Small Classes in the Early Grades:

One Policy – Multiple Outcomes

Jeremy D. Finn and Allison E. Suriani

The University at Buffalo – SUNY

Charles M. Achilles

Seton Hall University

November 27, 2007

Paper prepared for the National Invitational Conference of the Early Childhood Research Collaborative, sponsored by the University of Minnesota Center for Early Education and Development and the Federal Reserve Bank of Minneapolis: Minneapolis, MN, December 7—8, 2007. For further information, contact J. Finn at finn@buffalo.edu or (716) 645-2484 extension 1071.

Small Classes in the Early Grades:
One Policy – Multiple Outcomes

The idea that small classes are desirable for instruction is not new; indeed, it dates back at least to biblical times (Angrist & Lavy, 1996). In the 20th and 21st centuries, small classes, a distinguishing feature of American private schools and a fundamental component of all remedial programs, have been posed as a way to bolster student achievement in public schools generally.

Tennessee's Project STAR (Student Teacher Achievement Ratio) is not a 'program' in the sense of most programs discussed at this conference. It was a large-scale experiment to test the effectiveness of this basic principle of school organization, that is, arranging classes in the early grades so they have a relatively small number of students in a classroom with a full-time teacher.

Because of the reported effectiveness of small classes in STAR, the experiment has been followed by class size reduction (CSR) initiatives across the country. According to one recent survey, 33 states had statewide CSR programs for the early grades, of which 13 were mandatory (McCabe, 2006). The number of districts with their own programs is undocumented but probably quite large. President Clinton introduced a nationwide CSR program in 1998, which was discontinued by the Bush administration but remains in No Child Left Behind (NCLB) as an optional use of Title II funds.

STAR has also been followed by additional research on small classes, by several analyses of costs and benefits, and by some arguments that the research, or the conclusions drawn from the research, is faulty. This paper summarizes these

developments. Although Project STAR is emphasized, it is discussed in the context of class size research and policy more broadly. The paper has the following sections:

(a) A brief history of Project STAR;

(b) A description of the STAR experiment and its defining features;

(c) A summary of the short-term and long-term findings of STAR, and other evaluations that followed STAR;

(d) Attempts to explain how small classes benefit student achievement and behavior (the 'dynamics' of small classes);

(e) Attempts to evaluate the costs and benefits of class size reduction;

(f) A description of a new public access database containing information on the STAR students from kindergarten through high school.

Beware of Terminology: Heffalumps are not Elephants

Class size (CS) is *the number of students regularly in a teacher's classroom for whom that teacher is responsible*. Assessing class size is easy -- just count students in the room or count the names on the teacher's roster.[1] A class with 15 students and one teacher has class size of 15. A class with 28 students and one teacher has a class size of 28. A class with 28 students and 2 teachers and a full-time teacher aide has a class size of 28.[2] Class size is important to parents who want their children to attend a small class, and to teachers because it shapes their interactions with students. It is clearly related to the amount of material that can be learned – an observation supported by scientific data (see below).

[1] There may be some disagreement due to absenteeism.
[2] But a pupil-teacher ratio of about 9-to-1.

Seems simple, right? Guess again! A second "look-alike" construct, used primarily by economists, muddies the waters.

Pupil:teacher ratio (PTR) is *the ratio of the number of students in an educational unit to the number of full-time-equivalent professionals assigned to that unit.* The number of "teachers" may include teaching assistants, special education and Title I teachers, specialty teachers who do not have their own classes (e.g., music or foreign language teachers or librarians), and administrators.[3] Pupil:teacher ratios are usually computed for large educational units, for example, districts, states, or the entire nation. Many large urban districts have small PTRs (e.g., 13-to-1) because of the large number of ancillary staff members, <u>despite having overcrowded classrooms</u> (Boozer & Rouse, 2001; Lewit & Baker, 1997; Miles, 1995).

PTR does not describe the proximal setting in which teachers are teaching and students are learning. It does not play the same obvious role in any particular classroom – indeed, it may play no role at all. It is not surprising that the relationship of PTR with student achievement is weak at best (Ferguson & Ladd, 1996; Hanushek, 1989; Hedges, Laine, & Greenwald, 1994; Hedges & Greenwald, 1996).

PTRs and class size do not translate easily into one another because PTRs are usually computed for schools, districts and other larger units, obscuring variation from classroom to classroom, and because PTRs include personnel who are not responsible for classes of students on a regular or full-time basis.

[3] A ratio is obtained by division and the divisor is important.

Why is this distinction important? The distinction between PTR and class size is important because PTR data have been used erroneously to 'prove' that class size is unrelated to student achievement -- arguments echoed widely in public media. As a result, students for whom the benefits might be far-reaching may be denied the opportunities offered by small classes.

Three examples: In 1998, Hanushek published a monograph entitled "The Evidence on Class Size." The monograph summarizes studies of the relationship of PTRs with academic achievement, using data from the National Assessment of Educational Progress [NAEP] and other surveys that do not collect class size information. Further, the main results in the monograph are based on 277 estimates from econometric studies that "consider teacher-pupil ratios" (p. 21) -- a statement easily verified by examining the studies themselves. But the conclusion is: "the estimates are almost equally divided between those suggesting that *small classes* are better and those suggesting that they are worse." (pp. 21-22, emphasis added) The report has been cited and quoted by Hanushek[4] and others, including in testimony given to Congress, in the courts, and in the popular press.

Other high-profile publications also confuse the issue. A full issue of *Psychological Science in the Public Interest* was subtitled "Class Size and Student Achievement" (Ehrenberg et al., 2001a).[5] The monograph begins with the exhortation

[4] Table 6 of the monograph summarizes the results of 78 value-added studies of pupil:teacher ratios. The table is entitled "Percentage Distribution of Other Estimated Influences of Teacher-Pupil ratio on Student Performance…" (p. 24). The same table was published In a 1999 article in *Educational Evaluation and Policy Analysis* (Hanushek, 1999) but this time the title was: "Percentage Distribution of Effect of *Class Size* on Student Performance…" (Table 2, p. 148, emphasis added).

[5] A similar article by the same authors, entitled "Does Class Size Matter?" was published in *Scientific American* (Ehrenberg et al., 2001b).

that “Class size is not the same thing as the pupil/teacher ratio. Indeed it is quite different” (Ehrenberg et al., 2001a, p. 1) and notes that “class size is more closely linked to learning.” (p. 2) Yet much of the report summarizes research on pupil:teacher ratios (including a review of Hanushek’s findings). To its credit, the report also discusses Tennessee’s STAR experiment and asks the fundamental question “how can we explain the class size effects on achievement that have been reported?” (Enhrenberg et al., 2001a, p. 24). Contrary to the data on class size (and with very few data on salaries), the report concludes that funds might be better spent if used to increase teachers’ salaries (pp. 25-26).

And an overview of the correlates of human capital by Carneiro and Heckman (2003) was discussed in the *Wall Street Journal* (Heckman, 2006). The *Journal* column referenced a “cost-benefit analysis of classroom-size [sic][6] reduction on adult earnings” and concluded “While smaller classes raise the adult earnings of students, the earnings gains do not offset the costs” (p. A14). However, in the 2003 paper, the 2-1/2 pages that discuss the issue alternate between “class size” and “pupil-teacher ratio” (pp. 44-47). The report says explicitly that the cost-benefit analysis involved large-scale changes in pupil:teacher ratios. The discussion contains no reference to teaching or learning processes in classes, or whether decreasing the PTR ratio would result in fewer students in any given classroom.

The message is clear. Examine what you read about class size carefully. (The remainder of this paper is about class size.) If small classes are needed in your

[6] Class sizes are counted in numbers of persons. Classroom size is measured in square or cubic feet.

school(s), make sure that every targeted class is small. Extra effort may be required to assure that CSR reaches the classroom level.[7]

A Brief History of STAR[8]

In May of 1985, the Tennessee legislature passed House Bill (HB) 544, funding a policy study to determine the effects of class size on student achievement in the primary grades. The legislation directed that three questions be addressed:

(1) What are the effects of a reduced class size on the achievement (normed and criterion tests) and development (self-concept, attendance, etc.) of students in public elementary school grades (K-3)?

(2) Is there a cumulative effect of being in a small class over an extended time (4 years) as compared with a one-year effect for students in a small class for one year?

(3) Does effective use of teacher aides improve student performance as compared with teachers who have no special preparation for their altered conditions?

To design and conduct the study, the Tennessee State Department of Education formed a consortium of researchers from the Department, the State Board of Education, the State Superintendents' Association, and representatives from four Tennessee universities. Responsibility for direct contact with schools was delegated to the university representatives. The Consortium reviewed prior class size research as one basis for decisions about its own study: The study would begin in the earliest grades,

[7] Some states impacted by school funding litigation are implementing PTR reductions, or combinations of class size and PTR reductions, despite courts' judgments that class sizes should be reduced (e.g., New Jersey and New York).

[8] More complete histories are given in the *STAR Final Report* (Word et al., 1990), from which this section was adapted, in the *STAR Database User's Guide* (Finn, Boyd-Zaharias, Fish, & Gerber, 2007), and in Ritter and Boruch (1999). The *Final Report* and *User's Guide* are available on the website of HEROS, Inc.: www.heros-Inc.org

where small classes would be most likely to show positive effects; the small classes would have no fewer than 13 students and no more than 17 students; it would allow disaggregation of the data by school location (urbanicity), student race/ethnicity, gender, age, and socioeconomic status (SES).

Most importantly, the study was a randomized experiment in which causal connections between class size and student outcomes could be discerned. This design, often considered the gold standard of empirical research, would set STAR apart from prior (and, to date, all subsequent) research on class size, including observational studies and regression analyses of existing databases (e.g., production-function analyses).

The State paid the costs associated with the study, including the salaries of extra teachers required to reduce class sizes, and project teacher aides. The total cost of the four-year project, plus data analysis and reporting in the fifth year, was approximately $13 million.

Selection of Schools

All Tennessee school systems were invited to participate. Schools were to plan to participate in the project for four years, beginning with kindergarten in 1985-1986. All participating teachers had to be certified for the grade level they were teaching. Schools had to agree to the random assignment of teachers and students to different class conditions (i.e., class sizes).

Initially, 180 schools expressed an interest in participating, of which about 100 schools had enough kindergarten students to be eligible to participate. A minimum of

57 students was necessary, providing enough students for one class of each of three conditions (with 13, 22, and 22 students, respectively). Ultimately, 79 schools across the State were selected to participate[9] (17 inner-city, 16 suburban, 8 urban and 38 rural schools).

The STAR Experiment

The STAR experiment involved one cohort of students followed for four years – students entering kindergarten in 1985 (or those who began schooling in first grade in 1986). Within each school, all students entering kindergarten were assigned at random to one of three conditions: a small class (S) with 13-17 students, a regular class (R) with 22-25 students, or a regular class with a full-time teacher aide (RA) and 22-25 students. Teachers were assigned at random to the classes.[10]

The randomization of students among conditions meant that each class was comprised of a cross-section of students from the school population. Class arrangements were maintained all day (including kindergarten classes), all year long. The only interventions in STAR were class size and teacher aides. No other curricular changes or school policies were imposed on the STAR classes or teachers. All other school services, including special education programs, functioned as usual.

In total, 128 small classes, 101 regular classes, and 99 regular-aide classes were formed in kindergarten. Since kindergarten was not mandated in Tennessee at the

[9] Approximately 6,300 students from the 79 schools participated in STAR in the kindergarten year.
[10] The randomization was monitored regularly by the STAR Consortium. One school could not maintain the randomized classes throughout and withdrew from STAR after kindergarten.

time, a substantial number of students joined STAR when they entered first grade. They were assigned at random to the three conditions.

Once assigned to a class type, students remained in the assigned class type as long as they were in the experiment. A new teacher was assigned at random to the class in each subsequent grade. Students moving into STAR schools from non-STAR schools during the four-year experiment were assigned at random to one of the class types, with the constraint that small classes could not exceed 17 students. Students moving from one STAR school to another were assigned to the same type of class as they had participated in previously.

Students with the longest duration participated from kindergarten (1985-1986) through grade 3 (1988-1989). In all, 27% of the 11,601 STAR students participated for four consecutive years, and 42% remained in the experiment for 3 or more years.[11] Students who participated in STAR for one, two, or three years made it possible to study different patterns of small-class participation.

Measures

Academic achievement (norm-referenced and criterion-referenced) and self-concept/motivation were measured in the spring of each year (1986--1989). The norm-referenced tests were the Stanford Achievement tests (SATs); the criterion-referenced tests were the Basic Skills First (BSF) tests, developed by the Tennessee State Department of Education. Self-concept and motivation were assessed with the SCAMIN instrument (Milchus, Farrah, & Reitz, 1968). Additional data included logs of

[11] Students who were retained in grade or who moved out of STAR schools participated for fewer years.

time use, interviews with teachers and aides, and responses to the Teacher Problem Checklist (Cruickshank, 1980).

The experiment ended in the spring of 1989, when most students completed third grade. In fourth grade, all students returned to full-size classes. Researchers continued to collect data on the STAR students in subsequent grades. Five stages of additional data collection were undertaken:

- Academic achievement test scores were obtained annually in grades 4—8; scores were provided by the State Department of Education on the Comprehensive Tests of Basic Skills (CTBS) and on later-grade grade versions of the State's BSF tests.
- Teachers' ratings of students' classroom participation and behavior were obtained in grades 4 and 8, using the Student Participation Questionnaire (SPQ; Finn, Folger, & Cox, 1991).
- Students' identification with school was assessed in grade 8, using the Identification with School Scale (Voelkl, 1996).
- College-entrance examination scores were obtained for STAR students;[12]
- Complete high-school transcripts were obtained for approximately 5,300 STAR students, which included courses taken, grades, and graduation/dropout.

The primary data analyses were conducted by the STAR Consortium; secondary analyses have been conducted by other teams. A number of questions remain, and analyses using the STAR data continue today (see "Continuing Research" below).

The particular strengths of Project STAR

The STAR experiment had several strengths: (a) The within-school randomized experiment permitted causal conclusions to be drawn, and also controlled for

[12] This work was spearheaded by Princeton economist Alan Krueger.

differences among schools, e.g., populations served, resource availability, teacher morale, school size, and others; (b) It built on prior research findings about the timing of small classes and about recommended class sizes; (c) It was extensive, involving over 6,000 students each year, and almost 12,000 students in the 4-year span; and (d) The students were assessed regularly during the experimental years (K—3) and were followed through high school.

Short- and Long-term Findings

The most in-depth analyses of academic achievement are given in the STAR Final Report (Word et al., 1990), and in Finn and Achilles (1990), Krueger (1999), and Finn, Gerber, Achilles, and Boyd-Zaharias (2001).[13] Books and articles that review the STAR outcomes and other class size research in non-technical terms include Achilles (1999), Biddle and Berliner (2002), Finn (2002), Wang and Finn (2000), and Mosteller (1995).

Short-term outcomes

The short-term findings of STAR:

(1) STAR demonstrated that small classes have academic benefits in every grade, K—3, in every subject tested. The magnitude of the benefits are summarized in Table 1, both as effect sizes and in months of schooling. Overall effect sizes were in the range 0.20 to 0.30 standard deviations, which translated to benefits of ½ to 5-1/2 months of schooling (see Finn et al., 2001).

(2) Starting early and continuing in small classes for multiple years are important. Table 1 shows the greatest benefits for students who entered small classes in kindergarten and remained in small classes for 2 or more years—in general, the more the better.[14]

[13] Other publications, some listed in the following pages, address particular questions, e.g., the effects of teacher aides, the "whys" of small classes. The STAR *Database User's Guide* (Finn et al., 2007) includes a more complete bibliography.

[14] The analysis used a multitude of socioeconomic control variables to reduce the confounding with student mobility.

(3) In each grade, there were significant interactions with race/ethnicity or with school urbanicity.[15] The benefits of small classes were greater for minority students or students attending inner-city schools than for white students in non-urban schools. In terms of effect sizes, the advantages for minorities were often two to three times as great, thus reducing the white/minority achievement gap. This finding has led to recommendations that small classes should be "targeted" to schools serving low-income and minority students.

These findings have been corroborated in non-experimental research. Wisconsin's Project SAGE reported effect sizes similar to those in Table 1 (Molnar, Smith, & Zahorik, 1999; Molnar, et al. 2000). California's statewide CSR, evaluated mainly in third grade, yielded effect sizes similar to those of STAR for grade-3 students who had been in small classes for 1 or 2 years (see CSR Research Consortium, 2000). Benefits have been documented in evaluations in other high-poverty communities (Achilles, Harman, & Egelson, 1995; Achilles, Nye, & Zaharias, 1995), and in analyses of several large-scale databases, for example, with first-grade students (NICHD Early Child Care Research Network, 2004) and in a national sample of kindergarten students (Walston, West, & Rathbun, 2002).

In addition to achievement gains, STAR researchers found:

(4) Fewer small-class students were recommended for in-grade retention (being "held back") in grades 1, 2, and 3.[16] The differences were statistically significant in grade 1, where retentions are more common (7.5% compared to 11.9% in full-size classes), and in grade 3 (3.1% compared to 4.6%).

(5) No academic benefits of teacher-aide classes in any subject matter in any grade, *no matter what functions the aides performed in the classroom* (Boyd-Zaharias & Pate-Bain, 1998; Finn, Gerber, & Achilles, 2000; Gerber, Finn, Achilles, & Boyd-Zaharias, 2001).[17]

[15] Not shown in the Table.
[16] Retention data were not collected at the end of kindergarten.
[17] Note tha the addition of a teacher aide to a classroom decreases the PTR but does not alter the class size.

Long-term Outcomes

The academic benefits continued in all subsequent grades, after students returned to full-size classes (see Table 2 for effect sizes). In analyses of test scores for grades 4, 6, and 8:

(1) Students who had attended small classes in K—3 scored significantly better on all achievement tests in grade 4, on all tests except science in grade 6, and on four of six tests in grade 8.[18]

(2) The strength and duration of carryover effects depended on the number of years students had attended small classes. In any given grade (4, 6, or 8), the effects were mostly nonsignificant for students who had attended small classes for one year, marginally significant for students who had been in small classes for 2 years, and consistently significant for students who had been in small classes for 3 or 4 years.

(3) The interactions with socioeconomic status and race, found in the earlier grades, were no longer statistically significant. That is, the carryover effects were similar for students from all backgrounds.

Three analyses examined other (non-achievement) high school outcomes. The first used data on college entrance examinations (SATs and ACTs) to see if students who attended small classes in K-3 were more likely to take the exams than were students who had attended full-size classes (Krueger & Whitmore, 2001). The authors argued that taking college entrance examinations reflected aspirations to attend postsecondary schooling. The study found a significant increase in test-taking for students who attended small classes. The benefit for Black students was substantially greater than for white students, reducing the Black-white gap in college-entrance-test taking by 54%.

[18] Unfortunately, standardized testing in Tennessee ended in grade 8.

A second study examined enrollment in advanced course work (Finn, Fox, McClellan, Achilles, & Boyd-Zaharias, 2006). Using high school transcripts of STAR participants, researchers quantified the number and level of courses taken in mathematics, science, and foreign language. Small-class participation had a significant positive impact on the amount of foreign language taken, and the highest levels taken in foreign languages and mathematics. Although the effect sizes were small, the greatest course-taking benefits accrued to students who spent 3 or more years in small classes in grades K-3. No differential effect by socioeconomic status (SES) was found.

The third study examined graduation/dropout rates of students who had participated in STAR (Finn, Gerber, & Boyd-Zaharias, 2005). Graduation/dropout information was available for 4,948 STAR participants through high schools or State Education Department records. The small-class effect on graduation rates was greater with each additional year of small-class participation. For all students combined, the effects of attending small classes for 4 years increased the odds of graduation by about 80%. For students from low-income homes, the effect of attending small classes for 3 years increased the odds of graduating by approximately 67%, and for 4 years more than doubled the odds. The rates for low-income students with 3 or more years of small class participation were as high as those of higher-income students. These results, like those of Krueger and Whitmore (2001), support the recommendation that small classes be targeted to schools serving low-income or minority students.

Secondary Analyses of the Achievement Data

Other groups of researchers have reanalyzed the STAR achievement data using a variety of statistical approaches. The short-term effects were reexamined in Goldstein

and Blatchford (1998), Hedges, Nye, and Konstantopoulos (2000), and Krueger (1999). Long-term effects were reported in Hedges, Nye, and Konstantopoulos (1999, 2004) and Krueger and Whitmore (2001). Both general and specific results of these analyses were similar to the original STAR findings.

The Dynamics of Small Classes

Two questions are asked often: "What 'frame conditions' have to be present to realize the academic benefits of small classes?" and "How do small classes affect student performance?" Answers to the first question have come from the STAR design and findings. Answers to the second question have proven to be more elusive.

The Frame Conditions

The STAR experiment embodied principles that characterize interventions with sustained effects (Ramey & Ramey 1998):

(1) *Start early and continue*. STAR students began in kindergarten and continued in small classes through successive grades. Students who entered small classes in kindergarten had the greatest immediate impact on performance. Those who entered small classes in kindergarten and continued for three or four years had the strongest carryover effects (Finn et al., 2001). CSR programs that begin in later grades or have shorter duration (e.g., one year) are less likely to have the same benefits.

(2) *Program intensity*. STAR small classes were maintained throughout the day for the full school year. Kindergartens were full-day kindergartens.

Other conditions, if absent, could dilute the benefits of small classes:

(3) *Heterogeneous classes.* Through random assignment, each STAR class had a cross-section of students in a particular grade level in a school. Experience suggests that if the class is comprised only of students who are difficult to teach, the same benefits may not be realized. Strict randomization may not be necessary, but heterogeneity was an important feature of the STAR classes.

(4) *Cohort effects.* In the STAR experiment, the same class grouping was maintained across the years. Krueger and Whitmore's (2001) analysis suggests that this contributes to the impact of small classes.

(5) *Qualified teachers.* STAR demonstrated that certified teachers, in general, produced greater learning gains with small classes than with larger classes. We do not argue that small classes are the only factor that affects academic achievement, nor do we believe that small classes can compensate for poor teaching.

How Do Small Classes Affect Student Performance?

There are few data that show *how* small classes affect teaching and student achievement. The most common hypothesis is that teachers of small classes provide more individualized instruction. Teachers often report that small classes help them do this, but systematic interviews and classroom observations have not supported this perception (Evertson & Folger, 1989; Shapson, Eason, Wright, & Fitzgerald, 1980).

A broader argument has been made that "[S]mall classes let teachers use effectively what they already know…but large classes stifle or negate the effective use

of good teaching methods" (Achilles, 1999, p. 112). That is, small classes permit teachers to be better teachers. In the report of the National Commission on Teaching and America's Future, Darling-Hammond (1998) argued that "school reform cannot succeed unless it focuses on creating the conditions in which teachers can teach and teach well" (p. 6). Many of the conditions listed (pp. 6-11) characterize small classes.

Some evidence supports this argument. Studies have found that teachers of small classes spend more time on instruction and less on classroom management or matters of discipline (Achilles, Kiser-Kling, Aust, & Owen, 1995; Blatchford, 2003; Bourke, 1986; Molnar, Smith, & Zahorik, 1999). And data show consistently that teachers of small classes have more time to "listen to children, to get to know their personal lives and concerns" (Johnston, 1990, p. 12; also Finn, Forden, Verdinelli, & Pannozzo, 2001; Kiser-Kling, 1995).

However, the most visible and obvious changes when class sizes are reduced are not in teachers' behavior, but in *students' behavior*. The review by Finn, Pannozzo, and Achilles (2003) summarized theory and data to support this view. In the studies reviewed, 38 out of 42 measures of learning behavior indicated that students in small classes were more engaged in learning, and 17of 24 measures of pro- and antisocial behavior indicated that students in small classes were better behaved. With only one exception (in 66 measures), no significant difference favored large classes.

The authors discussed two principles that explain these effects. The first, visibility of the individual, is also termed the "firing line hypothesis:" In a small class, each student feels continuous pressure to participate; there are no 'back corners' in which to hide. Moreover, the teacher can't easily ignore any particular student. The second

principle is termed “sense of belonging.” Small classes are more cohesive groups than are large classes; there is a smaller proportion of nonparticipants, and splinter groups are rare. Further, students tend to be more supportive of one another and develop a stronger sense of identification with their classmates and with the group as a whole; this phenomenon has been termed “psychological sense of community” (Sarason, 1974).

Only 19 studies (with 66 different measures) were found that examined these principles in small and large classes. The studies were very diverse and varied in quality, but included several that were well designed and well conducted. In sum, observations show that student behavior changes, sometimes dramatically, when class sizes are reduced, but more research is needed on the mechanisms by which learning is affected.

The Economics of Class Size Reduction (CSR)

Several approaches have been taken to evaluating the costs and benefits of small classes. Simplistic evaluations are entirely cost-oriented; they assume that CSR can only be accomplished by increasing funds to add teachers and facilities to the budget. For better or worse, California accomplished its statewide reduction of K-3 class sizes in a matter of months by the rapid expenditure of about $1 billion (Bohrnstedt & Stecher, 1999). From this perspective, it is little surprise that decision makers view class size reduction as expensive and off-putting. It has also been shown that CSR can be accomplished by reallocating existing resources within a district, with little or no additional cost per pupil (see Achilles, 2005).

Other evaluations included the benefit side of equation. Some benefits accrue to the schools, and many are realized by the larger community, for example, the savings

associated with reduced dropout rates, reduced health care or welfare costs, or the increased earning power of higher-achieving students. School administrators may have difficulty taking these factors into account, especially when the benefits occur after students leave school.

Simplistic View of the Costs

In 1999, Brewer, Krop, Gill, and Reichardt estimated the operational costs, such as additional teachers, classrooms, and supplies that would be needed for nationwide CSR programs under various program alternatives. Alternatives included the degree of class size reduction (20, 18, or 15 students), the grades involved, and whether all students would be part of the program or just "targeted" students. Using data from several national databases, Brewer et al. (1999) modeled possible CSR programs in grades 1-3 for school years 1998/1999 through 2007/2008. At the high end, they estimated the cost of reducing class size to 15 for all students in grades 1-3 in the U. S. would be $11 billion. However, if only targeted students were included in the program, such as those with free and reduced-price lunches, the total cost dropped to approximately $2 billion. In either case, this analysis was based on the assumptions that operational costs were universal throughout the nation, and that no existing resources could be reallocated to support CSR.

On the other hand, it has been demonstrated that CSR can be implemented at much lower costs. In Burke County, North Carolina, a successful CSR program was implemented in grades 1-3 by creatively using funds, classroom space, and personnel (Achilles, Harman & Egelson, 1995; Egelson, Harman, Hood, & Achilles, 2005). Classrooms that housed 6th graders (who were moved to the middle school) were

converted to elementary rooms, mobile units were added, and, in some cases, older schools were reopened and remodeled. Through attrition, salaries of teaching assistants were used to fund teacher positions. The district's per-pupil expenditure remained nearly the same even with small classes and increased student achievement.

While the reallocation of resources to reduce class sizes was effective in a mid-sized, highly committed school district, it remains to be seen if such policies would encounter significant obstacles if attempted in other types of communities or on a larger scale. In either case, this work demonstrated that existing resources can be used to control the costs of CSR.

Academic Achievement, Dropping Out, and Grade Retention

More complete analyses have considered the cost savings produced by CSR. In an analysis of data from Project STAR, Krueger (2003) weighed the costs of CSR against the lifetime earnings of students. Costs were estimated based on current per-pupil expenditures and the number of additional classes that would be required. Benefits were based on previous research showing a statistical relationship between academic achievement and wages (Neal & Johnson, 1996).[19] Krueger (2003) used this relationship to calculate the internal rate of return of reduced size classes under a variety of economic scenarios (e.g., different rates of economic growth during the work life of the students).

[19] Neal and Johnson (1996) estimated that a 1-standard deviation increase in test scores translates into a 0.2-standard deviation increase in wages, a value accepted by most economists. On this basis, Krueger (2002) used STAR data to estimate that the average effect of attending small classes would translate to an 8% increase in earnings.

The analysis showed that reducing class size from 22 to 15 for grades K-3 would yield an internal rate of return of approximately 6% (Krueger, 2003). Assuming a 4% interest rate, the benefits of reducing class size to 15, in terms of lifetime earnings, would be *43% greater than the costs*, and 100% greater than the costs if real wages grow by 1 percent per year.[20] More recent analyses have found similar results, with estimated rates of return from class size reduction from 5% to 10%, with earnings exceeding the investment by $2,900 to $48,335 (Schanzenbach, 2006/2007). In 14 out of 15 economic scenarios shown by each author, the benefits in earnings were greater than the costs of class size reduction -- even without considering other types of payoff (e.g., reduced crime or welfare costs).

A wider range of outcomes was evaluated by Levin, Belfield, Muennig, and Rouse (2007), based on the impact of small classes on high school graduation rates. Potential savings were based on the fact that each adult cohort of 20 year olds includes approximately 700,000 dropouts. Levin et al. (2007) estimated the fiscal consequence of these dropouts to be $148 billion in lost tax revenues and additional public expenses over a lifetime.

The cost of class size reduction in grades K-3 was estimated to be $143,600 per expected high school graduate.[21] The benefits of increased graduation rates were estimated using multiple data sources. Graduates have higher earnings than do dropouts, which translate into higher tax revenue; lower crime rates and lower judicial costs;[22] lower public assistance payments; and less Medicaid and Medicare expenses

[20] Assuming that people would be in the workforce from ages 18 to 65.
[21] We believe this to be an overestimate, but the basic conclusions remain the same.
[22] Dropouts account for an estimated 50% of the state prison population (Bonczar, 2003).

(Levin et al., 2007). Graduates also enjoy improved health and lower mortality. Muennig and Woolf (2007) estimated that the health status of an average 45 year old graduate is comparable to that of an average 25 year old drop out.

Taking these factors into account, Levin et al. (2007) estimated that the average lifetime economic benefit per expected high school graduate is $209,100.[23] Even with a high estimate of the cost of CSR, this translates into a benefit to the community of $65,500 per dropout. Whether the estimates are precise or not, Levin et al. (2007), like Krueger (2003), concluded that class-size reduction produces significant economic benefits.

One of the most consistent precursors of dropping out (together with poor academic achievement) is in-grade retention (Goldschmidt & Wang, 1999; Kaufman & Bradby, 1992; Rumberger, 1995). This practice carries with it its own costs and also impacts the costs associated with dropping out. Shepard and Smith (1990) estimated that U.S. school districts spent almost $10 billion on the costs of retaining students—dollars spent on extra materials, classroom space, and personnel. In 1988/1989 in California alone, the cost of retaining 34,000 students in grades 1-6 was estimated at $212 million (Schwager, Mitchell, Mitchell, & Hecht, 1992). Class-size reduction has been shown to reduce rates of retention. To our knowledge, the savings that result from reducing retentions have not been factored into any cost-benefit analysis of CSR.

[23] This amount varies by gender and race with savings from $196,300 to $268,500 for males and from $143,000 to $174,600 for females.

An Additional Factor: Small Classes as Incentives

Although not considered in any cost-benefit analysis to date, small classes may be an incentive to teachers because of improved working conditions and increased student outcomes. Several surveys have documented this point. In 1989, 234 STAR teachers were asked to choose among three possibilities, a small class, a full-size class with an aide, or else a $2,500 bonus.[24] Of teachers who had been teaching small classes, 81% said they preferred a small class over a full-size class with an aide, and 70% chose the small class over the salary increase. Of teachers of teacher-aide classes, 56% preferred a small class to a full-size class with an aide, and 63% gave preference to a small class over the salary increase (Word et al., 1990, Table X-5). More recently, Howell, West and Peterson (2007) conducted the *2007 Education Next – PEPG survey* "What Americans think about their schools." One question asked: "Which do you think is a better use of our education dollars, increasing teacher salaries or decreasing class size?" (p. 24). Of all respondents, 23% chose salary increase and 77% chose decreasing class size. Among teachers, 19% chose salary increase and 81% chose class size reductions.

The economic and policy implications of findings such as these are considerable. Small classes may be one of the teaching incentives called for by Darling-Hammond (1998), Ehrenberg et al. (2001a), Hanushek (1998), and others, that promise to reduce teacher turnover and attract qualified candidates to the profession.

[24] A substantial sum in 1989.

Summary

Economists have long been interested in resources provided to schools and how those resources are related to student achievement. Analyzing the costs and benefits of educational policies such as class size reduction has sparked many debates in the education and economic communities. A full range of benefits, whether in terms of increased achievement, higher graduation rates, or the potential for improved health and lower crime rates, must be considered in any complete evaluation of large-scale interventions.

Continuing Research: A 13-year Public Access Database

A host of questions about small classes remains unanswered. For example, we know of no research that addresses the questions:

- Can small class benefits be increased if used in conjunction with other interventions (e.g., remedial reading programs; full-day kindergarten in place of half-day programs)?
- Will small classes in K-3 reduce or prevent the "fade effect" found in follow-up studies of effective preschool programs (including Head Start)?

To assist with continuing research, Project STAR has generated an extraordinary data set on almost 12,000 students. The data can be used to address policy issues and basic questions about child and adolescent development, and for developing statistical methodology.

With support from the W. T. Grant Foundation, the data have been cleaned and compiled into SPSS files now available on a public-access website.[25] A *User's Guide* is available in electronic and hard-copy form[26] (Finn, Boyd-Zaharias, Fish, & Gerber, 2007); it also includes a bibliography of studies conducted with the STAR data.

The files contain information on 11,601 students who participated in the experiment for at least one school year, their teachers and schools. This includes:

- Demographic variables;
- School and class identifiers;
- School and teacher background information;
- Experimental condition ("class type");
- Norm-referenced and criterion-referenced achievement test scores administered annually;
- Motivation and self-concept scores.

Data from follow-up studies include:

- Achievement test scores for the students when they were in grades 4 – 8, obtained from the Tennessee State Department of Education;
- Teachers' ratings of student behavior on the SPQ in grades 4 and 8;
- Students' self-reports of school engagement and peer effects in grade 8;
- Course taking in mathematics, science, and foreign language in high school, obtained from student transcripts;
- SAT/ACT participation and scores, obtained from ACT, Inc. and from Educational Testing Service;
- Graduation/dropout information, obtained from high school transcripts and the Tennessee State Department of Education.

Additional files include:

(1) Student data on 1780 students in grades 1 – 3 in 21 comparison schools, matched with STAR schools but not participating in the experiment;

(2) A school-level file with information about each of the 80 STAR elementary schools;

[25] The data files and documentation are available at www.heros-inc.org/data.htm; additional information may be obtained from STARDATA@heros-inc.org or by contacting the authors of this paper.
[26] At no cost.

(3) A school-level file with information about each high school attended by STAR students.

The database has already been used for a number of ancillary studies. We hope that it will continue to be a valuable resource to the research community.

References

Achilles, C. M. (1999). *Let's put kids first, finally: Getting class size right.* Thousand Oaks, CA: Corwin Press.

Achilles, C. M. (2005). *Financing class size reduction.* Greensboro, NC: University of North Carolina at Greensboro School of Education, Southeast Regional Educational Laboratory (SERVE).

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

Achilles, C. M., Kiser-Kling, K., Aust, A., & Owen, J. (1995). *A study of reduced class size in primary grades of a fully Chapter-I eligible school: Success starts small.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

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

Angrist, J. D., & Lavy, V. (1996). Using Maimonides' rule to estimate the effect of class size on children's academic achievement. *Quarterly Journal of Economics, 114*, 533-535.

Biddle, B. J., & Berliner, D. C. (2002). Small class size and its effects. *Educational Leadership, 59*(5), 12-23.

Blatchford, P. (2003). A systematic observational study of teachers' and pupils' behaviour in large and small classes. *Learning and Instruction, 13*, 569-595.

Bohrnstedt, G. W., & Stecher, B. M. (Eds.). (1999). *Class size reduction in California: Early evaluation findings, 1996-1998. (CSR Research Consortium, Year 1 Evaluation Report).* Palo Alto, CA: American Institutes for Research.

Boozer, M., & Rouse, C. (2001). Intra-school variation in class size: Patterns and implications. *Journal of Urban Economics, 50*, 163-189.

Bonczar, T. P. (2003). *Prevalence of imprisonment in the U.S. population, 1974- 2001, Special Report, NCJ 197976*. Washington DC: U.S. Department of Justice, Bureau of Justice Statistics.

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

Boyd-Zaharias, J., & Pate-Bain, H. (1998). *Teacher aides and student learning: Lessons from Project STAR.* Arlington, VA: Educational Research Service.

Brewer. D., Krop, C., Gill, B. P., & Reichardt, R. (1999). Estimating the costs of national class-size reductions under different policy alternatives. *Educational Evaluation and Policy Analysis, 21*, 179-92.

Carneiro, P., & Heckman, J. J. (2003). *Human capital policy*. Cambridge, MA: National Bureau of Economic Research Working Paper 9495. Retrieved from http://papers.nber.org/papers/w9495.

Cruickshank, D. R. (1980). *Teaching is tough*. Englewood Cliffs, NJ: Prentice Hall.

Darling-Hammond, L. (1998). Teachers and teaching: Testing policy hypotheses from a National Commission Report. *Educational Researcher, 27*, 5-15.

Egelson, P., & Harmon, P., Hood, A., & Achilles, C. M. (2005). *How class size makes a difference.* Greensboro, NC: University of North Carolina at Greensboro School of Education, Southeast Regional Educational Laboratory (SERVE).

Ehrenberg, R. G., Brewer, D. J., Gamoran, A. & Willms, J. D. (2001a). Class size and student achievement. *Psychological Science in the Public Interest, 2*, 1-29.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A. & Willms, J. D. (2001b). Does class size matter? *Scientific American, 385*(5), 32-40.

Evertson, C. M., & Folger, J. K. (1989). *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. (2002). Small classes in American schools: Research, practice, and politics. *Phi Delta Kappan, 83*, 551-560.

Finn, J. D., & Achilles, C. A. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*, 557-577.

Finn, J. D., Boyd-Zaharias, J., Fish, R. M., & Gerber, S. B. (2007). *Project STAR and Beyond: Database UsersGuide*. Lebanon, TN: HEROS, Inc. Retrieved from http://www.heros-inc.org/data.htm.

Finn, J. D., Gerber, S. B., & Achilles, C. A. (2000). Teacher aides: An alternative to small classes? In M. C. Wang & J. D. Finn (Eds.). *How small classes help teachers do their best* (pg 131-173). Philadelphia, PA: Temple University Center for Research on Human Development and Education and the U.S. Department of Education.

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

Finn, J. D., Forden, M. A., Verdinelli, S., & Pannozzo, G. M. (2001). *Evaluation of the class size reduction initiative—Buffalo Public Schools.* Buffalo, NY: University at Buffalo, Graduate School of Education.

Finn, J. D., Fox, J. D., McClellan, M., Achilles, C. M., & Boyd-Zaharias, J. (2006). Small classes in the early grades and course taking in high school. *International Journal of Education Policy and Leadership, 1*, 1-13. Retrieved from http://www.ijepl.org.

Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record, 103*, 145-183.

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology, 97*, 214-223.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. *Review of Educational Research, 73*, 321-368.

Gerber, S. B., Finn, J. D., Achilles, C. M., & Boyd-Zaharias, J. (2001). Teacher aides and students' academic achievement. *Educational Evaluation and Policy Analysis, 23*, 123-143.

Goldschmidt, P., & Wang, J. (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal, 36,* 715-738.

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

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45-65.

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

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*, 143-163.

Heckman, J. J. (2006, January 10). Catch 'em young. *Wall Street Journal*, p. A14.

Hedges, L. V., & Greenwald, R. (1996). Have times changed? The relationship between school resources and student performance. In G. Burtless (Ed.) *Does money matter? The effect of school resources on student achievement and adult success* (pp. 74-92). Washington, DC: Brookings Institution.

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*(3), 5-14.

Hedges, L. V., Nye, B., & 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.

Hedges, L. V., Nye, B., & Konstantopoulos, S. (2000). The effects of small classes on achievement: The results of the Tennessee class-size experiment. *American Educational Research Journal, 37*, 123-151.

Hedges, L. V., Nye, B., & Konstantopoulos, S. (2004). Do minorities experience greater lasting benefits from small classes: Evidence from a five year follow up of the Tennessee class size experiment. *Journal of Educational Research, 97*, 94-100.

Howell, W. C., West, M. R., & Peterson, P. (2007). What Americans think about their schools: The 2007 Education Next – PEPG survey. *Education Next, 7*(4), 13-26.

Johnston, J. M. (1990). *What are teachers' perceptions of teaching in different classroom contexts?* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (Eric Document Reproduction Service No. ED 320 867)

Kaufman, P., & Bradby, D. (1992). *Characteristics of at-risk students in NELS:88.* (National Center for Education Statistics Report No. 92-042). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Kiser-Kling, K. (1995*). A comparative study of life in first grade classrooms of 1:14 and 1:23 teacher-pupil ratios.* Unpublished doctoral dissertation, University of North Carolina at Greensboro.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114*, 497-532.

Krueger, A. B., Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *The Economic Journal, 111*, 1-28.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal, 113*, 34-63.

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

Levin, H., Belfield, C., Muennig, P., & Rouse, C. (2007). *The costs and benefits of an excellent education for all of America's children*. Retrieved from the Columbia University Web site: http://www.cbcse.org/media/download_gallery/

McCabe, M. (2006, January 4). State of the states*. Education Week, 26*(17), 72-76.

Milchus, N., Farrah, G., & Reitz, W. (1968). *The self-concept and motivation inventory: What face would you wear?* Dearborne Heights, MI: Person-0 Metrics.

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

Molnar, A., Smith, O., & Zahorik, J. (1999). *1998-99 evaluation results of the Student Achievement Guarantee in Education (SAGE) program*. Milwaukee, WI: University of Wisconsin School of Education.

Molnar, A., Smith, O., & Zahorik, J., Ehrle, K., Halback, A., & Kuehl, B. (2000). *1999-2000 evaluation results of the Student Achievement Guarantee in Education (SAGE) program.* Milwaukee, WI: 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*, 113-127.

Muennig, P., & Woolf, S. H. (2007). Health and economic benefits of reducing the number of students per classroom in US primary schools. *American Journal of Public Health, 97*(11), 2020-2027.

Neal, D., & Johnson, J. (1996). The role of premarket factors in black-white wage differentials. *Journal of Political Economy, 104*, 869-895.

NICHD Early Child Care Research Network (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes? *Developmental Psychology, 40*, 651-664.

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist, 53*, 109-120.

Ritter G. W., & Boruch, R. F. (1999). The political and institutional origins of a randomized controlled trial on elementary school class size: Tennessee Project STAR. *Educational Evaluation and Policy Analysis, 21*,111-125.

Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal, 32,* 583-625.

Sarason, S. B. (1974). *The psychological sense of community: Prospects for a community psychology.* San Francisco, CA: Jossey-Bass.

Schanzenbach, D. W. (2006). What have researchers learned from Project STAR? *Brookings Papers on Education Policy: 2006-2007*. Washington, DC: Brookings Institution, 205-228.

Schwager, M. T., Mitchell, D. E., Mitchell, T. K. & Hecht, J. B. (1992). How school district policy influences grade level retention in elementary schools. *Educational Evaluation and Policy Analysis, 14*, 421-438.

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

Shepard, L. A., & Smith, M. L. (1990). Synthesis of research on grade retention. *Educational Leadership, 47*(8), 84-88.

Voelkl, K. E. (1996). Measuring students' identification with school. *Educational and Psychological Measurement, 56*, 760-770.

Walston, J., West, J., & Rathbun, A. (2002). *Kindergarten program features and gains in reading and mathematics achievement of public school children: Class size, length of school day and classroom aides.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Wang, M. C., & Finn, J. D. (Eds). (2000). *How small classes help teachers do their best.* Philadelphia, PA: Temple University Center for Research on Human Development and Education and the U.S. Department of Education.

Word, E., Johnston, J., Bain, H. P., Fulton, D. B., Boyd-Zaharias, J., Lintz, M. N., Achilles, C. M., Folger, J., & Breda, C. (1990). *Final Report: Student/teacher achievement ratio (STAR): Tennessee's K-3 class-size study.* Nashville, TN: Tennessee State Department of Education. Retrieved from http://www.heros-inc.org.