achievement.

Class size is the number of students regularly in a teacher's room for whom that teacher is responsible each day of the school year. Class size is an important feature of the setting in which teachers teach and students learn. It limits the interactions that can take place, for example, the amount of attention available to any one student, the extent to which instruction can be individualized, the level and amount of disruptive behavior that can be tolerated, and more. Research on class size is predicated on the assumption that the most powerful antecedents of student outcomes are aspects of schooling proximal to the student that promote learning directly.

Pupil-teacher ratio is the ratio of the number of students in an educational unit to the number of full-time-equivalent education professionals assigned to that unit (Lewit & Baker, 1997). Rarely is the "unit" an individual classroom. Usually, pupil-teacher ratios are computed for entire schools or school districts and sometimes, as in some of the studies cited by Hanushek (1998), for entire states or nations. In addition to full-time classroom teachers, other classifications of professionals are included in pupil-teacher ratios. Teachers with classes designed to be small (e.g., special education and Title I classes) are always included. Teaching staff with no full-time classes of their own are counted as well, for example, reading specialists, music or art specialists, librarians, teachers who share responsibilities (team teach) in particular subject areas, substitute teachers, and others.

Why is the Difference Between Class Size and Pupil-Teacher Ratio Educationally Important?

Two differences between these approaches to counting pupils are significant. First, *pupil-teacher ratios do not generally describe the immediate teaching/learning setting for most students.* In fact, students in districts with low pupil-teacher ratios often spend most of their days in overcrowded classrooms. Pupil-teacher ratios are consistently lower than typical class sizes in an educational unit (e.g., Boozer & Rouse, 1995; Ferguson & Ladd, 1996; Flake, vonDohlen, & Gifford, 1995; Miles, 1995). The pupil-teacher ratio for public schools in the

United States in 1993–1994 was between 17 and 18.4 pupils per teacher, while the average class size was between 23.2 and 25.2 pupils (Lewit & Baker, 1997). The difference is attributable to the large number of teaching professionals not assigned to teach a full class of students each day. A Boston study (Miles, 1995) documented a pupil-teacher ratio of 13.2 but found that "most students spend the majority of their time in classes having more than 23 students" (p. 477).

The difference between actual class sizes and pupil-teacher ratios is more pronounced for some groups of pupils than others. Since self-contained classrooms are found mostly in elementary schools, the proportion of young students who attend large classes—even above 30—is highest in the early grades (Lewit & Baker, 1997). Large urban districts—districts with the most difficult educational task—tend to have the greatest discrepancies between class size and pupil-teacher ratio: large classes, small ratios (Boozer & Rouse, 1995). Despite small pupil-teacher ratios, Black and Hispanic students in these districts "attend schools with larger average class sizes" (p. 7). Project STAR showed the greatest benefits of reducing class sizes in inner-city schools and those serving large minority populations.

Changes in staffing patterns can have the opposite impact on class sizes and pupil-teacher ratios. In New York City schools, the size of the professional staff increased every year from 1991 to 1996, but average class sizes increased each year as well (New York State Education Department, 1997). As demands on schools increase, and as externally subsidized programs are created and expanded, the need for specialty teachers increases at a surprising rate. In Boston, it was revealed that more than 40% of teachers were working in specialty areas, including special education and bilingual programs (Miles, 1995). Lewit and Baker (1997) noted that "hiring such a large proportion of teachers to work with small numbers of students provides special services to many students but leaves regular classroom teachers with larger classes" (p. 114).

Second, the size of a class is related directly to the amount of time teachers spend on instruction and to pupils' engagement in learning. Project STAR and other studies have confirmed this connection. Larger classes present an additional burden to classroom teachers and constrain teaching/learning interactions. It is no surprise that class size is significantly related to pupils' academic performance. Confirmed by Project STAR, this connection was supported in the scores of studies of actual class size reviewed by Glass and Smith (1978), the Educational Research Service (Robinson, 1990; Robinson & Wittebols, 1986), and Slavin (1989), and continues to be replicated today.

Pupil-teacher ratio is an aggregate measure, usually computed for units larger than the individual classroom.[7] Other economists have studied pupil-teacher ratios, including some who disagree with Hanushek's conclusion of "no association with achievement" (e.g., Hedges, Laine, & Greenwald, 1994; Krueger, 1998, in press; Wenglinsky, 1997). This research generally finds weak but statistically significant relationships with test scores for a school or district. Researchers who compare class sizes and pupil-teacher ratios directly have found that class size is more strongly connected with academic achievement than is pupil-teacher ratio (e.g., Boozer & Rouse, 1995; Ferguson & Ladd, 1996).

These findings are also not surprising. Pupil-teacher ratios do not usually characterize the setting in which most students spend most of their school day. When the pupil-teacher ratio is computed for a school or district, it does not describe variation among classes within the unit or even whether some classes are very large or very small. That pupil-teacher ratio is not strongly related to students' academic performance does not refute that class size is!

Do STAR Findings Show That 1 Year of a Small Class Is Enough?

In the 1998 monograph and elsewhere, Hanushek has contended that STAR results do not show that small classes are beneficial "except perhaps at kindergarten" (1998, p. iii). The argument is advanced that the benefits of small classes found in kindergarten appear not to increase in subsequent grades, even though the published effect sizes for Grades 1–3 are larger than for kindergarten (Table 1).[8] Hanushek proposes that a value-added analysis of the data would show no additional gain after the first year. Thus, it is alleged, only 1 year of small classes is worthwhile.

This issue has profound implications for policymakers and the children who may be affected, and it must be examined carefully. An analysis of three assumptions underlying the 1-year recommendation refutes this interpretation; the third is the most telling.

First, it is assumed that effect sizes must increase over the grades in order to conclude that 2 or more years of intervention are beneficial. According to Hanushek (1998), "If resources had a continuing impact, we should observe a widening of achievement as more and more resources are applied" (p. 27). Is this assertion correct? Note that the benefits of small classes persisted throughout the experiment even though the material students learned was more complex and challenging, and the end-of-year tests were more difficult, in each successive grade. (Would we expect the best football team in the NFL not only to defeat every opponent but to outscore each successive opponent by a wider margin than the one before?) To demonstrate superior performance while facing new and more difficult challenges is itself evidence of continuing success.

Second, it is assumed that the impact of small classes remains stable (does not increase) throughout the grades. This is not the case. Because of scaling procedures used by commercial test publishers, cross-sectional analyses of data do not provide a complete picture of growth across the years. The IRT scale scores used in earlier STAR reports (see Tables 1 and 2) do not reflect increasing variability among students as they grow older; percentile scores computed within grades, used by Krueger (in press), do not reflect increasing means or variances from grade to grade. A true developmental scale reflects both. When these restrictions are lifted, it is clear that the benefits of small classes increase from year to year—both while resources are applied, and in Grade 4 and beyond when the resources are removed (see Table 3).

Third, the conclusion that 1 year of small classes is enough is not supported by any STAR results. The STAR analyses show that 3 to 4 years of small-class participation produce academic and behavior improvements that persist through Grade 7 and beyond. The experiment did not have a 1-year condition and provided *no evidence* that 1 year (or even 2 years) of small classes would produce enduring effects.

The field of education is replete with interventions that, because they were not of sufficient duration, did not have lasting benefits. Other disciplines recognize this principle as well. For example, antibiotics are prescribed for 5 (or 7 or 10) days. Although symptoms may improve on the first day, research shows that a longer regimen is needed to ensure that the infection is eradicated. It would be foolhardy to stop taking the medication after 1 or 2 days, even if additional improvement is not apparent. It is possible that fewer years of small classes would have some lasting benefits. However, the recommendation of 1 year is not founded on the current state of scientific knowledge.

Conclusion and Selected Recommendations

An experiment of the quality and magnitude of Tennessee's class size study is rare in education. That it has engendered a large number of school, district, and state initiatives is even more unique. Project STAR produced answers to questions that educators have long been asking and provided a perspective for reinterpreting previous research on the topic. We have learned that small classes in the primary grades are academically beneficial (especially for students at risk), have positive impacts on student behavior, and have benefits that last through ensuing years. Adding a full-time teacher aide to a regular-sized class, in contrast, does not affect the academic performance of the class.

A great deal remains to be learned. One question is paramount: Under what organizational and instructional conditions can the benefits of small classes be maximized? For example, are the benefits increased if small classes are employed in conjunction with other programs targeted to students having difficulty (e.g., preschool programs, full-day kindergartens, Title 1)? We know that teachers tend not to change their fundamental teaching strategies when given a small class. However, *should* they change their approaches to classroom management and instruction to take best advantage of the opportunities a small class presents?

The many schools undertaking class size reduction (CSR) initiatives can serve as natural laboratories for increasing our knowledge base. For the most part, this is not occurring. School and district leaders have been concerned with practical issues involved in getting the numbers down—not always an easy task. Our first recommendation is addressed to decision makers considering or implementing CSR efforts: Design evaluation studies that will inform us about the positive and negative experiences that accompany CSR and the potential for maximizing benefits as CSR initiatives are introduced.

Our second recommendation, addressed to policymakers and practitioners, is to base CSR efforts on what has been learned. Small classes are effective if introduced in the early grades; both the

theory of child development and findings such as those from Project STAR tell us that this is the place to start.[9] Small classes are most effective for students living in poverty; urban schools may be the best place to begin CSR initiatives. Small classes are beneficial because of their "smallness." A classroom with 40 pupils and 2 teachers, for example, cannot be expected to have the same effects on achievement as two classes each with 20 pupils and 1 teacher. Keep small classes small.

Our third recommendation is addressed to researchers, particularly those in a position to interpret the research for school personnel, policymakers, and parents: Be precise in specifying class sizes and in differentiating between class size and pupil-teacher ratio. The constructs are not the same. They represent different aspects of resource distribution among schools and should not be used interchangeably.

Notes

Portions of this article were presented at the annual meeting of the Association for Public Policy Analysis and Management, New York, October 1998. The work was supported in part by a grant from the Spencer Foundation. We are grateful to Susan Gerber for assistance with the statistical results reported in this article.

[1]Project STAR was directed by Elizabeth Word of the Tennessee Department of Education and conducted by a consortium of researchers from four Tennessee universities. The principal investigators were C. M. Achilles (University of Tennessee), H. P. Bain (Tennessee State University), J. Folger (Vanderbilt University), and J. Johnston (University of Memphis). Jeremy Finn was an external evaluator for the duration of the project.

[2]Some exceptions are explained in Finn and Achilles (1990).

[3]The extensive STAR database, comprising more than 10 years of data on approximately 12,000 pupils, continues to be analyzed to answer additional questions. Achievement data for the 4 years of experimentation (K–3) are now available on the Internet at **http://www.nashville.net/~heros/data.htm**.

[4]More recent analyses by these authors are using three-level hierarchical linear models (students within classrooms within schools).

[5]Recent analyses are examining developmental scales that have standard deviations that increase over the grades (see later discussion).

[6]The Grade 8 report contains obvious technical errors and thus is not included in Table 2.

[7]Even studies of pupil-teacher ratio at the classroom level usually involve classes in "normal" ranges of 25–30 pupils or so. In this limited range, the relationship with achievement may be attenuated.

[8]In his reanalysis of STAR data, Krueger (in press) concluded that the biggest benefit of small classes occurred in the first year of participation, whether it was kindergarten or first grade.

[9]California's recent decision to reduce class sizes in Grade 9 is a decision not based on current scientific knowledge.

References

Achilles, C. M., Egelson, P., & Harman, P. (1995). Using research results on class size to improve achievement outcomes. *Research in the Schools, 2*(2), 23–30.

Achilles, C. M., Finn, J. D., & Bain, H. P. (1997). Using class size to reduce the equity gap. *Educational Leadership, 55*(4), 40–43.

Achilles, C. M., Kiser-Kling, K., Owen, J., & Aust, A. (1994). *Success starts small: Life in a small class.* Greensboro: University of North Carolina.

Achilles, C. M., Nye, B. A., & Zaharias, J. B. (1995, April). *Policy use of research results: Tennessee's Project Challenge.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Achilles, C. M., Sharp, M., & Nye, B. A. (1998, February). *Attempting to understand the class size and pupil-teacher ratio (PTR) confusion: A pilot study.* Paper presented at the annual meeting of the American Association of School Administrators, San Diego, CA.

Angrist, J. D., & Lavy, V. (1996). *Using Maimonides' rule to estimate the effect of class size on children's academic achievement.* Unpublished manuscript, Department of Economics, Hebrew University, Jerusalem.

Boozer, M., & Rouse, C. (1995). *Intraschool variation in class size: Patterns and implications* (Working Paper No. 344). Washington, DC: National Bureau of Economic Research, Industrial Relations Section. (ERIC Document Reproduction Service No. ED 385 935)

Bourke, S. (1986). How smaller is better: Some relationships between class size, teaching practices, and student achievement. *American Educational Research Journal, 23,* 558–571.

Burket, G. R. (1984). Response to Hoover. *Educational Measurement: Issues and Practice, 3,* 15–16.

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 10,* 202–214.

Evertson, C. M., & Folger, J. K. (1989, March). *Small class, large class: What do teachers do differently?* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 265–298). Washington, DC: Brookings Institution.

Finn, J. D. (1998). *Class size and students at risk: What is known? What is next?* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. Available at **http://www.ed.gov/pubs/ClassSize/title.html**.

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., & Cox, D. (1992). Participation and withdrawal among fourth-grade pupils. *American Educational Research Journal, 29,* 141–162.

Finn, J. D., Folger, J., & Cox, D. (1991). Measuring participation among elementary grade students. *Educational and Psychological Measurement, 51,* 393–402.

Finn, J. D., Fulton, D., Zaharias, J., & Nye, B. A. (1989). Carry-over effects of small classes. *Peabody Journal of Education, 67,* 75–84.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *Elementary School Journal, 95,* 421–434.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology, 82,* 221–234.

Flake, J., vonDohlen, E., & Gifford, M. (1995). *Class size and student achievement: Is there a link?* Phoenix, AZ: Goldwater Institute.

Glass, G. V., & Smith, M. L. (1978). *Meta-analysis of research on the relationship of class size and achievement.* San Francisco: Far West Laboratory for Educational Research and Development.

Goldstein, H., & Blatchford, P. (1998). Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal, 24,* 255–268.

Grissmer, D., & Flanagan, A. (1998, November). *Improving the data and methodologies in educational research.* Paper presented at the U.S. Department of Education/RAND Conference on Analytic Issues in the Assessment of Student Achievement, Washington, DC.

Hanushek, E. A. (1996). School resources and student performance. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 43–73). Washington, DC: Brookings Institution.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19,* 141–164.

Hanushek, E. A. (1998). *The evidence on class size.* Rochester, NY: University of Rochester, W. Allen Wallis Institute of Political Economy.

Harvey, B. (1993). *An analysis of grade retention for pupils in K–3.* Unpublished doctoral dissertation, University of North Carolina, Greensboro.

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

Hoover, H. D. (1984). The most appropriate scores for measuring educational development in the elementary schools: GE's. *Educational Measurement: Issues and Practice, 3,* 8–14.

Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap.* Washington, DC: Brookings Institution Press.

Kashti, Y., Arieli, M., & Harel, Y. (1984). Classroom seating as a definition of situation: Observations in an elementary school in one development town. *Urban Education, 19,* 161–181.

Kiser-Kling, K. (1995). *Life in a small teacher-pupil ratio class.* Unpublished doctoral dissertation, University of North Carolina, Greensboro.

Krueger, A. B. (1998, March). Reassessing the view that American schools are broken. *FRBNY Economic Policy Review,* pp. 29–43.

Krueger, A. B. (in press). Experimental estimates of education production functions. *Quarterly Journal of Economics.*

Levine, J. M., & Moreland, R. L. (1998). Small groups. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 415–469). New York: McGraw-Hill.

Lewit, E. M., & Baker, L. S. (1997). Class size. *The Future of Children, 7*(3), 112–121.

Maier, P., Molnar, A., Percy, S., Smith, P., & Zahorik, J. (1997). *First year results of the Student Achievement Guarantee in Education Program.* Milwaukee: University of Wisconsin, Center for Urban Initiatives and Research.

McRobbie, J., Finn, J. D., & Harman, P. (1998). *Class size reduction: Lessons learned from experience* (Policy Brief No. 23). San Francisco: WestEd.

Miles, K. H. (1995). Freeing resources for improving schools: A case study of teacher allocation in Boston public schools. *Educational Evaluation and Policy Analysis, 17,* 476–493.

Molnar, A., Smith, P., & Zahorik, J. (1998). *1997–98 results of the Student Achievement Guarantee in Education (SAGE) program evaluation.* Milwaukee: University of Wisconsin, School of Education.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2), 113–127.

New York State Education Department. (1997). *New York: The state of learning. Statewide profile of the educational system.* Albany, NY: Author.

Nye, B. A., Boyd-Zaharias, J., Fulton, B. D., Achilles, C. M., Cain, V. A., & Tollett, D. A. (1994). *The Lasting Benefits Study: Seventh grade technical report.* Nashville: Tennessee State University, Center of Excellence for Research in Basic Skills.

Nye, B. A., Zaharias, J. B., Fulton, B. D., & Achilles, C. M. (1993). *The Lasting Benefits Study: Sixth grade*

technical report. Nashville: Tennessee State University, Center of Excellence for Research in Basic Skills.

Nye, B. A., Zaharias, J. B., Fulton, B. D., Wallenhorst, M. P., Achillles, C. M., & Hooper, R. (1992). *The Lasting Benefits Study: Fifth grade technical report.* Nashville: Tennessee State University, Center of Excellence for Research in Basic Skills.

Orlich, D. C. (1991). Brown v. Board of Education: Time for a reassessment. *Phi Delta Kappan, 72,* 631–632.

Peterson, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221–264). New York: Macmillan.

Robinson, G. E. (1990). Synthesis of research on effects of class size. *Educational Leadership, 47*(7), 80–90.

Robinson, G. E., & Wittebols, J. H. (1986). *Class size research: A related cluster analysis for decision making.* Arlington, VA: Educational Research Service.

Robinson, V. M. J. (1998). Methodology and the research-practice gap. *Educational Researcher, 27*(1), 17–26.

Shapson, S. M., Wright, E. N., Eason, G., & Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal, 17,* 141–152.

Shulz, E. M., & Nicewander, W. A. (1997). Grade equivalent and IRT representations of growth. *Journal of Educational Measurement, 34,* 315–331.

Slavin, R. E. (1989). Achievement effects of substantial reductions in class size. In R. E. Slavin (Ed.), *School and classroom organization* (pp. 247–257). Hillside, NJ: Erlbaum.

Veldman, D. J., & Worsham, M. (1983). Types of student classroom behavior. *Journal of Educational Research, 76,* 204–209.

Wenglinsky, H. (1997). *When money matters: How educational expenditures improve student performance and when they don't.* Princeton, NJ: Educational Testing Service, Policy Information Center.

Word, E., Johnson, J., Bain, H. P., Fulton, D. B., Zaharias, J. B., 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 State Department of Education.

Yen, W. M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement, 23,* 299–325.

Authors

JEREMY D. FINN is a professor of education, Graduate School of Education, State University of New York at Buffalo, 408 Christopher Baldy Hall, Buffalo, NY 14260. He specializes in classroom and school processes, educational equity, and multivariate analysis.

CHARLES M. ACHILLES is a professor of educational administration, School of Education, Eastern Michigan University, Ypsilanti, MI 48197. He specializes in educational administration.

Manuscript received January 13, 1999
Revision received February 22, 1999
Accepted February 23, 1999