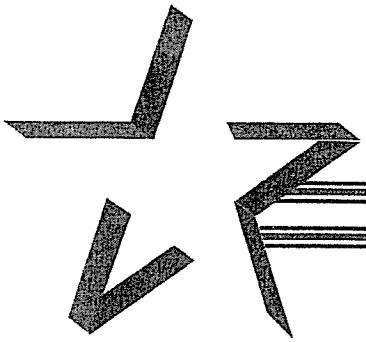


The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project

Final Summary Report
1985 – 1990

Prepared by:

Tennessee State Department of Education
Elizabeth Word, Project STAR Director

Memphis State University
John Johnston

Tennessee State University
Helen Pate Bain
B. DeWayne Fulton
Jayne Boyd Zaharias

University of Tennessee, Knoxville
Charles M. Achilles
Martha Nannette Lintz

Vanderbilt University
John Folger
Carolyn Breda

Summary Report

A. Background

In the spring of 1984, the Tennessee State Legislature adopted comprehensive education reform called the Better Schools Program. Although the media gave most attention to the career ladder for teachers, the Tennessee Center for Excellence program provided higher education with a means to work toward improving education in Tennessee. The Tennessee State University (TSU) Center for the Teaching of Basic Skills to the Economically and Educationally Disadvantaged began a modest program on the effects of small class size in one Metro Nashville school. The director of that project, Dr. Helen Bain, encouraged the legislature to adopt a reduced class size program statewide. One model for what might be done in Tennessee was a program in Indiana, Project Prime Time, which studied reduced class size in grades K-2.

In the spring of 1985, information about the Indiana and the TSU studies was shared with the Tennessee State Board of Education Chairman and staff and the members of the House and Senate Education Committees. Steve Cobb, a member of the Tennessee House of Representatives, became interested in the issue of the optimum class size in the early elementary grades. The literature, particularly the Glass Meta-Analysis (1982), suggests that class size must be reduced to about 15 to 1 to have a noticeable effect on student achievement. Glass' analysis has been criticized because the type of school and student characteristics in small classes are unrepresentative of the average public school student, and some of the "small classes" were tutoring groups (Educational Research Service, 1980). Because the research results were not conclusive and because the cost of a major reduction in class size would be very large, Representative Cobb wanted the state to conduct a well-designed study of class size before investing in a costly new program. With legislation, House Bill (HB) 544, passed in May 1985, the Tennessee Legislature authorized and funded a major policy study to consider the effects of class size on students in primary (K-3) grades. There was an appropriation of $3,000,000 for the first year of the four-year study.

B. Organization to Conduct the Study

The Tennessee State Department of Education organized to conduct the legislated study of reduced student/teacher ratio and adopted the name STAR, an acronym for Student/ Teacher Achievement Ratio. The Department employed Elizabeth Word, an experienced elementary principal, as project director and asked personnel from four universities (Memphis State University, Tennessee State University, the University of Tennessee at Knoxville, and Vanderbilt University) to develop the study design, plan the research, analyze the data, and prepare periodic reports of progress for the State Board of Education and the legislature. The State Department of Education retained management and budget control of the project, and the universities had both an advisory and an operational role. Responsibilities for direct contact with the 79 STAR schools were divided among the universities. Personnel from each university supplied assigned schools with information, collected data, and observed testing and other activities. The project director contacted all schools directly concerning administrative and fiscal matters and some research issues.

Thus, the organization to conduct the study consisted of a consortium of persons from the Tennessee State Department of Education, STAR staff, the four universities, and a

representative each from the State Board of Education and the State Superintendents' Association. The term "consortium" refers to the total group that guided project activities.

The State paid salary costs for the extra teachers required to reduce class size and the project teacher aides and provided modest contracts to each of the four universities in the consortium. Major costs, about 2.5 of the 3 million dollars per year, were for additional teachers and aides in the project. During the third year of the study it was decided to have a follow-up year to collect information about the persistence of STAR achievement gains and to complete the data analysis. Funding for an additional year was requested and the legislature approved $389,500 for this purpose.

Two nationally recognized experts on class-size research and measurement served as an external review and advisory committee. They were Dr. Doris Ryan of the Ontario Institute for Studies in Education and later at St. Johns University, New Brunswick, who has extensive experience in the conduct of class-size studies, and Dr. Roy Forbes of East Carolina University (and later at the University of North Carolina, Greensboro) and former director of the National Assessment of Educational Progress. Several researchers from Memphis State University, Tennessee State University, the University of Tennessee, Knoxville and Vanderbilt University reviewed the project's design and the work plans and suggested ways to improve the design. As a result of their suggestions, the comparison schools were added to the design. The consultants reacted favorably to the within-school design and the study plans. Dr. Jeremy Finn, a nationally acclaimed educational statistician from the State University of New York at Buffalo, assumed responsibility for primary analyses of class-size effects for each year of the project.

C. Sample Selection

The project timeline (legislation in May, director appointed in July, schools opened in August) required the consortium to decide upon a design and get students placed quickly. The first task, even while the design was being developed, was to identify school districts and schools to participate in the study. The ideal would have been that all school districts would opt to participate and that all choices (select districts to participate from among all districts in the state, then select schools, teachers, students, etc.) be made randomly.

1. Selection of Project Schools

The legislation specified that the project should include "inner city, suburban, urban, and rural schools" to assess the effects of class size in different school locations. No existing designation of schools used the categories specified above, so the consortium developed designations using various criteria.

Inner-city and **suburban** schools were all located in metropolitan areas. Schools that had more than half of their students on free or reduced cost lunch (indicative of a low-income family background) were tentatively defined as **inner city.** Schools in the outlying areas of metropolitan cities were classified as **suburban**.

In non-metropolitan areas, schools were classified as **urban** or **rural** depending on the location of the school. If located in a town of over 2,500 and serving primarily an urban population (the census definition of urban), the school was classified as **urban**. All other schools were classified as **rural**. All classifications were checked with local school officials to see if they agreed with the designation of their school. The application of these rules led to the classification of schools shown in Table 1.

In kindergarten there were 17 inner-city schools and 16 suburban schools drawn from four metropolitan areas: Knoxville, Nashville, Memphis, and Chattanooga. Fifteen of the 17

inner-city schools were located in Memphis. There were 8 urban schools that served non-metropolitan cities and large towns (for example, Manchester and Maryville). There were 38 rural schools.

Schools were spread across the state, not clustered in one section. The Commissioner of Education invited all Tennessee school systems to participate and sent guidelines for participation to each local system. These guidelines indicated that the state would cover additional costs for project teachers and teacher aides, but that local systems would furnish any additional classroom space needed. The project schools would not receive any special considerations other than class size--the students would use the regular district or school curriculum, supplies, texts, etc. There should be no major changes in process, organization, etc, other than class sizes. Schools should plan to remain in the project for four years; the project would start in kindergarten in 1985-86 and follow students successively through grades one, two and three.

TABLE 1

Project STAR Schools by School Type
Kindergarten Through Grade 3 (1985-1989)

	Kindergarten	Grade 1	Grade 2	Grade 3
Inner City	17	15	15	15
Suburban	16	15	15	15
Rural	38	38	38	38
Urban	8	8	7	7
Total	**79**	**76**	**75**	**75**

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 the different class conditions. Initially, 180 schools in about 50 of the state's 141 school systems expressed interest in participating. Only about 100 schools had enough students in kindergarten (a minimum of 57) to meet the size criterion for participation. The size criterion was necessary because the project utilized a "within-school" design. The final selection of schools was based on (a) including at least one school from each district that had volunteered and (b) including enough schools from all four school types and all three regions of the state to permit comparisons between school types, as specified in the legislation. After discussion and negotiation, 79 schools in 42 systems became participants in the first year. The goal was to have approximately 100 small, 100 regular, and 100 regular with aide classes. This objective was met. In the 1985-86 year, the project had 128 small classes (approximately 1,900 students), 101 regular classes, (approximately 2,300 students), and 99 regular classes with teacher aides (approximately 2,200 students).

2. Selection of Comparison Schools

In addition to the project schools, information was needed about the performance of a comparable group of students whose teachers were carrying out the regular school program in average-size classes. Sometimes an experiment in a school will affect all students and all teachers. The use of a comparison group helped researchers to identify such effects. The superintendent of each system having project schools was asked if non-STAR elementary

schools would administer the same tests used in STAR schools to provide comparison scores. Seventeen systems identified 39 possible comparison schools. Twenty-two schools with 51 regular classes and approximately 1,100 students became a comparison group. The 22 comparison schools, drawn from 17 STAR school systems, administered the same tests that the project schools administered. Comparison schools allowed STAR researchers to check on the possibility that project schools were influenced by the Hawthorne Effect.

D. Data Collection Plan and Data Base

A major first-year task was to plan and implement a comprehensive data collection plan for the first and subsequent years. The design and data formats allowed researchers to follow individual students for four years. Subjects were assigned individual identification numbers. Data were collected for students, teachers, principals, teacher aides, schools, and systems. Each child in the appropriate grade in comparison schools received an identification number and information was collected about race, sex, age, free or reduced lunch (one indicator of socioeconomic status), and test scores.

In seeking information about why a small class might affect student learning, researchers collected data about how teachers teach, about student-teacher interactions, etc. Data were also collected on factors that might affect the results: the number and distribution of special education children, pull-out programs, and adults other than the teacher who participate in the instructional program.

E. General Operating Guidelines

Two general guidelines helped project personnel with operational decisions.

1. Participation in STAR would not cause any child to receive fewer services than if the child/school did not participate. (Participation in STAR would not put any child "at risk" in any way.)

2. STAR would not dictate changes (e.g., curriculum, materials, schedule) to the school; STAR efforts would work within the regular school framework (with the exceptions of student and teacher assignment, ability grouping across classes, testing, etc.) as much as possible. STAR would minimize disruptions to the school's regular routine. Schools would maintain the random assignments, and basic instruction would be carried out primarily in the classes to which students were assigned.

F. Teacher Orientation

Orientation sessions were conducted for teachers at 20 schools entering the project in kindergarten. The orientation idea was later refined and used for all principals and all teachers entering the project. The person conducting the orientation described the project, its purposes and processes, and answered questions. The orientation process for new teachers entering the project at each grade level was also expanded after the first year and made more comprehensive.

G. Study Design

The STAR design had to provide answers for questions required by the legislation, meet requirements for a longitudinal study, review one-year or cross-sectional effects of the treatment, and answer questions of interest. Two key design decisions were to have a within-school design and random assignment of both teachers and students to class types.

STAR was a randomized experiment conducted **in situ**. The control-group design was Campbell and Stanley (1963) Design Number 6, a randomized experiment employing post-test analysis only. The primary analysis was built on post-test only design. Additional analyses employed other analytic models.

H. Choice of Within-School Design

Because of potentially large differences between schools (i.e., school effects) in such items as resources, teachers and students, the consortium chose a within-school design. A within-school design reduced major sources of possible variation in student achievement attributable to school effects. This decision required that each school have sufficient enrollment in each grade (at least 57 students) to provide at least one small (13-17 enrollment), one regular (22-25), and one regular with a full-time aide (22-25) class. In schools with larger enrollments, additional classes were established. This design assured that there would be the same kinds of students, curriculum, principal, policy, schedule, expenditures, etc., for each class type by school and avoided the problem of control groups that were not motivated to attend carefully to project needs since they probably would gain nothing by remaining in the project. In the within-school design the control classes participated fully in all testing, etc., since it was part of the project. An entire project school might do better than expected due to project participation (the halo or Hawthorne effect). Reciprocally, it was also possible that competition could occur within the school whereby the control teacher(s) would work exceptionally hard (the John Henry effect).

After initial selection of participating systems, the choice of schools within systems was partly a function of school size. Grade-level enrollment determined the number of classes of each type established in each school. For example, the 79 schools selected to participate in Project STAR (kindergarten) provided enough classes (small, regular, regular w/aide) to meet the design estimate of approximately 100 classes of each type.

I. Selection into the Three Conditions

The 79 project elementary schools selected in the first year served rural, urban, suburban and inner-city students. The within-school design required each participating school to have three or more classes. Larger schools had more classes distributed among the three class types. Table 2 shows the design configurations for establishing classes in schools of various sizes. A student in a small class in kindergarten remained in the small class for grades one, two and three, to assist the measurement of cumulative effect of the class type. In kindergarten (1985-1986), there were 128 small classes, 101 regular classes, and 99 regular classes with full-time teacher aides. Approximately 6,500 students participated in Project STAR in kindergarten.

TABLE 2

Plan for Distribution of Students and Classes in Within-School Design: Project STAR (1985-1986)

Design Type	Enrollment (ADM)	Classes (N)	Class Types	Extra Room Needed
One	57-67	(3)	S,R,R/A	No
Two	68-78	(4)	S,S,R,R/A	Yes
Three	79-92	(4)	S,R,R/A,R/A or S,R,R,R/A	No
Four	93-109	(5)	S,S,R,R,R/A or S,S,R,R/A,R/A	Yes
Five	110-134	(6)	S,S,R,R,R/A,R/A	Yes
Six	135+	(7+)	Individually Designed	Yes

S=Small Class (1:13-17);R=Regular Class (1:22-25);
R/A=Regular Class with a Full-time Teacher Aide (1:22-25)

The plan described in Table 2 was used to govern the selection of class condition throughout the study. Once assigned to a class type a student was to remain in the assigned class type as long as he/she was in the project. Due primarily to teacher-identified discipline problems and some parent complaints, the STAR consortium had to revise this procedure after the kindergarten year. Since there were no differences on any measure for students in regular and regular with aide classes, students who had been in these class types in kindergarten were reassigned randomly within the two class types for first grade. The external advisory committee informed STAR that this interchanging could create problems in conducting longitudinal analysis. Therefore, first grade was the only grade in which students in regular and regular with aide classes were permitted to interchange. No further changes were made after first grade.

J. Modifications in Study Design

In a large-scale field project some changes occur that cannot be anticipated. Schools may drop out of the project; classes will gain or lose students; and in some cases these changes will make a class too small or too large for the design. The researchers took these possibilities into consideration by over designing the project. A power test at the beginning of the project indicated that it would be possible to detect a small achievement difference (3% or more) with only 80 classes of each type, or a total of 240 classes, rather than the 329 that actually participated. At the end of kindergarten, 34 classes had either too many or too few students for the original design (e.g., a small class may have ended up with 12 students rather than staying within the 13-17 range). Data were analyzed both including and excluding the 34 classes and results of both analyses were substantially the same. Oversampling was necessary because of the expected attrition of students and schools over the project's four years.

K. Methodology (Primary Analysis)

Project STAR's primary analysis consisted of a cross-sectional analysis of data from all students participating in project classes at each grade level, and two longitudinal analyses. For the latter, data were analyzed for students who were in the project in the same class type for four consecutive years (K-1-2-3). Analyses-of-variance procedures were employed to address the major questions of the study as follows:

(1) **Class Type** (Small/Regular/Aide) was assumed to be a fixed dimension; mean differences among class types comprise the most important question of the investigation.
(2) **School Type** (Inner City/ Urban/ Suburban/ Rural) was assumed to be a fixed dimension, crossed with class type.
(3) **Schools** were treated as a random dimension, nested within locations, but crossed with class type, since all three class types were present in each school. This is an important aspect of the design to account for the influence of shared conditions on all project classes within a school.
(4) **Classes** were treated as a random dimension when there were more than one class of a given type within a particular school.
(5) **Students** were treated as a random sample, nested within each class.

When all of the main effects and interactions of these factors are assembled into an analysis-of-variance model and expected mean squares evaluated, the resulting tests of significance are those given in Table 3.

TABLE 3
Analysis of Variance Source Table

Source of Variation	Error Term
Fixed effects:	
Location	Schools
Class Type	Location X Class Type
Schools X Class Types	
Random Effects:	
Schools	
Schools X Class Types	
Classes	
Students	

Since the error terms needed to test the significance of the fixed effects in the data are variations attributable to Schools and the School-X-Type interaction, student-level data were not required for this portion of the STAR analysis. Thus, data were aggregated to the level of class means before the analysis was conducted, to reduce the magnitude of statistical computations. Table 3 also shows that the correct error degrees of freedom for the primary questions of the study are proportional to the number of schools -- in the neighborhood of 75 for some tests and 150 for others -- and not the number of students. The exact degrees of freedom for each computer run were affected slightly by the pattern of missing data on the particular instrument.

A parallel analysis was conducted with sex (grades K and 1 only) and race (all grades) as an additional factor of classification. Since both males and females are present in each class and, potentially, both white and minority students, these factors were treated as fixed effects, crossed with all other dimensions in the design. For these tests, means of all males and all females in each class, or all white and all minority students in each class were used as the units of analysis. Race and Sex were analyzed in parallel computer runs, so that no analysis of both factors was conducted simultaneously.*

The design has unequal N's and many empty cells. A general linear model approach for nonorthogonal designs was employed.

In each year, data from the measurement instruments were analyzed in subsets: the SAT achievement scales, the BSF performance tests (beginning in grade 1), and the SCAMIN self-concept and motivation scales. Since the measures are intercorrelated, multivariate test statistics (Wilks' likelihood ratios) were employed for each subset.

Prior to all analyses, the distributions of the criterion measures were examined for skewness and outliers. This resulted in only a few deletions of data that were obviously erroneous, and a rescaling of the BSF reading and mathematics scale. Individual students were scored as pass or fail, based on whether or not they passed 80 percent of the objectives covered on the respective test. At the class level, the percentage of students passing each test was obtained (P). Since these were not normally distributed, a "log-odds index" was obtained for each class, ln(P/100-P).

The distribution of the index was normal and used for tests of significance. Descriptive tables in this report, however, give BSF results just as average percent of objectives mastered.

The longitudinal analysis used the same basic design, but in a "repeated measures" form, and with just that subset of students who were in the same experimental condition for three consecutive years. The dependent variables were differences in mean performance between K and grade 1 and between grades 1 and 2; in the second longitudinal analysis, they were differences between grades 1 and 2 and between 2 and 3. Only the SAT measures were scaled as to permit grade-to-grade comparisons of this sort.

The original three years of data are intercorrelated, because they are obtained on the same individuals over time. As a result, the two difference scores are correlated as well. Thus, multivariate repeated measures analyses were used to control statistical errors, in the manner described by Bock (1975). Individual year-to-year growth was examined, or its interaction with other corresponding factors in the design, only when the corresponding overall test was statistically significant.

While the global analyses used the procedures outlined above, other more specific analyses employed a variety of statistical methodologies. These are described in the following sections of the report, together with the results that were obtained. The analysis procedures employed were conservative and should have provided significant results only when there were considerable differences.

---

* This decision was made because means of all white males, minority males, white females and minority females would be based on very small and unreliable groups of youngsters. Also, the magnitude of a combined analysis would be unwieldy.

L. Achievement Results

1. Kindergarten Class Size Effect

STAR's kindergarten results showed definite advantage for small classes in achievement but no significant advantage for the use of a teacher aide. The overall superiority of the performance of students in small classes and the similarity of performance of students in regular and regular/aide classes are shown graphically in Figures 1 and 2 which present SAT scaled scores and percentile ranks on Total Reading and Total Math by class type and by grade.

2. First Grade Class-Size Effect

At the end of first grade, Project STAR students in small classes were outperforming students in regular and in regular/aide classes by substantial (statistically and educationally significant) margins on standardized tests and also on the state's Basic Skills First (BSF) test of reading and math. Small-class students scored at the 64th percentile in reading and the 59th percentile in math at the end of the first grade, while students in regular classes scored at the 53rd percentile (11 points lower) in reading and at the 47th percentile (12 points lower) in math. Students in regular classes with a full-time teacher aide outperformed students in regular classes in both reading and math. The presence of a teacher aide in grade one benefits student achievement but not as much as the small-class condition.(See Figures 1 and 2.)

3. Second Grade Class-Size Effect

Students in small classes continued to outperform students in regular and regular with a full-time aide classes on all tests in the second grade. There were significant advantages for students in small classes on the SAT in Reading, Math, Listening, and Word Study Skills, and a similar advantage on the Tennessee BSF tests in Reading and Math.

Although students in regular/aide classes outperformed students in regular classes, the differences were not significant. Students in regular/aide classes maintained their small achievement advantage over students in regular classes but did not increase their advantage. There is less consistency in the regular/aide advantage than in the small-class advantage.

Figures 1 and 2 present the scaled SAT scores and percentiles on Total Reading and Total Math by class type. Due to similarity of results on all subtests, the summary results presented here are confined to Total Reading and Total Math.

4. Third Grade Class-Size Effect

By grade three the pattern of results established in kindergarten had become firmly fixed. A strong class-size effect is evident in all school locations (urban, rural, inner-city, and suburban) and for all students on standardized and criterion-referenced achievement measures. The SAT scaled scores and percentiles in each of the three class types in third

Figure 1

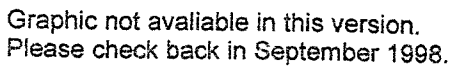
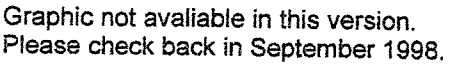


Figure 2



grade are shown for Total Reading and Total Math in Figures 1 and 2. The consistency of the finding of the small-class effect across all measures is important. The absence of a statistically significant teacher aide effect is consistent.

5. Summary of the Principal Analyses, Grades K-3.

A comparison of results for grades K, 1, 2 and 3 provides a picture of routine consistency. The classes of inner-city students consistently score lower on achievement measures than classes in the other three locations. (Note that most of the free-lunch students and a majority of the minority students were in the inner-city classes). The small-class effect is extremely strong (significant $p < .001$) in all contrasts. Students benefit from small classes wherever the small classes are located.

The effect of a regular class with a full-time teacher aide on student outcomes is less powerful and less consistent. There is some benefit to being in a class with a teacher aide in grade one, but that effect loses significance in other grades. A summary of the analyses showing significance levels (.05, .01, .001) is in Table 4.

Trained and untrained teachers did equally well across all class types. The small-class advantage and the absence of Regular/Aide effect is found equally in all four locations for trained and untrained teachers. There was no training main effect, or training-by-type interaction.

The small-class advantage and all effects found for total class generally apply equally to white and minority students, especially in grade 2. The race difference was statistically significant for all measures and multivariate sets, but not for most interactions (LxR, TRxR, TxR, LxTxR, or TRxTxR).

TABLE 4

Analysis of Variance for Cognitive Outcomes, STAR, Grades 1, 2, & 3, Sig. Levels p<=.05* or greater are Tabled. (All levels are <=.)

Effect/a	Grade	Reading			Mathematics		
		Multi-variate[b]	SAT Read	BSF Read	Multi-variate[b]	SAT Math	BSF Math
Location (Loc)	K	.01	.02	N/A	.01	.05	N/A
	1	.01	.06		.05		
	2	.001	.001	.001		.001	.001
	3	.001	.001	.001	.001	.001	.001
Race (R)	1	.001	.001	.001	.001	.001	.001
	2	.001	.001	.001	.001	.001	.001
Type (T)	K	.05	.001	N/A	.05	.02	N/A
	1	.001	.001	.001	.001	.001	.05
	2	.001	.001	.05	.001	.001	.05
	3	.001	.001	.001	.001	.001	.001
Train (TR)	2						
Loc X Race	1	.05	.05				
	2						
Loc X Type	K						
	1	All N/S. The class-size effect is found equally in all					
	2	locations--Inner City, Suburban, Urban, and Rural schools.					
	3						
Race X Type	1	.05	.05	.01			
	2						
LocxRxT	1	.05	.01				
	2						
LocxTRxT	2	.05	.01	.05	.05	.05	.01

* NOTE: Only statistically significant (<=.05) results are shown. aThe nonorthagonal design required tests in several orders (Finn and Bock, 1985). Results were obtained as follows: each main effect was tested eliminating both other main effects; Loc x race tested eliminating main effects and loc x type; loc x type tested eliminating main effects and loc x race; race x type tested eliminating main effects and other two-way interactions, and loc x race x type tested eliminating all else (Finn and Achilles, 1989). bObtained from F-approximation from Wilks' likelihood ratio. Essentially, no statistically significant differences were obtained on the self-concept and/or motivation (SCAMIN) measures.

6. Longitudinal Achievement Results

Project STAR researchers hoped that enough students would remain in the study to allow a strong longitudinal analysis. Although each year of the study included more than 6000 students, only 1842 were in the same class-size condition for all four years (K-3; 1985-1989) of the study. Since kindergarten is not mandatory in Tennessee, there was a fairly large influx of new students in first grade. Since at the end of kindergarten there were no differences between results of students in regular and regular with aide classes, students in regular and regular with aide classes were reassigned at random; students in small classes were not reassigned. This reduced the number of students who met the conditions for the longitudinal analysis, as newly entering students were excluded because they lacked kindergarten scores. Thus, researchers decided to do a longitudinal analysis that had two parts: Kindergarten-Grade 1 (K-1) and Grades 1, 2, and 3 (1-3). This decision produced more students, schools and classes for the analyses. (See Table 5)

TABLE 5

Number of Schools, Pupils and Classes by Type, Longitudinal Data Base: STAR, 1985-1989*

Groups	Schools	Students	Classes							
			Small		Regular		Regular/Aide		Total	
	N	N	N	%	N	%	N	%	N	%
K-3	54	1842	91	44%	51	25%	65	31%	207	100%
K-1	74	2416	115	38%	102	33%	90	29%	307	100%
1-3	60	2571	99	42%	64	27%	73	31%	236	100%

* In STAR in the same class type, for 4 years (K-3), or K-1 and 1-3.

To be considered in the K-3 longitudinal analysis, a student had to be in the same class type (small, regular or regular with a full-time teacher aide) for the entire project, and have the appropriate test scores. The revised analyses (K-1, 1-3) held to the same general rules: a student was in the study for the requisite number of years and had all of the required data points. Longitudinal analysis results should be treated tentatively due to the restricted subsamples (about one-third of the total group) in each analysis. Table 4 shows the longitudinal analysis of variance results expressed as significance levels.

TABLE 6

Analysis of Variance Results Expressed as Significance Levels, Project STAR, Longitudinal Analysis (1985-1989) Showing the Total Class Results and the Class Results by Race

	Word Study Skills		Total Reading		Total Math		Total Listening		
	K-1	**1-3**	**K-1**	**1-3**	**K-1**	**1-3**	**K-1**	**1-3**	
GRADE	.01	.001	.01	.001	.01	.001	.01	.001	[A]*
LOC X GRADE	.01	N.S.	.01	.01	N.S.	N.S.	---	N.S.	[B]
TYPE	.01	.001	.01	.001	.01	.001	---	.01	[C]
TYPE X GRADE	.05	N.S.	.01	N.S.	N.S.	N.S.	---	N.S.	[D]
LOC X TYPE X GRADE	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	---	N.S.	[E]
RACE	---	.01	---	.001	---	.01	---	.01	[F]
RACE X GRADE	.01	N.S.	.05	N.S.	N.S.	N.S.	---	N.S.	[G]
RACE X LOC X GRADE	.05	N.S.	N.S.	N.S.	N.S.	N.S.	---	N.S.	[G]
RACE X TYPE	---	N.S.	---	N.S.	???	N.S.	---	N.S.	[H]
RACE X TYPE X GRADE	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	---	N.S.	[I]
RACE X LOC X TYPE X GRADE	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	---	N.S.	[I]

* Results are discussed on the following pages using the designations [A]-[I] to identify the results being discussed.

N.S.= Not Significant; significance levels p<.001, .01, or .05 reported.

Discussion of the Longitudinal ANOVA Results (Table 6)

[A] There was statistically significant student growth on the standardized tests on all four measures and at all grade levels. This does not address class size.

[B] There were no statistically significant differences in pupil growth between/among the classes in the various locations (Inner City, Suburban, Rural, Urban) except in total reading for the 1-2-3 analysis where inner-city gained significantly more from G1 to G2 and from G2 to G3 than all other locations. In the K-1 analysis, there were statistically significant gains between/among class types in locations, with inner-city students gaining most in Total Reading and Word Study Skills. This result does not address class size. Note that the gain in all locations was fairly similar, with a range of 77.2 to 105.7 favoring the inner city. Inner-city class results were consistently the lowest and except for K, rural classes had the highest results. Also the largest difference between inner city score (lowest) and the highest score in any given year fluctuates from 49.4 to 24.4 with the superior gains in the inner city from G1 to G2 and from G2 to G3 reducing the differences.

[C] Small-regular contrast was significant on all scales at or beyond $p<.01$; aide-regular contrast is not significant for any scale.

[D] There was no interaction with class type over years 1-3 of the study. All class types grow equally, on the average. That is, the small-class advantage which originated in K neither increased or decreased in a statistically significant manner over the subsequent three years.

[E] There were no statistically significant Location x Class Type x Grade interactions on any measures.

[F] Race effects (1-3) significant on all scales at or beyond $p<.01$. Whites do better than minorities on all these measures. K-1 analysis was not run.

[G] In general, grade-to-grade growth in 1-3 was similar for whites (W) and minorities (M), although the differences for the average scores for W and M were considerably less on all four measures for small classes than for the other two class types. In K-1, whites' gains, on average, exceeded gains for minorities on word study and reading. Generally, grade-to-grade growth was the same for whites and minorities, regardless of location.

[H] There was no statistically significant differential impact of small classes on whites or minorities.

[I] There was no evidence of a differential impact of small classes on whites or minorities, as small classes affect "growth" in each year equally. That is, there may be differential impact on end-of-year performance but not on the total amount of change from K to 1, or 1 to 2 to 3 when students in the project are considered over time. There is no significant Race x Location x Class Type x Grade interaction. However, since there were only a few locations (i.e., school types) that had both white and minority students, the test of this effect is based on very small segments of the data.

Although each yearly analysis continued to identify the benefits of a student's being in a small class, the results for the small (about 33 percent) subsample of students in the same class size for 2 years (K-1) and 3 years (1-3) showed that the small class effect does not have a continuing cumulative effect after the large gains in K and in grade 1. The results showed that the large and statistically significant gains favoring the small classes made in the first year (i.e., K in the K-1 comparison and Grade 1 in the 1-3 comparison) were still evident in later years, but that there were no statistically significant gains in future years. Likewise, the average scores on measures of achievement used for the longitudinal analyses showed that the minority students in small classes achieved higher scores than minority students in the other class conditions, but the non-minority students continued to outperform the minority students in all class types and locations.

Combining year-by-year and longitudinal results suggests that 1) a student's achievement and development are greatly improved if the student is in a small class, 2) the small-class experience is more successful if in K or Grade 1, and 3) small-class condition gains remain in the small-class condition.

M. Summary of Non-Cognitive Results

Being in a small class did not have an impact on student self-concept and motivation as measured by the SCAMIN. Students in the inner city had somewhat higher self-concept scores than students in the other locations. Self-concept measurement of young children is difficult and results may become more stable in later years.

Students in small classes in kindergarten had significantly higher self-concept scores but not higher academic motivation scores. Classes effective in improving achievement measures are not necessarily effective in achieving positive non-cognitive results (X^2=11.71, $p<.05$, df2). There are positive ($p<.05$) relationships between each of the achievement measures and self-concept but not between achievement measures and the non-cognitive measure of achievement motivation.

The self-concept (SCAMIN) results in grade one generally were not significant based upon class size, but there was a statistically significant result based upon school location with inner-city students scoring higher than students for other locations. Essentially the same pattern of results (with minor variation) was found for the SCAMIN results in kindergarten and grades two and three.

Approximately 77 percent of the small-class average scores in first grade were some higher (not significantly) than the regular or regular/aide class average scores on the self-concept measures (SCAMIN). Thus, the conclusion is that self-concept was the same for students in small classes, regular with full-time teacher aide classes and in regular classes. In second grade self-concept and motivation differences as measured by SCAMIN results tended to be minimal and non-significant, but students in the inner city (primarily minority students) continued to have higher self-concept scores than did students in the other three locations.

In third grade the differences in SCAMIN results by location were considerably more marked than in K, 1 and 2 and showed that the inner-city students had significantly higher scores than did the students in classes in the other three locations. There was no significant class-size effect for SCAMIN results; students in all three class types scored about the same wherever the classes were located. By grade three, inner-city students had higher self-concept and motivation scores as shown on the SCAMIN. The inner-city students were predominantly minority in the STAR database.

N. Summary of Achievement Results Based on Effect Sizes*

1. Students in small classes have higher performance than regular and regular/aide classes in all locations and at every grade level.

Each of the four years, small-class students in both reading and math (as well as in other SAT subtests) achieved significantly higher test scores than students in regular classes. Figure 3 shows these differences expressed as effect sizes, for both reading and math. Small classes were constantly higher in performance.

There was a significant positive small-class effect for both reading and math at the end of kindergarten, the effect increased at Grade 1, then declined in Grades 2 and 3. Analysis of grade-to-grade gains showed that score gains in the first grade were about 15 percent larger in small classes than in regular classes, but that after the first grade, gains for both reading and math were as large, or slightly larger in regular classes as in small classes.

2. Small-class effects diminish after first grade

The small-class effect is concentrated in kindergarten and Grade 1. Thereafter the small-class effect declines slightly, but is still significant at the end of Grade 3.

This finding suggests that class size reduction should be concentrated in kindergarten and Grade 1, where effects will be greatest. This reasoning is confirmed by an analysis of the class size effect for new students who entered the project each year. The new entrants to the project allow class size effects each year to be compared with the cumulative effects for students who have been in the project from the beginning. The effect size for **new** students is about the same in reading in kindergarten and Grade 1, declines slightly in Grade 2, and is very small at Grade 3. For math, the class size effect is highest at Grade 1, not significant at Grade 2, and is fairly similar for kindergarten and Grade 3.

New student effect sizes also suggest that small classes should be concentrated in kindergarten and Grade 1. Effect sizes for the continuing students are always larger than the effect sizes for the new students, which is to be expected, because continuing students have had the benefit of the small class for more than one year. The effect size "advantage" of the continuing students over the new students averaged over math and reading is at approximately the same level in Grades 1, 2, and 3. This also indicates that there is no additional class size effect after Grade 1. (See Figure 4)

There are numerous possible explanations for larger effects in kindergarten and Grade 1, one is that it is more difficult to manage students who are not well socialized to the classroom routines. By the time children get to the second and third grades, they are better socialized, and the teacher can manage a larger group effectively. Another is that one year in a small class may serve to get a student "on track" or "up to speed" and subsequent years did not add to this benefit. This explanation would be similar to results obtained in the Reading Recovery projects.

---

* Results reported here are based upon analyses conducted by Dr. John Folger, Vanderbilt University. Dr. Folger employed slightly different decision rules in selecting a sample for analysis from the STAR database. For example, as there were no differences between student performance in classes of trained and untrained teachers, Dr. Folger retained the classes of trained teachers; the primary analysis excluded them. The parallel analyses were confirmatory; they produced essentially identical results.

Figure 3

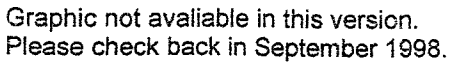
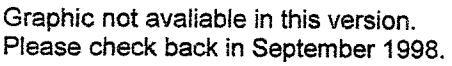


Figure 4



3. Aides were less effective than small classes in enhancing student performance at each grade level.

Classes with a full-time aide had higher achievement scores than regular classes in kindergarten through grade two but the differences were small and not statistically significant in kindergarten and second grade. In grade three the regular/aide classes' scores were slightly lower than the regular classes. In the first grade, regular/aide classes were significantly higher than regular classes in both reading and math.

In grades one, two and three regular classes had the part-time services of Basic Skills aides; on the average they were available to each regular class about 25-33 percent of the time. The basic comparison is between a regular class with one-fourth to one-third time services of an aide, and a class with a full-time aide.

Aides performed a wide variety of clerical, custodial, and instructional tasks. The pattern of aide activities was not related to student achievement. Aides who performed mostly instructional tasks did not enhance student achievement any more than aides who did only clerical tasks.

Teachers liked teacher aides. In a forced choice, about 45 percent of teachers who had an aide preferred an aide to a small class, and 55 percent favored the small class, but the bottom line is that teacher aides did not have much effect on student learning in Project STAR.

4. Math and reading effects are similar.

In a meta analysis of well controlled studies Glass (1984) estimated the **average** effect size for reading was .11, and for math it was .22 (reduction in size of 43 percent, from 35 to 20). The Project STAR effect size (averaged over four years) is .26 for reading, and .23 for math. Glass speculates that reading effects are smaller because reading instruction is done in small groups, where the overall size of the class makes less difference. Math instruction, on the other hand, is done whole group, and class size makes more of a difference. Glass's explanation did not fit Project STAR, where nearly all teachers used small groups for reading instruction but math instruction was almost all whole class. Project STAR found that class-size reduction had similar effects for all of the SAT subtests; it did not have differential effects in different subjects. Effect sizes in Project STAR were larger than those found in other well controlled studies. Slavin (1986) estimated an average effect size for smaller classes of .13, about half the Project STAR effect size. Since more positive effects of small classes have been reported for early elementary grades (Robinson, 1990), Project STAR's larger effect sizes may be because it was limited to Kindergarten through Grade 3.

5. Small classes help low socioeconomic student achievement, but they help high SES student achievement about as much.

In reading at each grade level, effect sizes for low SES students exceeded those for high SES students. At Grade 2 the difference was substantial (see Table 7). In math, by contrast, effect sizes for high SES students exceeded those for low SES students except at Grade 2 where they were about the same.

Low socioeconomic students scored lower than high SES students on the average, but there were many exceptions. To study the effect of small classes on low academic achievers, the scores of students in the bottom quartile were compared to their scores at the end of the next year to determine if a small class helped them more than a regular class.

TABLE 7

Effect Sizes for Small Classes by Grade, SES, and Achievement Level Reading and Math

Test and Group	Small-Regular Effect Size			
	Kindergarten	Grade 1	Grade 2	Grade 3
Reading				
All	.21	.34	.26	.24
High SES	.19	.32	.20	.21
Low SES	.23	.35	.33	.25
Bottom quartile, previous year	---	.26	.12	.12
Math				
All	.17	.33	.23	.21
High SES	.20	.34	.21	.20
Low SES	.14	.30	.22	.18
Bottom quartile, previous year	---	.09	.25	.23

The effect sizes for the lower quartile students were below the overall effect sizes for reading at each grade, and for math at Grade 1. At Grades 2 and 3 math effect sizes were about the same for the lower quartile and all students (see Table 7).

These results indicate that there is no differential effect of a small class that favors low achieving or low SES students over average students or high SES students. The class size effect is "across the board" for all students.

O. Small Classes Reduce Grade Retention.

A smaller percent of students in small classes are retained each year than students in regular classes. Since grade retention has been shown by previous research (Shepard and Smith 1989, CPRE, 1990) to reduce students' chances of graduating, compared to equal ability students who are not retained, this is an advantage of small classes. Teachers were more willing to promote marginal students in small classes. Over the four years of Project STAR, 19.8 percent of the small-class students were retained, as compared with 27.4 percent of students in regular classes. Seven and one-half percent fewer students had to repeat a grade in the small classes, this meant about a two percent a year reduction in cost per grade. It could also save costs later because promoted students have a greater chance of completing school, and avoiding delinquency and unemployment.

P. Teacher In-service Training Did Not Improve Student Achievement.

One of the reasons offered in the literature for class size not making a difference is that teachers do not change the way they teach when they have a smaller class (Robinson and Wittebols, 1986). Project STAR specified that there should be training for teachers, so a subgroup of 57 teachers in thirteen randomly selected schools in Grade 2 , and another 57 teachers in the same schools in Grade 3 were given three days of in-service training before school started. The training was designed to help them to teach more effectively in whatever

class type they had been randomly assigned to teach. There were not significant differences in student achievement in reading or math in either the second or third grade between classes where the teachers were trained and all the other classes where the teachers had not received special training. (See Table 8)

TABLE 8

Stanford Achievement Test Scaled Score Gains in Reading and Math for Students in Classes where Teachers were Trained or Not Trained in STAR Training Program

	Total Reading		Total Math	
	Trained	Not Trained	Trained	Not Trained
Grade Two	58.6	58.2	46.5	45.3
Grade Three	25.7	27.4	31.9	34.1

In exit interviews at the end of the year, about half the STAR trained second grade teachers said they had not modified their teaching as a result of the training. It is not surprising that the training program did not lead to improved student performance under these conditions. Although the statistical finding for differences in teacher behavior between class sizes and for trained and untrained teachers were not strong, many valuable findings emerged:

1. If instructional goals are to increase the development of higher-order thinking skills, creativity, and personal responsibility for learning, a reduced teacher/student ratio may be important to enable teachers to achieve these objectives effectively. Fewer rote tasks, more reading and writing in context, more problem-solving activities-- all will require more teacher/student interaction than the present curriculum. If such broad changes in learning outcomes are desired, changing class size and training teachers alone will not be enough; these changes must be coupled with a curriculum focused on these objectives.

2. Teachers with small classes must be willing to receive training and be committed to try new skills and procedures.

3. Training should include effective in-service that provides:

   a. time for teachers to visit other teachers who have had success in teaching small classes and

   b. training in the following skills:

      (1) Ability to establish effective communication with the home.

      (2) Ability to involve the family in the education of their children.

      (3) Ability to make home visits. These should be made during in-service time or during school time with a substitute provided.

4. This improvement effort must be encouraged and strongly reinforced by principals, local system supervisors, and state department personnel.

Q. Teacher Grouping Practices

The Instructional Grouping Practices Questionnaire collected data from K-3 teachers. Few differences in grouping practices were observed among K-3 small, regular, and regular/aide

class teachers. All teachers formed small groups for reading and most often reported using three groups. Math instruction was generally carried out with the whole class, but small and regular/aide class teachers reported using two or more groups for math more often than the regular class teachers. Teachers in small classes reported using fewer reading groups than the regular and regular/aide class teachers. However, neither of these differences was statistically significant. Almost no teachers reported the use of instructional groups for science or social science. Students, who were most often assigned to groups based on their skill level, were occasionally moved from one group to another.

R. Parent/Volunteer-Teacher Interaction

The Parent/Volunteer-Teacher Interaction questionnaire, a self-report instrument completed by Project STAR K-3 teachers, indicated the weekly, monthly, and yearly frequency of various types of contacts with parents and other classroom volunteers. In addition, teachers reported the monthly and annual number of home visits and indicated their overall satisfaction with the level of parent-teacher interaction. Regular class teachers reported more frequent involvement of parents in classroom activities and support than did small or regular/aide class teachers. Throughout grades K-3, having a full-time teacher aide appeared to reduce the need for and hence the frequency of involvement of parents or volunteers in classroom activities. Small class teachers were more likely to phone, write, or confer with parents about student accomplishments and good behavior than were regular or regular/aide class teachers. In addition, small class teachers reported less frequent communication with parents regarding student misbehavior or learning problems. One explanation for this finding is that small class teachers reported that they were better able to prevent problem behavior from happening, were better able to solve misbehavior problems in class, and had more time available to solve problems when they arose. Thus, small-class teachers may not have felt the need to involve parents in solving classroom behavior problems. Regular/Aide class teachers, with their built-in source of clerical assistance, made less frequent use of parents or volunteers than did small or regular class teachers. However, no difference discussed above was statistically significant. Only 10 to 15 percent of the K-3 teachers reported making professional visits to students' homes and no class-type differences were observed.

S. Teacher Problems

To provide the project with data on teachers' perceptions of their work- related problems, Project STAR teachers completed a slightly modified version of the Teacher Problems Checklist (Cruickshank & Myers, 1980). This instrument, modified by the addition of a single item regarding teacher aides, consists of 61 problem statements. For each of the 61 specific problem statements, teachers provided information about the extent to which the problem was perceived to be bothersome and the frequency with which the problem was experienced.

The three problem statements: (a) I have a problem having enough time to teach and also to diagnose and evaluate learning, (b) I have a problem having enough preparation time, and (c) I have a problem having enough free time, were consistently the top-ranked problems both for Bothersomeness and Frequency for all teachers, K-3. No significant differences were observed between class type and teacher-perceived problems.

T. Teacher Exit Interview Data

Project STAR K-3 teachers were interviewed by consortium staff at the end of each school year from 1986 through 1989. Over a four-year period 1,003 kindergarten, first, second, and third grade teachers were interviewed (Table 9).

TABLE 9

Teacher Interviews

	Kindergarten	Grade 1	Grade 2	Grade 3
SMALL	128	126	86	88
REGULAR	101	113	54	55
REGULAR/AIDE	99	107	71	70
TOTAL	328	346	211	213

The number of interviews decreased in second and third grades in accordance with a consortium decision.

Kindergarten Teacher Interview Procedures

The primary question explored in depth with each kindergarten teacher was: If your experience was different this year than last year, then how was it different? A three-stage iterative analysis was performed on the kindergarten data. In the first stage, teachers' interview response statements were examined to define common themes. In the second stage, interview statements were categorized along the dimensions of those themes. In the third stage a random set of responses, which had been set aside at the outset of the analysis, was used to check the reliability of the theme categories and the coding process. Fourteen categories which identified teachers' perceptions of teaching either in a small class, a regular class with no aide, or in a regular class with a full-time aide are described below.

1. Grouping - described classroom grouping practices, e.g., number, membership and purpose of groups.
2. Physical environment - described the amount and use of classroom space, furniture arrangement, heat, light, noise level, and traffic patterns.
3. Learning centers - described the use, quality, and perceived effects of learning centers.
4. Social climate - referred to social interactions among students and between teacher and student: references to cooperation among children, and teacher knowledge of each child's personal and academic strengths and weaknesses.
5. Enrichment activities - described those experiences that provided additional learning opportunities, e.g., cooking activities, special art, music or drama, field trips, and invited guests in the classroom. It also included references to planning and carrying out enrichment activities.
6. Classroom management - described problem behavior and techniques to prevent and deal with it.
7. Monitoring and evaluating student progress - described ways used to monitor student progress and give students feedback about their progress.
8. Morale and attitude of teachers - described positive or negative/energy, outlook, level of frustration and stress, degree of satisfaction, physical and mental health and well-being.

9. Amount or rate of student progress - described the amount of material covered and how quickly students grasped the material.

10. Parent-teacher relationships - described how parents were involved, parent-teacher communication, and home-environment factors.

11. Teacher aides - described responses about having or not having an aide, quality of aide, aide duties, and aide characteristics.

12. Instruction - described use of instructional time, purposes, curriculum, instructional goals, teaching methods and techniques, and degree of structure.

13. Teacher planning and preparation - described planning class activities, preparation of teaching materials with references to paperwork, copying, duplicating, stapling, record keeping, collecting money, etc.

14. Individual attention to students - described one-on-one attention or instruction of students, with references to reteaching and reinforcement of content as well as student counseling.

First Through Third Grade Teacher Interview Procedures

The **first grade** interview forms included the 14 areas identified from the kindergarten interviews. Two additional questions were added:

1. If you had your choice, which teaching situation would you choose:
   a. a small class with 15 children
   b. a regular class with 25 children with a full-time aide
2. If you had your choice, which teaching situation would you choose:
   a. a small class with 15 children
   b. a $2,500 salary increase

The **second grade** interview form contained basically the same areas used the previous two years. Questions from the second grade interview asked teachers to describe differences, if any, that they perceived regarding the following dimensions: (a) amount of content covered, (b) amount of instructional time on task, (c) monitoring children's work, (d) ability to match level of instruction to the ability of individual students, (e) pacing of instruction, (f) degree of active student-teacher academic interaction, (g) individual attention to children, (h) classroom social climate, (i) demands on available teacher time, and (j) use of full-time teacher aide. The **third grade** exit interview form was essentially the same as the second grade.

Based on four years of interviews, patterns emerged in kindergarten and continued through the third grade. The following advantages were apparent for instruction in small and regular/aide classes:

1. basic instruction was completed more quickly, providing more time for covering additional basic material,

2. use of supplemental text and enrichment activities,

3. more in-depth instruction regarding the basic content,

4. more frequent opportunities for children to engage in first-hand learning activities using concrete materials,

5. increased use of learning centers and

6. increased use of highly desirable primary grade practices.

Improved individualization of instruction emerged as a dominant theme in small and regular/aide class teachers' perceptions. Teachers reported: 1) increased monitoring of student behavior and learning, 2) opportunities for more immediate and more individualized reteaching or enrichment, 3) more frequent interactions with each child, 4) a better match between each child's ability and the instructional opportunities provided, 5) a more detailed knowledge of each child's needs as a learner, and 6) the necessary time to meet individual learner's needs using a variety of instructional approaches. Significant reduction of class size or the addition of a full-time teacher aide also made positive changes in the physical, social, and emotional environments in primary grade classrooms. Classrooms were more pleasant work environments for both teachers and students. Teachers and students were under less stress, and learning occurred in a more relaxed atmosphere. Students were less likely to get lost in the crowd and were more likely to have their own unique needs met by adults who had a good understanding of them as individuals. The extent to which teachers, aides, and children were friendly, supportive, and trusting of one another was an indicator of the classroom morale and the sense of team spirit that is characteristic of effective elementary schools.

The teachers' perceptions of the value of small class size can be seen in the third grade teachers' choices of a small class, a full-time aide, or a salary increase (see Table 10 and Table 11.

TABLE 10
Preferred Teaching Situation Of
Small, Regular, and Regular/Full-Time Aide Teachers

	Class Type							
Teacher Preference	**Small**		**Regular**		**Regular/Aide**		**Total**	
Small Class	88	(81%)	29	(71%)	46	(56%)	163	(71%)
Regular/Aide Class	20	(19%)	12	(29%)	36	(44%)	68	(29%)
Total	108	(100%)	41	(100%)	82	(100%)	231	(100%)

TABLE 11
Teacher Preference for a Small Class or a Salary Increase

	Class Type							
Teacher Preference	**Small**		**Regular**		**Regular/Aide**		**Total**	
Small Class	73	(70%)	22	(48%)	52	(63%)	147	(63%)
$2,500 Salary Increase	32	(30%)	24	(52%)	31	(37%)	87	(37%)
Total	105	(100%)	46	(100%)	83	(100%)	234	(100%)

U. Although reducing class size is more expensive than adding a full-time teacher aide, it is more cost effective.

The cost of reducing class size by one third is primarily the additional salary cost of adding teachers, and the capital costs for new classrooms that must be added. Reducing class size from 23:1 to 16:1 statewide in K-3 would require about 175-180 million dollars in additional operating expenses. If we assume that 20% of these classes are available in schools now, the additional capital costs would be 21-25 million each year amortized over 30 years for a total annual cost of 196-205 million. The need for additional classrooms could be eliminated by the implementation of year round schools. Reducing class size just in K and 1 would cost a little less than half the total (kindergarten is about 10% smaller than Grade 1) or about a 100 million dollars. It would add about 30-32 percent to the current cost per student. Adding a full-time aide in Grades K-3 would add about 75 million dollars, if the aide were only added in Grade 1 where the only aide effect was found, the cost would be about 19-20 million dollars.

If a reduction in class size is to be done in phases the program should begin in grade one with classes of 1 to 15 because that is where the greatest small-class effect was found and where the cost effectiveness would be greater. Small classes will have the greatest cost effectiveness when teachers use those teaching practices best suited for small classes. A small class provides an opportunity to do things better and differently and break out of the "more of the same" mindset. Teachers can use new teaching strategies. Home visits and increased involvement of adults or parents in the education of their children, team learning strategies, individual programming (and remediation) for each student, improved screening for physical and learning disabilities are all possible with small (1:15) classes. Small classes may be seen as a minimum foundation program which will allow variations or additions previously desired but untried due to excessive "case loads" for classroom teachers. These types of changes may require extensive training and practice before substantial benefits are achieved. The Star training program pointed out the need for more in-service with a new approach.

V. Estimates of the Magnitudes of the Differences (Grades K,1,2,3)

One important question in this study was "How large are any small class and regular with teacher-aide class advantages?" The magnitude of difference begins to get at the policy questions upon which this study was founded and to explore the educational significance of the statistically significant results obtained.

The "small-class" advantage is evident; it increases in K and 1 and decreases thereafter. Gains realized in K and Grade 1 remain evident, but decreased in grades 2 and 3. The teacher-aide advantage, like the small-class advantage, is most pronounced in grade 1, and it declined thereafter. There is no important teacher-aide advantage in K.

There is a consistent and fairly large scaled score difference favoring the small class over the regular class at each grade (approximately 10-12 in total reading and 8-11 in total math). This difference is also reflected in the higher percent of BSF criterion-referenced test items answered correctly by students in the small-class condition. These results are summarized in Tables 12 and 13 for the differences in performance of white and minority and all students in small and regular classes for the SAT Total Reading and Total Math (K-3) and the percent passing difference on the BSF (1-3; no K test). The SAT differences are effect sizes; the BSF are percents.

TABLE 12

Summary of Estimates of Small Class Effect Sizes on Total Reading and Total Math, Grades K-3 Project STAR, 1985-1989.

	Group	Kindergarten	Grade 1	Grade 2	Grade 3
Total Reading	White	.18	.25	.19	.17
	Minority	.25	.52	.42	.32
	ALL	.21	.34	.26	.24
Total Mathematics	White	.20	.25	.19	.17
	Minority	.09	.38	.27	.22
	ALL	.15	.33	.23	.21

TABLE 13

Differences in Average Percent Passing BSF Test of Reading and Math Between Small Classes and Other STAR Classes, Grades 1, 2, and 3

	Group	Grade 1	Grade 2	Grade 3
BSF - Reading	White	4.8%	1.6%	4.0%
	Minority	17.3%	12.7%	9.3%
	ALL	9.6%	6.9%	7.2%
BSF - Mathematics	White	3.1%	1.2%	4.4%
	Minority	7.0%	9.9%	8.3%
	ALL	5.9%	4.7%	6.7%

W. Conclusions

The design and magnitude of Tennessee's randomized class size experiment (STAR) allow researchers to make, with high levels of confidence, statements about class-size effects. Here are some examples from prior reports. "This research leaves no doubt that small classes have an advantage over larger classes in reading and mathematics in the early primary grades" (Finn and Achilles, 1989:21). "This experiment yields an unambiguous answer to the question of the existence of a class-size effect, as well as estimates of the magnitudes of the effect for early primary grades" (p.22). "These data confirm that a small-class effect, while not immense, is found in two basic subject areas, at four grade levels, and in all four school settings...Few, if any, other classroom-level interventions have been identified that have a consistent impact of this sort" (Finn, et al., 1989: 15-16).

Numerous papers have been developed and presented at national, regional and state meetings and conferences. Some articles based on STAR data and concepts have been disseminated. These and other detailed papers and technical reports are available from