



---

The Tennessee Study of Class Size in the Early School Grades

Author(s): Frederick Mosteller

Source: *Bulletin of the American Academy of Arts and Sciences*, May, 1997, Vol. 50, No. 7 (May, 1997), pp. 14-25

Published by: American Academy of Arts & Sciences

Stable URL: <https://www.jstor.org/stable/3824562>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

American Academy of Arts & Sciences is collaborating with JSTOR to digitize, preserve and extend access to *Bulletin of the American Academy of Arts and Sciences*

## Project Report

### The Tennessee Study of Class Size in the Early School Grades

Frederick Mosteller

*The Evaluation Center of the Academy's Initiatives for Children Program analyzes interventions directed at improving the status of children. The Center's director, Frederick Mosteller (Harvard University), has examined the findings on elementary school class size in a large multiyear experiment in Tennessee. The following report is an abridgment of a paper evaluating the results of that study, published in the Summer/Fall 1995 issue of The Future of Children.*

*Under Mr. Mosteller's leadership, the Center has also completed studies on other topics, including the effects of skill grouping on learning, the efficacy of computer-assisted language learning, and the findings from research on student evaluations of college teachers. Currently, he and a working group are continuing studies of educational interventions and examining interventions in the realm of juvenile justice, starting with the Massachusetts CHINS (Children in Need of Services) program.*

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

---

Abridged, with permission, from Mosteller F., "The Tennessee Study of Class Size in the Early School Grades," *The Future of Children* 5(2): page 113, 1995. © 1995 by the Packard Center for the Future of Children.

“the Great,” is believed to have had more than one student per class.

The size of the class is largely under the 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, second, and third grades. In that

project, (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 being 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 colleagues, 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, known by the acronym STAR (for Student-Teacher Achievement Ratio).

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 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. 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, minimizing the number of new classrooms needed. No new textbooks or curricula were to be introduced.

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

A teacher's aide had no specific duties but helped each teacher of a regular-sized class however 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.

Analysts report that attendance was about 95 percent, independent of school location, type of class, or minority or nonminority sta-

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

### *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. Student performance in small classes was compared with that in regular-sized classes with and without aides.

In small classes, both math and reading scores on the standardized tests showed a benefit of about one-fourth of a standard deviation. On the curriculum-based tests, reading scores in small classes improved by about one-fifth of a standard deviation, and math scores improved 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. A gain of one-fourth of a standard deviation would move such a pupil 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 a considerable gain in performance.

A comparison between the performance of regular-sized classes with and without an aide showed that the average gain associated with an aide is about 35% of that 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 type 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 a 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 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,



second, 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 test results gives the percentile ranks for the average score, based on national norms for the test. Averaged over the four grades, the small classes gained a little more than eight percentiles over the regular-sized classes without aides in reading and a little less than eight percentiles in mathematics. The addition of an aide to a regular-sized class resulted 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 several-year follow-up observation called the Lasting Benefits Study. As 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.

Achilles and colleagues reported that 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 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 ob-



served 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. 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.

As a consequence of the systematic findings of improvement in the 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 Challenge in the seventeen 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 and first, second, and third grades, both reading and math scores improved, compared with previous performance by children in those districts and with other schools in the state. Before the small classes were introduced, those districts had been performing well below the average rank for the school districts of the state in mathematics; after the intervention, they moved to above the average rank.

It should be noted that these gains 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.

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 those 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 seventeen school districts, no further observations were made about the in-classroom value of paid teacher's aides.

### *Special Concerns*

As noted earlier, a substantial 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. That change, applied to all who had entered the experiment in kindergarten, created four rather than two regular-sized groups of classes for analysis and comparison. 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.

One benefit reported in 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. Thus, 108 of 1,678 students moved from small classes. This move was intended to separate incompatible

children and “to achieve sexual and racial balance,” the latter a puzzling goal in view of the purported emphasis on randomization. The investigators make no mention 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.

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

Between the second and third years of Project STAR, the researchers 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. When one considers that 30 percent of these teachers already had twenty years of teaching experience and only four had fewer than three 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 those with untrained teachers.

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 investigators report that 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*

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 a 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.

The experiment showed that the minority groups gained more than others in the first two years of the experiment. In the last two years, although the minority children showed benefits comparable with those of the majority, there was a falling off of benefit. A report by Word and colleagues suggests that much of the gain from the small classes was achieved in the first two years.

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 very small classes of one or two pupils, even if they were affordable. Learning to work in a group is important and requires the presence of others.

The Lasting Benefits Study, the second phase of the Tennessee class size project, has evaluated the performance of students from small classes as compared with the perfor-

mance of students from regular-sized classes with or without an aide after all students had returned to regular-sized classes. The results have always favored the students from smaller classes. Year after year, from the fourth through seventh and later grades, the students who were originally in smaller classes have continued to perform better than the students from regular-sized classes with or without a teacher's aide.

### *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 that 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 policy makers 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 controlled experiment as a valuable device for assessing educational interventions and thereby improving school systems.

*The unabridged version of this report may be obtained at <http://www.futureofchildren.org> on the Internet. Fellows may also request complimentary copies of the journal issue containing the article by writing to Lisa Aguilar, David and Lucile Packard Foundation, Center for the Future of Children, 300 Second Street, Suite 102, Los Altos, CA 94022 (separate reprints are not available).*