

The Tennessee Study of Class Size in the Early School Grades

Frederick Mosteller

Abstract

The Tennessee class size project is a three-phase study designed to determine the effect of smaller class size in the earliest grades on short-term and long-term pupil performance. The first phase of this project, termed Project STAR (for Student-Teacher Achievement Ratio), was begun in 1985, when Lamar Alexander was governor of Tennessee. Governor Alexander, who later served as secretary of education in the cabinet of President George Bush, had made education a top priority for his second term. The legislature and the educational community of Tennessee were mindful of a promising study of the benefits of small class size carried out in nearby Indiana, but were also aware of the costs associated with additional classrooms and teachers. Wishing to obtain data on the effectiveness of reduced class size before committing additional funds, the Tennessee legislature authorized this four-year study in which results obtained in kindergarten, first, second, and third grade classrooms of 13 to 17 pupils were compared with those obtained in classrooms of 22 to 25 pupils and in classrooms of this larger size where the teacher was assisted by a paid aide. Both standardized and curriculum-based tests were used to assess and compare the performance of some 6,500 pupils in about 330 classrooms at approximately 80 schools in the areas of reading, mathematics, and basic study skills. After four years, it was clear that smaller classes did produce substantial improvement in early learning and cognitive studies and that the effect of small class size on the achievement of minority children was initially about double that observed for majority children, but in later years, it was about the same.

The second phase of the project, called the Lasting Benefits Study, was begun in 1989 to determine whether these perceived benefits persisted. Observations made as a part of this phase confirmed that the children who were originally enrolled in smaller classes continued to perform better than their grade-mates (whose school experience had begun in larger classes) when they were returned to regular-sized classes in later grades. Under the third phase, Project Challenge, the 17 economically poorest school districts were given small classes in kindergarten, first, second, and third grades. These districts improved their end-of-year standing in rank among the 139 districts from well below average to above average in reading and mathematics. This article briefly summarizes the Tennessee class size project, a controlled experiment which is one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools.

Frederick Mosteller, Ph.D., is a professor emeritus of mathematical statistics at the departments of Statistics and of Health Policy and Management at Harvard University.

Because we have all gone to school, we each have ideas about how to improve the system. For example, James Garfield once said that a pine log with a student on one end and Mark Hopkins, a beloved president of Williams College, on the other would be an ideal university. But if we want to improve school systems, we need to consider what changes may be practical and effective. Setting aside the discomfort of outdoor logs during New England winters, would Garfield's design have made effective use of President Hopkins's time? Aristotle, even when tutoring the young Alexander before he was called "the Great," is believed to have had more than one student per class.

The size of the class is largely under control of the school system, and its choice influences the size and number of classrooms and the number of teachers required, and so class size is naturally a concern of parents, teachers, and school administrators. Everyone is concerned that the pupils receive adequate attention and that the teachers are able to control their classes. Some courses seem to need more teachers per student than others. For example, classes in carpentry or cooking, in which hazardous tools and equipment are used, may require closer supervision than a class in arithmetic.

The effects of class size on children's learning have been studied, usually without reaching definitive conclusions. Most research on class size has compared the performance of pupils in classes of different sizes in such cognitive subjects as reading, mathematics, or social studies. Designing and executing these studies is difficult not only because parents may object to variation in the treatment of children but also because of the constraints that must be imposed if anything of value is to be learned from the investigation. Groups to be compared following different treatments need to be equivalent at the start. The treatments must be carefully described and delivered. Suitable measures of performance must be chosen. Beyond all this, a healthy atmosphere toward the investigation must be created; otherwise, the study can be easily sabotaged. It does not take many unwilling workers or full-time grumblers to spoil a research program.

In the 1980s, conditions favorable for a study of class size evolved in the state of Tennessee. Governor Lamar Alexander had established education as a top priority for his second term. Members of both the state legislature and the educational community in Tennessee had been intrigued by a modest-sized study in the state of Indiana, called Project Prime Time, which investigated the effect of reduced class sizes in kindergarten and first and second grades. For example, Bain and Achilles¹ report that, in Project Prime Time, (1) students in smaller classes scored higher on standardized tests than did those in larger classes, (2) the smaller classes had fewer behavioral problems, and (3) teachers of smaller classes reported themselves as more productive and efficient than they were when they taught larger classes.

The Tennessee legislators and teachers were also aware of an investigation by Glass and colleagues² which reviewed the vast literature on the effects of class size on learning using a special quantitative method called

meta-analysis. The results of this investigation suggested that a class size of 15 or fewer would be needed to make a noticeable improvement in classroom performance. At the time of the Glass study, the effect of class size on performance was controversial because many studies in the literature differed in their outcomes. The new methods used by Glass and his colleagues were not accepted by all professional groups. At the same time, there were ongoing discussions about the lesser cost and possibly equal effectiveness of placing paid teachers' aides in elementary classrooms. Because of the additional expense associated with a reduction in class size for early grades, members of the Tennessee legislature decided that any proposed innovation should be based on solid information and, therefore, authorized a four-year study of class size which would also examine the cost-effectiveness of teachers' aides. The legislature appropriated \$3 million in the first year for a study of pupils in kindergarten and then appropriated similar amounts in subsequent years for the project, which carried the acronym STAR (for Student-Teacher Achievement Ratio).³

The study was carried out in three kinds of groups: (1) classes one-third smaller than regular-sized classes, (2) regular-sized classes without a teacher's aide, and (3) regular-sized classes with a teacher's aide. By comparing average pupil performance in the different kinds of classes, researchers were able to assess the relative benefits of small class size and the presence of a teacher's aide. The experiment involved many schools and classes from inner-city, urban, suburban, and rural areas so that the progress of children from different backgrounds could be evaluated.

Study Design and Execution

Personnel from four Tennessee universities helped design and execute the Tennessee study, which was carried out in three phases (see Box 1). Each year, \$2.5 million was spent on additional teachers and teachers' aides. The remaining funds were used to gather and analyze the data and to carry out other obligations imposed by the enabling legislation.

Legislation for the STAR experiment required that studies be made of classes in inner-city, suburban, urban, and rural schools. Because the legislators did not define these types of residential areas, the study makers had to invent categories appropriate for Tennessee and their experiment. To do so, they placed inner-city and suburban schools in the category of *metropolitan* areas. Inner-city schools were defined as those in which more than half of the students received free or reduced-price lunches. Schools in outlying areas of metropolitan cities were called *suburban*. In nonmetropolitan areas,

schools in towns of more than 2,500 serving primarily an "urban" population were called *urban*, and the rest were classified as *rural*.

To be eligible to participate in the experiment, a school was required to sign up for four years and to have at least 57 children for any given grade (to comprise a small class of 13 and two classes of 22). This constraint assured the ability to make comparisons among the three kinds of classes within a single school. Participating schools received no extra support other than funds for additional teachers and aides and had to supply the extra classrooms. In any given calendar year, the experiment was carried out in one grade only, and this minimized the number of new classrooms needed. No new textbooks or curricula were to be introduced. Although 180 schools offered to participate, only 100 were large enough to qualify, and 79 actually participated in the kindergarten year.

The treatments planned for the program were started in 1985, beginning with

Box 1

The Tennessee Class Size Project

The Tennessee project on the effectiveness of small classes and of teachers' aides has had three phases:

Phase 1

1985–1989. The educational system of Tennessee carried out a four-year experiment, called Project STAR (for Student-Teacher Achievement Ratio), to assess the effectiveness of small classes compared with regular-sized classes and of teachers' aides in regular-sized classes on improving cognitive achievement in kindergarten and in the first, second, and third grades.

Phase 2

1989–. The Lasting Benefits Study (LBS) was an observational study of the consequences of the experimental program on children when they returned to regular-sized classes in the fourth, fifth, and sixth grades and beyond. This research phase asked whether the children who started in the smaller classes performed better in later grades. Only students who had been in the experiment (Phase 1) could contribute data to this second phase.

Phase 3

1989–. Project Challenge implemented the small classes in kindergarten and in the first, second, and third grades in the 17 districts of Tennessee where children are highly at risk of dropping out early. These districts have the lowest average incomes in the state.

kindergarten and continuing each year through first, second, and third grades. The classes were of three types: (1) small, 13 to 17 pupils; (2) regular size, 22 to 25 pupils; and (3) regular size with a teacher's aide. The small classes averaged

A teacher's aide had no specific duties but helped each teacher of a regular-sized class in whatever way the teacher wished. Some aides participated in teaching, others prepared materials and kept records, and some carried out all of these duties. Teachers' aides were paid.

The study findings apply to poor and well-to-do, farm and city, minority and majority children.

15 pupils, down about 35% from the average regular size of about 22 or 23. During the first year, the study involved about 6,400 pupils in 108 small classes, 101 regular-sized classes, and 99 regular-sized classes with teachers' aides.

Within a school, pupils and teachers were assigned to classes at random each year to ensure that classes came from equivalent populations and that teachers did not choose their classes. In a study of this kind, randomization protects against all variables that might matter, whether they have been identified or not.

Analysts report that attendance was about 95%, independent of school location, type of class, or minority or nonminority status.

Table 1 indicates the composition of the experimental groups by giving a breakdown of schools by city type and of classes by city type and ethnicity at the end of the first grade (second year of the experiment). This table shows participation by 6,572 pupils in 331 classes at 76 schools and is important because it indicates that enough pupils were studied to enable researchers to reach a conclusion. Ultimately, the findings from the investigations repeated themselves at least qualitatively in nearly every large cell of Table 1, suggesting that the study findings apply to poor and well-to-do, farm and city, minority and majority children. The magnitude,

Table 1

Composition of the First Grade Cross-Sectional Sample in the Second Year of the Tennessee Experiment				
	Location			
	Inner City	Urban	Suburban	Rural
Number of schools	15	8	15	38
Number of classes				
All majority students	0	18	28	119
All minority students	65	0	13	0
Mixed classes	5	23	21	39
TOTAL CLASSES	70	41	62	158
Number of students	1,495	804	1,214	3,059

Source: Finn, J.D., and Achilles, C.M. Answers and questions about class size: A statewide experiment. *American Educational Research Journal* (1990) 27,3:557-77.

control, and duration of the experiment illustrate the sort of investigations that are needed to improve education in the schools.

Examining and Interpreting the Findings

In assessing student performance, two types of tests were used: (1) standardized tests, which have the advantage of being used nationally but the disadvantage of not being directly related to any particular curriculum or course of study; and (2) curriculum-based tests, which reverse the advantages and disadvantages of standardized tests. Curriculum-based tests measure more directly the student's increased knowledge of what was actually taught, but they give little indication of where local results stand in the national picture.

The first graders took two standardized tests in reading: (1) the Stanford Achievement Test (SAT) for word study skills and reading, and (2) the Tennessee Basic Skills First (BSF) test for reading, a curriculum-based measure. In mathematics, first graders took one SAT (standardized) and one BSF (curriculum-based) test.

When an experiment applies a new treatment or employs a new method, one

way of comparing the effects of this new approach with those previously achieved using old treatments or methods is by expressing individual test scores in terms of standard deviation (see Box 2) and then expressing group differences as effect sizes (see Box 3). Here, *effect size* is defined as the difference between means divided by the standard deviation for individuals in the regular classes without aides. Thus Table 2 shows the effect sizes for small classes compared with the average of the

Both math and reading scores show a benefit of about one-fourth of a standard deviation.

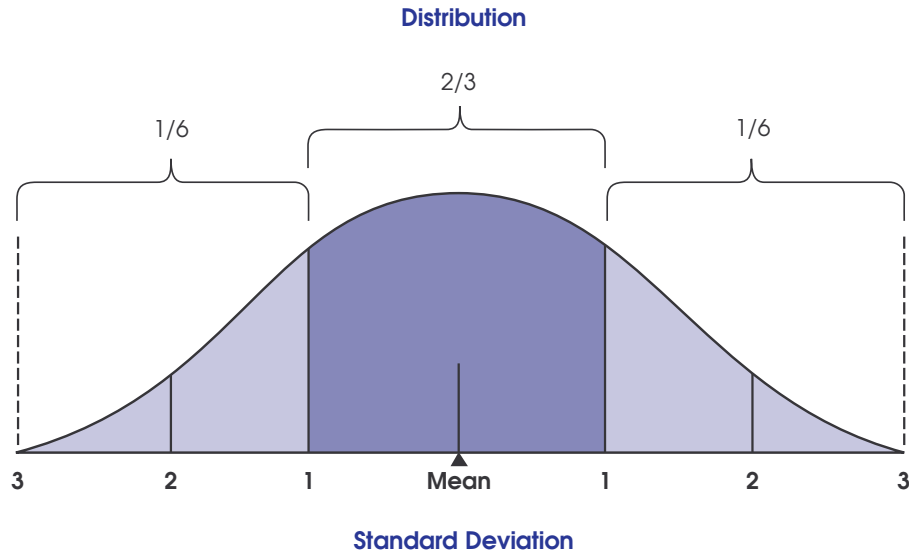
performance of the regular-sized classes with and without aides for the standardized tests. Both math and reading scores show a benefit of about one-fourth of a standard deviation. On the curriculum-based tests (BSF), reading scores improve by about one-fifth of a standard deviation and math scores by only one-twelfth.

To interpret the gains represented by these effects, it is useful to consider a pupil who, without a special treatment such as attending small classes, would achieve about the average score, say at the midpoint or 50th percentile, of all students.

Box 2

Standard Deviation

When considering distributions of quantities such as heights of people, family incomes, and scores on standardized tests, it is often useful to think first of the typical person, family, or score and then to represent that typical one by either the *mean* (average) of the numbers or the *median* (value of the middle measurement).



This drawing is of a distribution about the mean. The total area between the curve and the horizontal axis is one (or 100% of the measurements, incomes, or scores). For distributions that are approximately symmetrical, about half of the measurements lie to the right of the mean and half to the left. The slightly asymmetrical mountain-shaped (or bell-shaped) curve indicates roughly the way that many types of measurements distribute themselves in large populations, with the height of the curve representing the density of the scores at various positions. Typically, the distributions are dense in the middle and are less dense as one moves farther from the middle in either direction.

For many common distributions of everyday quantities, it is convenient to relate the mean and a measure of variability called the *standard deviation* to the fraction of measurements falling within a symmetrical interval about the mean. For example, this drawing indicates that the proportion of measurements falling in the interval that goes from one standard deviation to the left of the mean to one standard deviation to its right is about two-thirds. This number is not exact but is a rough approximation for distributions that are shaped generally like the one pictured.

What about the interval that includes the mean plus or minus two standard deviations? In the same approximate sense, this interval contains about 95% of the measurements for many distributions occurring in practice. If the interval is extended to three standard deviations each way from the mean, it will include nearly all—almost 100%—of the measurements.

What would a gain of one-fourth of a standard deviation do for such a pupil? That pupil would move from the 50th percentile of all pupils up to the 60th percentile, thus surpassing an additional 10% of the population beyond the 50% that were exceeded originally. Thus, an increase of one-fourth of a standard deviation can amount to considerable gain in performance.

In the study report, the average performance of small classes was compared with the average for all regular-sized classes with or without an aide. The resulting gain is shown in the first line of Table 2. The second line of that table shows the effect size of the gain from having an aide in the regular-sized class compared with the performance in the regular-sized class without an aide. When the effect of the small class is compared with that of the regular-sized class without an aide, the numbers in the first row of Table 2 increase to 0.30, 0.25, 0.32, and 0.15, respectively.

When performance of classes with an aide is compared with that of regular-sized classes without an aide, the gain averages about one-twelfth of a standard deviation. In other words, the average gain associated with an aide is about 35% of the gain achieved by reducing class size from regular to small.

Of special interest is the effect of class size on minority students. At the end of the second year of the experiment, in small classes compared with regular-sized classes and regular-sized classes with an aide, the effect size for minorities was about double that for majorities, averaged over the four tests. This extra gain occurred only in the first two years of the experiment; thereafter, the gains of both groups were about the same.

The original plan of the study was that all students would remain in their class types for all four years of the experiment. But after the first year, parents of students in regular classes objected to the continuation of the assignments. As a result of discussions with parents and with the people guiding the experiment, in the second year, students in the regular-sized classes with and without the teacher's aide were randomly reassigned half to classes with a teacher's aide and half to ones without,

but the assignments to small classes remained unchanged. Such changes were not allowed in later years. It was the view of the advisory group from the four universities that continued changes would make it impossible to interpret the results of the experiment. As a result of the changes that had been allowed, at the end of the second year, there were four situations in the regular classes for those who had attended kindergarten and first grade: (1) two years without an aide, (2) two years with an aide, (3) first year without an aide and second year with an aide, and (4) first year with an aide and second year without.

Schools had an influx of children in first grade who had not attended kindergarten the first year of the experiment. (Subsequently, kindergarten became required in Tennessee.) These children had to be assigned to the experiment in participating schools. This led to some separate analyses of results from kindergarten and first grade (for years one and two of the experiment) and of results from first, sec-

Of special interest is the effect of class size on minority students. The effect size for minorities was about double that for majorities.

ond, and third grades (for years two, three, and four of the experiment) to increase the numbers of students who experienced the same circumstances. (The rerandomization before the second year of the experiment shuffled some people between regular-sized classes with an aide and those without. Consequently, starting in the second year of the experiment, pupils could be classified according to their having experienced regular-sized classes with and without an aide.)

One way of summarizing results gives the percentile ranks for the average score based on national norms for the test. Table 3 shows the results for small classes, regular-sized classes, and regular-sized classes with a teacher's aide for both Total Reading SAT and Total Math SAT. Averaged over the four grades, the small classes gained a little more than eight percentiles over the regular-sized classes

Box 3

Effect Size

When an experiment applies a new treatment whose consequences are to be compared with those of the old or standard treatment, the difference in their consequences is often called the *size of effect* of the new treatment. For standardized tests, information is usually available which gives the distribution of scores for members of large populations who take the tests. Frequently, these distributions look like the common distributions described in Box 2. They are shaped approximately like distributions called Gaussian, or normal, in English-speaking countries. (When used in this way, the term *normal* means “usual, customary, or related to the norm” and does not connote an ideal situation or a desirable state of being.) The shapes of these curves are often well described by a formula that requires knowing only their mean and standard deviation.

Suppose that the national mean of a certain test is 500 and that its standard deviation is 100. Suppose as well that a new method of teaching produces higher test scores in an experimental group than would have been achieved without it, say a distribution with a mean of 550 instead of the usual 500. One way of thinking about this situation is to view the effect as shifting the original distribution to the right by 50 points—essentially adding 50 points to everyone’s score.

To interpret the value of this gain requires knowing how variable the scores are. If, for example, the standard deviation is 1,000 instead of 100, then 50 points does not look like much of a gain; but if the standard deviation is 10, a gain of 50 points is astounding because it represents a gain of five standard deviations, when a gain of only three standard deviations would take a student from an average score to one of the best scores that had ever been made.

One interpretable quantity is the gain represented as a fraction of the standard deviation of the original distribution. In this example, the fractional gain would be $50/100 = 0.5$, or half a standard deviation. An improvement of half a standard deviation would move people who were originally at the mean, which is also about the 50% point on these distributions, up to about the 69% point. Thus, a person who originally scored higher than half the population would now score higher than 69%.

This particular ratio of gain to the standard deviation is often called the *effect size*, a technical term that has a more specific meaning for such tests than the general notion of *size of effect*, which refers to any method of describing changes. In practice, effect sizes of half a standard deviation are rare.

Although effect sizes of the magnitude of 0.1, 0.2, or 0.3 may not seem to be impressive gains for a single individual, for a population they can be quite substantial. For example, a 0.2 effect size corresponds in the United States to the difference between the average heights of 15-year-old versus 16-year-old girls. For large numbers of girls of each age, this average difference may seem small, but most people notice it.^a An effect size of 0.3 corresponds to about 30 points on a SAT verbal or mathematics standardized test.

How much does computer-based instruction help students learn when it is offered as an adjunct to traditional teaching in certain settings? A review of 59 studies finds a mean effect size of 0.25 for computer-based instruction.^b And, as a result of this finding, computer-based instruction is viewed as an extraordinarily promising innovation—one that might revolutionize education.

Sources:

^a Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Erlbaum, 1988.

^b Kulik, J.A., Kulik, C.C., and Cohen, P.A. Effectiveness of computer-based college teaching: A meta-analysis of findings. *Review of Educational Research* (1980) 50:525–44.

Table 2

Gains in Effect Sizes from Small Classes				
Gains in effect sizes from small classes in first grade compared with all regular-sized classes and from regular-sized classes with an aide compared with regular-sized classes without an aide				
	SAT Reading	BSF Reading	SAT Math	BSF Math
The effect size on performance in small classes compared with performance in regular-sized classes with or without an aide	.23	.21	.27	.13
The effect size on performance in regular-sized classes with an aide compared with regular-sized classes without an aide	.14	.08	.10	.05

Source: Finn, J.D., and Achilles, C.M. Answers and questions about class size: A statewide experiment. *American Educational Research Journal* (1990) 27,3:557-77, Table 5.

without aides in reading and a little less than eight percentiles in mathematics. The addition of an aide to a regular-sized class results in a slight gain in both reading and math over the regular-sized class without an aide.

In the third year of the four-year study, questions were raised about the persistence of effects when children returned to regular-sized classes, as they would in fourth grade, and so an additional sum was appropriated for a three-year follow-up observation called the Lasting Benefits Study (LBS). As a part of this study, researchers observed the performance of children who had been in the three types of experimental classes during kindergarten and the first, second, and third grades after they returned to regular-sized classes in the fourth, fifth, sixth, and later grades.

In a paper presented at a meeting of the North Carolina Association for Research in Education at Greensboro, North Carolina, Achilles and colleagues reported on the Lasting Benefits Study.⁴ These authors found that, in the fourth and fifth grades, the children who had originally been in small classes scored higher than those who had been in regu-

lar-sized classes or in regular-sized classes with a teacher's aide. In the fourth grade—the first year after return to regular-sized classes—the effect size was about one-eighth of a standard deviation, averaged across six different cognitive subjects studied, and in the fifth grade, it was nearly two-tenths of a standard deviation, again averaged across six subjects. Within each grade, the different subjects produced almost the same effect size, though the

In the fourth and fifth grades, the children who had originally been in small classes scored higher than those who had been in regular-sized classes.

observed gain was somewhat larger for the fifth grade. Curiously, in both of these years, the effect size systematically favored the regular-sized classes previously without a teacher's aide over those previously with an aide, though the difference was small, averaging about 0.03 over all subjects in both grades. The encouraging finding is that early experience with the smaller class size seems to have had a continued effect beyond the moment when the children returned to regular-sized classes.

Table 3

Summary of Project STAR Results in Terms of the Percentile Ranks of Average Scores Based on National Test Norms				
	Percentile^a			
Grade level	K	1	2	3
Total reading SAT				
Small	59	64	61	62
Regular without an aide	53	53	52	55
Regular with an aide	54	58	54	54
Total math SAT				
Small	66	59	76	76
Regular without an aide	61	48	68	69
Regular with an aide	61	51	69	68

^a Percentile ranks are based on Stanford's multilevel norms.

Source: Word, E., Johnston, J., Bain, H.P., et al. *Student/Teacher Achievement Ratio (STAR): Tennessee's K-3 class size study*. Nashville: Tennessee Department of Education, Figures 1 and 2.

As a consequence of the systematic findings of improvement in performance of pupils in small classes over those in regular-sized classes, Tennessee implemented reduced class sizes for beginning students in kindergarten and first, second, and third grades in a program called Project

ematics. Before the small classes were introduced, these districts had been performing well below the average for the state in mathematics; after the intervention, they moved above the average.

It should be noted that the gains recorded here are not part of a carefully controlled experiment; they are consequences of installing the program. For this reason, the comparisons are not as well equated as they were in the original investigation. To measure experiment gains would require carrying out new class size experiments in the districts where the program is being implemented. Belief in the continuing benefits of the program is based on the uniform improvement found in the experiment for all types of classes in all types of cities. The additional evidence based on norms during the implementation phase, while reassuring, must be regarded as weaker because this new investigation is less well controlled.

An additional way to report the progress gives the average rank of the test scores of the 17 Tennessee districts in Project Challenge (among the 139 districts) for the years reported so far (1989–1993) in reading and mathematics. The results reported by Achilles, Nye, and Zaharias⁶ for the second grade are shown in Table 4.

Belief in the continuing benefits of the program is based on the uniform improvement found in the experiment for all types of classes in all types of cities.

Challenge (refer to the description of Phase 3 in Box 1) in the 17 school districts with the lowest per capita income and the highest percent of free or reduced-price lunch participation among students.

In the summary report for Project Challenge, Nye and colleagues observe that, in the school districts where small classes were installed in kindergarten, first, second, and third grades, both the reading scores and the math scores improved, compared with previous performance by children in these districts and with other schools in the state.⁵ The gains in effect sizes were 0.4 for reading and 0.6 for math-

Table 4

Average Second Grade Ranks for the 17 Districts Among the 139 School Districts for Early Years of Project Challenge				
Subject	Year			
	1989–90	1990–91	1991–92	1992–93
Reading	99	94	87	78
Mathematics	85	79	60	56

Source: Achilles, C.M., Nye, B.A., and Zaharias, J.B. Policy use of research results: Tennessee's Project Challenge. Paper presented at the Annual Convention of the American Educational Research Association, San Francisco, April 1995. Available from the Center of Excellence for Research in Basic Skills, College of Education, Tennessee State University.

When these districts are ranked from 1 to 139, where 1 indicates best academic performance and 139 indicates the worst, the average rank for all districts is 70. Note that in mathematics, the average rank for 1991–92 and for 1992–93 is below 60 (and so above the median) so that the 17 districts have shown a startling improvement as well as a gain of 20 ranks in reading for second grade. The same report mentions that the corresponding analysis of first grade shows that the 17 districts were better than average in both reading and mathematics in 1992.⁷

In summary, the evidence is strong that smaller class size at the beginning of the school experience does improve the performance of children on cognitive tests. Observations from the Lasting Benefits Study confirm that the effect continues into later grades when children are returned to regular-sized classes. In addition, the implementation of the program for the economically poorest districts seems to be improving the performance of children in these districts by noticeable amounts. In regular-sized classes, an aide produced some gain in kindergarten and in the first, second, and third grades; but when students returned to regular-sized classes, the gain from aides did not persist. After the small classes were implemented in all 17 school districts, no further observations were made about the in-classroom value of paid teachers' aides.

Special Concerns

During the course of the experiment, researchers made two substantial departures from the basic plan: they rerandomized regular-sized classes during the second year and moved incompatible children. In addition, researchers instituted a teacher training program between the second and third year.

As reported earlier, one departure from the original plan occurred in the second year, when the children in regular-sized classes were rerandomized to regular-sized classes with an aide and regular-sized classes without an aide. Such a change applied to all who had entered the experiment in kindergarten. From the point of view of

Second-Year Rerandomization in Regular-Sized Classes

those beginning in kindergarten, it created four rather than two regular-sized groups of classes for analysis and comparison, as described above. After the second year, the children in regular-sized classes continued with their second-year assignment. This change complicates the analysis for all children except those whose assignments remained unchanged and makes it difficult to assess accurately the effectiveness of having or not having a teacher's aide.

The evidence is strong that smaller class size at the beginning of the school experience does improve the performance of children on cognitive tests.

those beginning in kindergarten, it created four rather than two regular-sized groups of classes for analysis and comparison, as described above. After the second year, the children in regular-sized classes continued with their second-year assignment. This change complicates the analysis for all children except those whose assignments remained unchanged and makes it difficult to assess accurately the effectiveness of having or not having a teacher's aide.

© Courtesy of Tennessee State University/Project STAR



Moving Incompatible Children

One benefit reported from the Indiana study was that behavioral problems were reduced in the smaller classes. Nevertheless, in Project STAR at the end of the first year, 48 students moved from small kindergarten classes to regular classes with an aide, and 60 moved to regular classes without an aide.⁸ Thus, the number of students moved from small classes was 108 of 1,678 students.⁹ This move was intended to separate incompatible chil-

It is impossible to assess the impact of this reassignment on the experiment; and, in fact, it may have had little impact because the affected students may have been removed from the analysis altogether.

The Teacher Training Program

The added feature in Project STAR came between the second and third years, when it was decided to give a special training course to 57 teachers. The enabling legislation had specified teacher training. Essentially, all teachers were getting some additional training as a routine matter in Tennessee, but apparently it was felt that the legislation called for something special. The participating teachers in 15 selected Project STAR schools were all given a total of three days of special training. The training was the same for all teachers selected; their assignment to small or regular-sized classes had not yet been made.⁸ When one considers that 30% of these teachers already had 20 years of teaching experience and only four had fewer than 3 years of experience, a three-day training program seems modest. As it turned out after the training, the classes with trained teachers performed the same as did the classes with untrained teachers.

Having fewer children in class reduces the distractions in the room and gives the teacher more time to devote to each child.

dren and “to achieve sexual and racial balance,”¹⁰ the latter a puzzling remark in view of the purported emphasis on randomization. No mention is made of what was done about incompatible students who were already in regular-sized classes. Perhaps there was nowhere to move them if there was only one small class or perhaps children seem more incompatible in small classes. A school administration planning to reduce class sizes might want to keep this potential difficulty in mind.

Class Size Drift

In addition, the sizes of the classes drifted a bit as time went on. Some small classes

became larger than their intended upper bound, and some regular-sized classes became smaller than their intended lower bound. The overall outcome of these violations of the original distributions should be to underestimate the effectiveness of the small classes compared with that of the regular-sized classes.

Assessing the Implications of the Study

Smaller Class Size

Why does smaller class size help teaching and learning? Reducing a class from 23 to 15 reduces the number of children in the room by about one-third. Having fewer children in class reduces the distractions in the room and gives the teacher more time to devote to each child. However, the impression one gets from reading papers emerging from Project STAR is that at least some teachers and administrators engaged in the study think of themselves as dealing with a start-up phenomenon. When children first come to school, they are confronted with many changes and much confusion. They come into this new setting from a variety of homes and circumstances. Many need training in paying attention, carrying out tasks, and interacting with others in a working situation. In other words, when children start school, they need to learn to cooperate with others, to learn to learn, and generally to get oriented to being students. These observations fit neatly with several current theories of education, including the idea of frames and scripts.^{11–16}

The experiment showed that the minority groups gained more than others in the first two years of the experiment; and although the last two years showed benefits comparable with those of the majority, there was a falling off of benefit. Some statements in the report by Word and colleagues³ suggest that much of the gain from the small classes was achieved in the first two years. The data presented in Table 3 do not show the falling off, but other summary tables from the study might.

Optimum Class Size

The idea of an ideal, or optimum, class size is open to question. This investigation

did not provide information about a variety of class sizes. Within the ranges of what is affordable, it is reasonable to suppose that smaller classes are preferable for beginners. But some desired training probably could not be accomplished in classes of such small sizes as one or two pupils even if they were affordable. Learning to work in a group is important and requires the presence of others.

Persistence of Beneficial Effects

In the Lasting Benefits Study,⁴ a continuation of studies evaluated the performance of students from small classes as compared with the performance of students from regular-sized classes or regular-sized classes with an aide after all students had returned to regular-sized classes. The results always favored the students from smaller classes. One year later (1989–90), the effect sizes ranged from 0.11 to 0.16 ($n = 4, 230$) in the fourth grade, and then, in subsequent years, from 0.17 to 0.34 ($n = 4, 639$) in the fifth grade, from 0.14 to 0.26 ($n = 4, 333$) in the sixth grade, and

The students who were originally in smaller classes continued to perform better than the students from regular-sized classes with or without a teacher's aide.

from 0.08 to 0.16 ($n = 4, 944$) in the seventh grade. Data from the eighth grade have been gathered and are being analyzed. Thus, year after year, the students who were originally in smaller classes continued to perform better than the students from regular-sized classes with or without a teacher's aide.¹⁷

Conclusion

Compelling evidence that smaller classes help, at least in early grades, and that the benefits derived from these smaller classes persist leaves open the possibility that additional or different educational devices could lead to still further gains. For example, applying to small classes the technique of within-class grouping in which the teacher handles each small group separately for short periods could strengthen

the educational process (essentially a second-order use of small class size). The point is that small classes can be used jointly with other teaching techniques which may add further gains.

Because a controlled education experiment (as distinct from a sample survey) of this quality, magnitude, and duration is a rarity, it is important that both educators and policymakers have access to its statistical information and understand its implications. Thought should be given by both public and private organizations to making sure that this information is preserved and well documented and that access to it is encouraged. The Tennessee three-phase study calls attention to the statewide con-

trolled experiment as a valuable device for assessing educational interventions and, thereby, improving school systems.

The preparation of this material was supported in part by a grant from the Andrew W. Mellon Foundation to the American Academy of Arts and Sciences in support of the Center for Evaluation of its Initiatives for Children program. The author's efforts have been helped by the kind responses of people who worked on the Tennessee class size project and have kept him in touch with project publications as they have appeared. Professors C.M. Achilles and J.D. Finn have given helpful advice and information. In addition, suggestions from John Emerson, Richard Light, Marjorie Olson, Jori Raymond, Jason Sachs, and Cleo Youtz improved early versions of the manuscript.

1. Bain, H.P., and Achilles, C.M. Interesting developments in class size. *Phi Delta Kappan* (1986) 67:662–65.
2. Glass, G.V., Cahen, L.S., Smith, M.L., and Filby, N.N. *School class size: Research and policy*. Beverly Hills, CA: Sage, 1982.
3. Word, E.R., Johnston, J., Bain, H.P., et al. *The state of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical report 1985–1990*. Nashville: Tennessee State Department of Education, 1994.
4. Achilles, C.M., Nye, B.A., Zaharias, J.B., and Fulton, B.D. The Lasting Benefits Study (LBS) in grades 4 and 5 (1990–1991): A legacy from Tennessee's four-year (K–3) class-size study (1985–1989), Project STAR. Paper presented at the North Carolina Association for Research in Education. Greensboro, North Carolina, January 14, 1993.
5. Nye, B.A., Achilles, C.M., Zaharias, J.B., and Fulton, B.D. Project Challenge third-year summary report: An initial evaluation of the Tennessee Department of Education "At-Risk" Student/Teacher Ratio Reduction Project in seventeen counties 1989–90 through 1991–92. Nashville: Center of Excellence for Research in Basic Skills, College of Education, Tennessee State University, 1993.
6. Achilles, C.M., Nye, B.A., and Zaharias, J.B. Policy use of research results: Tennessee's Project Challenge. Paper presented at the Annual Convention of the American Educational Research Association. San Francisco, April 1995. Available from the Center of Excellence for Research in Basic Skills, College of Education, Tennessee State University, Nashville.
7. See note no. 6, Achilles, Nye, and Zaharias, Appendix B, p. B-1.
8. Word, E., Johnston, J., Bain, H.P., et al. *Student/Teacher Achievement Ratio (STAR): Tennessee's K–3 class size study*. Final summary report 1985–1990. Nashville: Tennessee Department of Education, 1990.
9. See note no. 3, Word, Johnston, Bain, et al., p. 45, Table IV-2.
10. See note no. 3, Word, Johnston, Bain, et al., p. 59.
11. Wertsch, J.V., and Smolka, A.L.B. Continuing the dialogue: Vygotsky, Bakhtin, and Lotman. In *Charting the agenda: Educational activity after Vygotsky*. H. Daniels, ed. London: Routledge, 1993, pp. 69–92.
12. Wertsch, J.V. *Voices of the mind: A sociocultural approach to mediated action*. Cambridge: Harvard University Press, 1991.
13. Bruner, J.S. *Acts of meaning*. Cambridge: Harvard University Press, 1990.
14. Elkind, D. *Miseducation: Preschoolers at risk*. New York: Knopf, 1987.

15. Nelson, K. Characteristics of children's scripts for familiar events. Paper presented at the meeting of the American Psychological Association. Montreal, September 1980.
16. Schank, R.C., and Abelson, R.P. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum, 1977. Distributed by the Halsted Press Division of John Wiley and Sons.
17. Nye, B.A., Zaharias, J.B., Fulton, B.D., et al. *The Lasting Benefits Study: A continuing analysis of the effect of small class size in kindergarten through third grade on student achievement test scores in subsequent grade levels*. Seventh grade technical report. Nashville: Center of Excellence for Research in Basic Skills, Tennessee State University, 1994.