

DOCUMENT RESUME

ED 390 913

TM 024 262

AUTHOR Achilles, C. M.; And Others  
 TITLE The Multiple Benefits of Class-Size Research: A Review of STAR's Legacy, Subsidiary and Ancillary Studies.  
 PUB DATE Nov 94  
 NOTE 30p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Nashville, TN, November 9-11, 1994).  
 PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)  
 EDRS PRICE MF01/PC02 Plus Postage.  
 DESCRIPTORS \*Academic Achievement; \*Classroom Research; \*Class Size; Databases; Educationally Disadvantaged; Low Achievement; \*Primary Education; Racial Differences; Research Reports; \*Research Utilization; School Districts; School Size; \*State Programs  
 IDENTIFIERS Project Challenge TN; \*Student Teacher Achievement Ratio Project TN

ABSTRACT

This review of research connected to Project Student Teacher Achievement Ratio (STAR) is a summary of the project's ongoing work in the form of a letter from researchers to a hypothetical colleague. Project STAR has investigated the effect on student achievement and development of small classes in the primary grades (K through 3). A research consortium of four universities and the State Education Agency was formed to conduct and monitor the study with the aid of an advisory panel and outside consultants. Forty-two districts and 79 schools (later 76) participated. Three major subsidiary studies have built on STAR: (1) the Lasting Benefits study that tracks students; (2) Project Challenge, a policy application of small classes in 16 low-performing school systems; and (3) the Grade 4 (and eventually Grade 8) participation study. The database established for STAR, which now includes about 9,000 students randomly assigned to small-class, regular-class, and regular-class with aide conditions, is serving as the basis for other research, including examinations of achievement, racial differences, the interaction of school size and class size, and other state initiatives. Three appendixes present tables of research design, analysis of variance for cognitive outcomes, and rankings of Challenge districts. (Contains 8 tables and 33 references.) (SLD)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

# THE MULTIPLE BENEFITS OF CLASS-SIZE RESEARCH: A REVIEW OF STAR'S LEGACY. SUBSIDIARY AND ANCILLARY STUDIES

A Paper Presented at the Annual Meeting of the  
Mid-South Educational Research Association  
(MSERA)

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. M. ACHILLES

Nashville, Tennessee  
November 9-12, 1994

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)

BY

C.M. Achilles

Eastern Michigan University

B.A. Nye

Tennessee State University, and the Center

J. Boyd-Zaharias

of Excellence for Research on the

D. Fulton

Basic Skills at Tennessee State University

V. Cain

## Precis

Using an epistolary format, the researchers explain to a doubting peer some of the continuing work on the Project STAR data, and on data arising from added class-size studies that build upon the STAR database. The research, begun in 1985, is still going on and being extended. Reanalyses of the original data are providing added questions and answers regarding class sizes and pupil learning as measured by various tests.

BEST COPY AVAILABLE

1024262

## Abstract

### The Multiple Benefits of Class-Size Research: A Review of STAR's Legacy. Subsidiary and Ancillary Studies

Although Project STAR has been around for nearly a decade, it is only now reaching the "maturity" when it can be engaged to answer a wide variety of questions. Each year, researchers use the STAR design and data to build a knowledge base on class-size issues. Tentative plans are to follow a large STAR cohort through grade 12.

Authors will briefly review STAR, Tennessee's longitudinal randomized experiment (1985-90). They will discuss three major subsidiary studies that build upon STAR (Lasting Benefits Study or LBS that is tracking STAR students; Project Challenge that is a policy application of small classes in 16 low-performing Tennessee systems; and the Grade-Four -- and Grade-Eight if it is ready -- Participation Study). They will describe how the large, carefully developed STAR database offers a "mother lode" of opportunity to mine rich ancillary questions relating especially to early elementary education. Ancillary studies use the STAR database, but are further removed than the subsidiary studies.

Besides the reviews of design and study results, researchers discuss how a database established specifically for one purpose supports a stream of research deriving from the primary research question. Two major criticisms of STAR encouraged the researchers to review the study, to make adjustments and re-analyze some data. The revised results substantiated and strengthened the original analyses, nearly doubling some Effect Sizes (ES).

The STAR/LBS database includes approximately 9,000 pupils randomly assigned in K-3 to one of three conditions: Small class (S) of about 1:15; Regular class (R) of about 1:25; and a Regular class with a full-time Aide (RA). Districts in STAR provided test-score results from 21 "comparison" schools of similar demographics to STAR schools for grades K-3. LBS has followed approximately 4,500 pupils from STAR into grade 8 (1994). Data include test scores on the Stanford Achievement Tests (SAT) and Tennessee Comprehensive Assessment Program (TCAP), responses to questionnaires and interviews, and demographics.

Employing purposive subsamples from STAR data, from comparison schools and from LBS data, researchers have used descriptive, multivariate and univariate analyses to answer to such questions as: 1) What is the test-score "value" of K to pupils (by race) in grades 1, 2 and 3?; 2) Is small-class placement a remedial strategy for test-score gaps between white and non-white pupils? 3) How do class size and school size interact in early grades?; 4) How does early class-size placement influence pupil participation in and identification with school?; 5) Does small-class placement a) reduce retention in grade, and/or b) help achievement of pupils once retained?; 6) How long do early test-score benefits remain for pupils from the (S) condition and how much do they fade (LBS)?; How do STAR/LBS results apply to Project Challenge? How does the Tennessee Value-Added Assessment System (TVAAS) support Project Challenge? Researchers are interested in group discussions and the potential that participants can identify new questions for analysis using the databases. Researchers hope to engage the audience in considering ways to make small classes (1:15) more palatable to policy and funding agencies.

The Omphalos Research Center

April 1, 1984

Dr. Ima Snipe  
Skeptos Institute  
Showme, Missouri

Dear Dr. Snipe:

Although this may sound insincere, we actually thank you for belatedly raising questions about the line of research that we have been exploring and reporting for the past eight years. Belatedly, we believe, because your questions go back to the initial analyses that we reported beginning more than five years ago. We have been continuing our work and have moved well beyond some of your concerns. Had you contacted us directly for clarification (rather than making tenuous assertions and unsupported generalizations in non-refereed publications and "in-house" documents) we might have answered your questions more exactly and in a timely fashion.

On the other hand, careful critique keeps researchers honest and humble. Without your skepticism and questioning, we probably would not have returned to the original, excellent database -- and that would have been an error. The tone of your questioning served only to fan our ardor and set a "we'll show 'em" tone that ensured more thorough analyses.

Yet, in spite of the elation of "completing" a study there always are twinges of underlying doubt in good research: Did we do everything right? So, while some people may tire of their seemingly endless repetitions, peer review and critique have clear benefits. Not only does peer review help "keep us honest," it brings to the problem a new set of eyes; it may ask questions beyond those that we were satisfied to answer; it may interject ideas based upon a different reading and interpretation of the research

and literature supporting the original line of inquiry; it may identify gaps in the rigor or breadth of the analyses; and may make connections with other studies and related fields that we skipped over, etc.

In this serious vein, then, we have considered carefully the concerns that you raised. Your questions started us on the task of re-analyzing the initial database in slightly different ways, and we reformulated some of the initial questions in slightly different ways. Perhaps we get ahead of ourselves; let's return to the original purpose of the initial study and start there to take another journey through the database and analyses. In constructing this revisitation, I have followed an outline format. For ease in following the narrative, here is the outline format.

- I. Some History.
  - A. Background
  - B. Original Study Purpose
  - C. Synopsis of Methods (Appendix A)
  - D. Synopsis of Results (Appendix B)
- II. The Derivative Studies (LBS and Challenge).
- III. Some Ancillary Studies (class size/school size, test-score value of K, retention in grade, test-score "gap" reduction, homogeneous vs. heterogeneous grouping and achievement, etc.).
- IV. Questions and Answers Raised by Your Critique; References and Bibliography.

#### I. Some History

##### A. Background

In 1985 the Tennessee legislature commissioned Project STAR (Student Teacher Achievement Ratio) to try to get some answers to the basic question, "What is the effect on pupil achievement (and development) of small classes (e.g., about 1:15) in early primary grades (K-3)?" A research consortium of four major Tennessee universities and the State Education Agency (SEA) was formed to conduct and monitor the study aided by an advisory panel and outside consultant help. All Tennessee districts (n=140) were invited to participate, resulting in 42 districts and 79 schools (later reduced to 76). Selected districts did not differ from others in the state except slightly in size (the three largest systems were in the

final sample) and fairly represented urban, rural, inner-city and suburban areas. Districts agreed to STAR procedures and to remain in the study for four years. Researchers manipulated only one variable, class size, and were to assure that no pupil received any diminution of services by being in STAR.

#### B. Original Study Purpose

The study was to answer the basic research question (see IA above) and to try to get a definitive answer to the class-size issue and debate (e.g., Glass & Smith, 1978; Glass et al., 1982; Cahen et al., 1983; Education Research Service or ERS, 1978 & 1980). The policy makers in Tennessee wanted this information as a basis for setting state regulations on class sizes (Tomlinson, 1988, 1990).

#### C. Synopsis of Methods

Researchers assigned pupils in kindergarten (K) in the participating schools in 1985 to one of three class-size conditions: a Small class (S) of about  $n=15$  with a range of 13-17; a Regular class (R) of about 24 with a range of 22-26, and a Regular class with a full time aide (RA). Once designated as S, R, or RA a class remained so designated for the duration of the study and pupils remained in the class conditions. Pupil mobility was handled by random replacement. Since Tennessee did not have mandatory K during STAR, the increase in students to STAR in grade one required researchers to establish some additional classes. Use of an "in-school design" where each participating school had at least one class of each condition (S,R,RA) helped control for building-level and district-level variables. In all analyses there were approximately 100 classes of each of the three conditions. A multivariate analysis of variance (MANOVA) was used (Finn & Bock, 1985). The basic design has been reported elsewhere (e.g., Finn & Achilles, 1990) and is in most papers and documents generated for the study (e.g., Word et al., 1990). Appendix A is a summary of design and analysis steps and is reproduced from Achilles et al., (1993, pp. 617-618).

#### D. Synopsis of Primary Findings.

The rather straight forward analysis was of the class-size means (although data were collected on pupils on both norm-referenced tests or NRT and criterion-referenced tests or CRT, as well as on other measures, data were analyzed as class means as this was a study of class-size effects) of the three class-size conditions (S,R,RA). Results consistently (grades K-3) showed that pupils in the (S) condition

outperformed pupils in other conditions. Generally the (R) and (RA) results were quite similar. These results have been reported in detail elsewhere (e.g., Finn & Achilles, 1990; Word et al., 1990). Appendix B is a summary of results reproduced from Achilles et al. (1993, p. 619).

Small classes in this study outperformed (R) and (RA) classes in all locations and on all cognitive measures, generally at or beyond  $p \leq .001$ . This translated into effect sizes (ES) ranging from .25 to .44, depending upon the comparisons. These results came from a carefully designed longitudinal (K-3) experiment with random assignment and the manipulation of just one variable, class size. Since the results were so clear and consistent (note Appendix B), and the answer to the basic question so unambiguous, researchers essentially left the STAR database and turned their attention elsewhere.

## II. The Derivative Studies (LBS and Challenge)

After answering the class-size issue posed in STAR's enabling legislation, researchers sought and obtained modest funding to follow as many STAR pupils as possible as they moved out of the experimental conditions and into "regular" classes at grade 4 and beyond. This was called the Lasting Benefits Study (LBS). By 1993-94 many STAR pupils were in grade 8. Data analyses for grades 4-6 show continuing achievement benefits ( $p \leq .01$  and ES from .15-.25) for pupils who had been in (S) classes even three years after their return to "regular" classes. (Analyses are in process for grade 7.)

Starting in 1989 leaders in Tennessee made funds available in 17 of the state's poor counties for broad-scale class-size reduction (Project Challenge). Using only the gross measure of ranking, on average these 17 systems have increased their rank 12 places in reading scores and 26 places in math scores (grade-two data) between 1989-90 and 1991-92 among the state's 138 districts. (See Appendix C for a summary.)

Some researchers recognized that STAR and LBS were generating a useful database and so they framed other questions that the database could help answer. We are just getting started with this series of studies which we are calling the Ancillary Studies. Even though the database is again attracting modest attention, researchers had not decided to return to the original STAR analyses until there were some questions raised about STAR in other writings (e.g., Tomlinson, 1988 & Mitchell et al., 1983). Then, of

course, there are the concerns that you have raised and some nagging notions, reflected also in Robinson (1990) that we did not learn enough in STAR. But first, just a note on the Ancillary Studies.

### III. The Ancillary Studies

The mass of data collected for STAR and LBS surely could help in exploring additional questions of concern to educators. Not often does one have available a longitudinal database of over 10,000 pupils, many of whom were at one time randomly assigned, etc. At the outset of STAR the researchers selected a set of "comparison" schools from districts that had a school in STAR. Each comparison school was to be as similar as possible to the STAR school (demographics). There were no interventions; researchers only collected the test data for the parallel classes (K-3) for the STAR years (1985-86 to 1988-89).

As it moved ahead, for example from grade 1 to grade 2, the STAR cohort encountered those pupils who had been retained in grade 2 the previous year. Each year, all new students in STAR cohort sites (including retainees) were assigned to classes (S,R,RA) at random. Researchers also collected a file of data about each pupil, each teacher, each school, etc. In grade 4 (LBS) researchers collected data on student participation in school (Finn & Cox, 1992; Finn, 1993) from the available STAR pupils. The participation study became the first of the formal "ancillary" studies.

Given the data and the opportunity, researchers have recently initiated the series of Ancillary Studies. The following are very brief summaries of some Ancillary Studies to date.

- A. Participation. Students from (S) conditions were more actively involved (participating) in grade 4 than were students from (R) or (RA) conditions ( $p \leq .05$ ) (Finn & Cox, 1992). We have repeated this in grade 8 (Finn et al., in process) and hope to repeat the study again in high school.
- B. Retention in grade. Small classes do not help students who have been previously retained. Retainees do poorly (academically) in all class conditions, including the (S) condition. Initial placement in (S) seems to prevent or deter retention; (S) is not a treatment for remediation later (Harvey, 1993).
- C. "Gap reduction" between White and minority pupils. As in the case of retention, the (S) seems to prevent a large gap opening (K-3) between scores of White and minority pupils. The (S) condition helps minority pupils proportionately more than it helps White pupils. Once the test-score gap opens, (S) is not

an effective treatment or a remedy (Bingham, 1993). However, (S) placement seems to help keep this gap from opening.

D. Homogeneous or heterogeneous? Using the randomly-assigned STAR (R) and the comparison school (non-random assigned) students as comparisons, test-score achievement results favor the randomly assigned pupils (ANOVA, ANCOVA) increasingly from K to grade 3 (Zaharias, 1993).

E. Class size/school size. Numerous studies (e.g., Kounin & Gump, 1944; Fowler & Walberg, 1991) have shown that students in large schools get lower test scores and have lower participation rates than do students in small schools. Nye (in process) is exploring if small class placement tends to ameliorate the school-size effect. Initial results suggest that this is the case.

F. Other studies are planned in this series, including one about discipline, one on school effects, and an entire series on the teacher aide issue. A study of the "Test score value of kindergarten in later years: Grades 1, 2 and 3" has shown substantial (ES .40 to .50 or more) benefits to pupils, especially in (S) classes (Nye, Achilles, & Bain, 1994). Researchers hope to explore the range of questions inherent in "Is the (S) treatment a preventive or a remedial event?"

G. By Grade 7 pupil data were being entered into the state monitor system, so we're working on studies of behavior (discipline).

H. Tennessee has established an important test-data system, the Tennessee Value-Added Assessment System (TVAAS). We are comparing our LBS and Challenge results with this database. So far, all analyses are confirmatory.

I. A re-analysis of the teacher aide question is planned, as is a study of the effects of class-size reduction on teacher classroom behavior. Data for this were actually collected (pretest or pre-intervention data) in 1985-86, with post-test data collected in 1986-87 for Grade 1.

#### **IV. Questions and Answers Raised by Your Critique**

This background is important as a base for understanding the processes we followed to explore answers to the questions and issues that you raised. As I understand it, a general issue was that although STAR showed the benefits of (S) for pupil achievement, you were disappointed that STAR was a "mundane" study in that there were no steps to explain such things as why the (S) pupils may do better.

or to discuss what teachers may have done differently. Indeed, it seems inconsequential to advocate for (S) since other interventions (e.g., Success for All; Slavin et al., 1990; Madden et al., 1993) get considerably larger effect sizes (ES), and since there were no additional, probing analyses to try to explain why we got what we got. (We attribute it to class size -- period!) Indeed, on some of these points you are correct: we answered the question put before us and, with little ado, moved to the Derivative and Ancillary Studies. Some points that you raise cannot easily be resolved in a letter with a brief "yes" or "no"; they need a symposium-like atmosphere for discussion and they beg for lively interaction with the "on-line" STAR and LBS data. Only then, guided by both critics and supporters of the research might we get some resolution. Nevertheless, since you have raised the questions and issues, let me try to address them.

1. You point out that in STAR most gains were in K and 1, and that in later years (grades 2 and 3) pupils seem only to hold the gains from the early years. On the surface, this seems substantially correct, but our year-to-year analyses (cross-sectional) did show annual differences.

The "learned more each year" issue is not easily answered, and we continue to search through the data for clues. We do know that each year on the CRT (new objectives) the (S) students did better than those in (R) or (RA). We do know that the (S) pupils achieved what they did in reading and math in considerably less time per day than the other pupils. Evertson and Folger (1989) noted:

Teachers in the small classes devoted an average of an hour (64 min.) to reading instruction, while teachers in regular classes spent an hour and twenty-four minutes (84 min.). This might be expected considering that teachers in regular classes were instructing 1/3 more students. In fact, the time spent in the small classes reflects an increase of time per individual pupil of nearly a minute. (p. 7)

Since (S) pupils learned more in less time, we presume that they did not stop learning in the time saved. The math here is impressive: 20 min/day x 150 days = 3000 minutes or 50 hours/year. They learned more, we think. Tests in later years show (S) students ahead of (R) and (RA) students in all subjects tested (not just reading and math).

On the topic of additiveness, I believe that Mitchell et al. (1989) and I would agree (in theory) that there should be a cumulative effect (p. 38), so the lack of an increase in effect size in the original analyses

is disturbing. Avenues that we are exploring relate to retention in grade, special education placement and "passing" scores. For example, a smaller percentage of pupils is retained in the (S) classes than in (R) and the range of scores of promoted students is larger in (S) than in (R), 19 to 8. (See Table 1.) This means that as the (S) class moved along as a cohort it accumulated lower scoring pupils than did the (R) class. (Also, it had higher-scoring pupils due to the class-size impact . The ratio may be key.) As STAR cohorts moved through the grades (K-1-2-3) they would "pick up" pupils retained in grade the previous years. Again, the ratio is important. One major issue here, however, is a minor design flaw. As the years of the study went by, the line between (S) and (R) became blurred. (More on this later.)

Table 1. Range of scores for promotion/retention by class type for STAR, Kindergarten to Grade 1. (Scores on SESAT.)

	<u>S</u>	<u>R</u>	<u>RA</u>	<u>DIFFERENCES</u>		
				<u>S-R</u>	<u>S-RA</u>	<u>R-RA</u>
Promote	441	435	436	6	5	-1
Retain	422	427	421	-5	1	6
Difference	19	8	15	11	4	-7

Since the original analysis answered the original questions, we probably would not have revisited the STAR database if you (and others) had not raised some questions. Our responsibility to the State Board and legislature was to research the question in the legislation (which we did). The answer clearly was that students in (S) classes do statistically (and educationally) better than students in (R) and (RA) classes in all locations. When people began to ask about the size of the difference (new questions) we have returned to the data. It is very clear that the original reported results were quite conservative. However, discussion specifically about the 1:15 ratio spurred us to run for public consumption the frequency distributions of class sizes in the S,R,RA conditions over the four years (K,1,2,3) of the study. Table 2 provides the frequency distribution information. Several points are important here.

Table 2 about here

- a. A class was designated S,R,RA based upon the K distribution. If a (S) class grew "out of range," we still analyzed it as (S) as the pupils were in (S) in K. A class designated (R) was still treated as (R) even though it may have shrunk "out of range," a particular possibility with large levels of retention in grade, and a possible explanation for a diminution of differences between scores of (S) and (R) or (RA) in later grades. These analyses will mean that we'll lose some classes; but a class (n) of 50 or so should still be enough for this analysis.
- b. Most in-range (12-17) small classes have 16 or 17 pupils, toward the "large" end of (S), and larger than 1:15 (except grade 2), the preponderance of in-range (22-26) regular classes have 22-23 pupils and are toward the "small" end of the (R) range (except grade 3). Thus, many (S) classes in the

original analyses had more than 15 pupils and many (R) classes had fewer pupils than 24.

Classes were "out of range," and the debate still spoke of 1:15.

- c. The distribution of "out of range" classes (n=18-21) seems to favor (R) or (RA), but this remains to be tested. Even though we have "lost" some classes, early analyses show that the results are significant and show larger (S) benefits than found in the original analyses. Tables 3 and 4 show basic Pearson (r) for class size with average scores for each class size using the composite of (S) and (R) (Table 3) and just the RA classes (Table 4). Column A in Tables 3 and 4 shows the "legal" correlation using all classes ranging in size from 12-27; column B (Table 3 only) is based on in-range classes and suggests that removal of out-of-range classes may have some merit.
- d. Note that most of the "Most Effective" teachers based on gain scores in Grades 3 were in (S) (Table 5), and for this analysis, (S) was 13-17; some were above 1:15.

---

Tables 3,4, and 5 about here

---

2. Your second point is that you consider the finding of class size effects is, by itself, not very exciting -- indeed, "mundane" is your word. Study results do say that small classes by themselves lead to improved achievement and to get that achievement, teachers needn't do anything special. Finn's notion of "participation" may be important here. The work of Slavin and others (Slavin et al., 1990; Madden et al., 1993) shows that if teachers do other things, achievement goes up. Our research suggests that some of Slavin's et al. ES may well be due only to starting with a base of n=15 or so. In fact, probably one-third to one-half of the ES found in the Slavin et al. work may be due only to class size, and that this condition needs to be met prior to the benefits of other interventions.

We should not argue over your use of "mundane." If "mundane" serves your purposes, surely you should use that descriptor. We are, however, gratified to find among the poor research (or data) out there often called "education research," at least one fairly secure finding supported by a strong design, longitudinal analysis, etc. We probably could argue on this for awhile, but -- frankly -- we find most of the one-shot questionnaire stuff pretty useless; indeed, mundane. In fact, the OERI "Background

Document on Proposed Funding" for CSEARS notes: "Ideally, in order for scientific findings to be considered 'definitive,' they need to be executed with rigorous techniques. . . ; they need to be sufficiently robust to be generalizable beyond the study sample. . . . Because few social science studies meet these criteria. . . ." (p. 14). The rigorous STAR/LBS study at least meets this criterion of "definitive" -- even if it is "mundane." Again, we're content with the conservative results of STAR.

This brings us to the multivariate analysis (MANOVA) that we conducted (Finn & Bock, 1985) after we removed the out-of-range classes. (See Table 2.) The results of this analysis using classes of 13-15 as (S) and of 23-27 as (R) (pupils tested, not pupils assigned, so results will still be conservative) appear in Table 6. Although the results are statistically significant (usually beyond  $p \leq .01$ ), the key findings are probably reflected in the reported effect sizes. They range from .33 to .71 and generally show a trend of increasing, grades 1 to 3. This analysis shows a greater effect of (S) once the out-of-range classes are cleaned out of the analysis and now we have an ES consistently over .50 at nearly all grades. This compares favorably with some planned interventions to cause achievement gain -- a question totally different from the one addressed by STAR.

---

Table 6 about here

---

3. You ask why the STAR and LBS studies found continuing benefits for (S) for pupil achievement while prior studies had not found such continuing benefits. One reason is that most STAR pupils were in (S) for K-3, or at least for several years. Other studies were short term, even "one-shot."
4. Some of the "puzzling findings" that you note have also been found in other studies of early intervention, such as the Head Start evaluations and other studies where there is a "fade" in later grades (Weikart, 1989; Zigler, 1992; McMasters, 1991; and other sources). The initial STAR analysis seemed to find the same thing. One way to explain it is to consider the students who would be new to STAR each year (often retainees) and that in (S) teachers kept lower scoring students moving ahead while in (R) the lower scoring students were more likely to be retained (Table 1). Another approach is to re-analyze the

STAR data after removing the out-of-range classes and considering the proportion (%) of (S) and (R) classes in the top and in the bottom quartiles of classes based upon test scores.

Table 7 shows what percent of (S) and (R) classes are in the top 25%, middle 50% and bottom 25% of classes based on reading test scores. Notice that consistently the STAR (S) classes are over-represented (from 8% to 11%) in the top-scoring classes and under-represented in the bottom, while the reverse is true for (R) and that there isn't much variation in the middle 50% group.

Table 8 examines the question after removing the out-of-range classes and comparing the percent of (S) and (R) classes in the top 25% and bottom 25% as a proportion of the percent of (S) and of (R) classes at each grade. The (S) classes are 39% of all K classes, but 51% of the top 25%, or a difference of +12% from what might be expected. The constant positive differences for (S) in the top (from +11% to +14%) and constant minus differences for (S) in the bottom (-4 to -14) and the opposite for the (R) classes show the consistent and generally increasing benefits of the (S) condition even with the retainee and low-scoring student phenomena mentioned earlier. This, of course, raises a new set of questions for exploration, such as removal of retainees prior to the analyses, etc.

---

Tables 7 and 8 about here

---

So, we trust that this helps clarify some of your questions. If you (and others) had not raised questions -- often about the power of the differences that we initially found -- we probably would not have revisited the analysis. In fact, the benefit of your questions (and, we'll admit, we thought that you were stretching things a bit to find fault!) is that we have decided to do what we should have considered before -- some secondary analyses to answer questions that we should have anticipated. A major database, collected carefully and at great expense, should help with nagging questions. We hope that this letter will give you fodder for future questions. Next time, please contact us directly so that we can all benefit from the question asking and problem finding, not just from the problem solving.

Today's children (Hamburg, 1992; Hodgkinson, 1991 and 1992) are different from yesterday's, and their needs are far different. A small-class start in school seems to be an outstanding option for

preventing future problems (e.g., Finn, 1993; Bingham, 1993; Harvey, 1993; etc.). Tutoring is a one-to-one option, the ultimate in class-size reduction; it is quite expensive compared to 1:15. Yet, given the condition of young children entering schools today (e.g., Hodgkinson, 1992) we know of few options that provide the positive results of class-size reductions, and we know of few "programs" successful in raising scores of pupils that do not start with small groups or classes.

We can speculate from STAR/LBS results that the use of (S) in early primary grades has the potential not of being expensive, but of actually offering a path to cost savings:

- a) By reducing the need for grade retention (an expensive and non-productive form of witchcraft in the field).
- b) By reducing the pesky gap in achievement between the poor and/or minority pupil and the mainstream pupils who have traditionally benefited most from formal schooling.
- c) By establishing work conditions that support the idea of mainstreaming and that help teachers to early identification of pupil learning difficulties, thus saving on special-education costs.

Additionally, we should not exclude the potential for teacher satisfactions gained by successes achieved through working with a manageable group size in one classroom. We shall continue to inform you of our work as it progresses.

Sincerely,

### Author Notes

C.M. Achilles is Professor, Educational Leadership at Eastern Michigan University, Ypsilanti, MI 48197. Barbara A. Nye is Director, Center of Excellence for Research in the Basic Skills, Tennessee State University, Nashville, TN 37203-3401 (615-963-7238). Dr. J. Zaharias, D. Fulton, and V. Cain are researchers at the Center. Among other things the Center is managing the Student-Teacher Achievement Ratio (STAR) database, conducting the Lasting Benefits Study (LBS) which is still following a major cohort of STAR students who are now in grade 8, providing some evaluation support for Project Challenge, which is a general application in 17 poor Tennessee counties of small classes (1:15) in the line with STAR findings.

C.M. Achilles was one of the original principal investigators on STAR (H.P. Bain, J. Folger, J. Johnston, F. Bellott, E. Word were the others). He continues to re-analyze the STAR database and to try to address new questions through the STAR, LBS and Challenge connections.

The authors wish to thank all of the persons who have reviewed prior papers that provided the base for the present one. Dr. Jeremy Finn assisted with analysis; Drs. Richard Hooper, G. Bobett and H. Bain assisted with analysis and interpretation. Special thanks to the various skeptics who have been combined into the mythic, composite Dr. Snipe, to whom this letter is addressed.

## References/Bibliography

- Achilles, C.M., Kiser-Kling, K., Owen, J., & Aust, A. (1994). *Success Starts Small. Final Report. A Small-Grant School-Based Research Project.* Chapel Hill, NC. The Board of Governors of the University of North Carolina System.
- Achilles, C.M., Nye, B.A., Zaharias, J.B., & Fulton, D. (1993, November). Creating successful schools for all children: A proven step. *Journal of School Leadership*, 3(6), 606-621.
- Bingham, S.\* (1993, December). An examination of small class as a "gap reduction" strategy for achievement differences in groups of students, K-3. Unpublished Ed.D. study. University of North Carolina, Greensboro.
- Cahen, L.S., Filby, N., McCutcheon, G., & Kyle, D.W. (1983). *Class size and instruction.* New York: Longman.
- Campbell, D.T., & Stanley, J.C. (1986). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.
- Education Research Service or ERS. (1978). *Class size: A summary of research.* Arlington, VA: Author.
- Education Research Service or ERS. (1980). *Class size research: A critique of recent meta-analysis.* Arlington, VA: Author.
- Everston, C.m., & Folger, J.K. (1989). Small class, large class: What do teachers do differently? Paper at American Educational Research Association, March, 1989, San Francisco.
- Finn, J.D. (1993, August). *School engagement and students at risk.* Washington, DC: National Center for Educational Statistics (93-470).
- Finn, J.D., & Achilles, C.M. (1990, Fall). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27(3), 557-577.
- Finn, J.D., & Bock, R.D. (1985). *Multivariate VII user's guide.* Chicago, IL: Scientific Software, Inc.
- Finn, J.D., & Cox, D. (1992, Spring). Participation and withdrawal among fourth-grade pupils. *American Educational Research Journal*, 29(1), 141-162.
- Glass, G.V., & Smith, M.L. (1978). *Meta-analysis of research on the relationship of class size and achievement.* San Francisco: Far West Laboratory for Educational Research and Development.
- Glass, G.V., Cahen, L.S., Smith, M.L., & Filby, N.N. (1982). *School class size. Research and policy.* Beverly Hills: Sage Publications.
- Hamburg, D.A. (1992). *Today's children.* New York: Time Books, Random house.
- Harvey, B.\* (1993, December). An analysis of grade retention in pupils in K-3. Unpublished Ed.D. Study, University of North Carolina at Greensboro.

---

\* These studies are extending the STAR, LBS and Challenge data as a way to try to understand the impact of either a small class or a full-time teacher aide on the conditions being studied. They are a part of a series of dissertations and studies.

- Hibbs, B.F.\* (In Process). Relationships among discipline factors and early student placement in small (1:15), regular (1:25) and regular-with-aide classes.
- Hodgkinson, H. (1991, Sept.). Reform vs reality. Phi Delta Kappan, 73(1), 8-16.
- Hodgkinson, H. (1992, June). A demographic look at tomorrow. Washington, DC: Institute for Educational Leadership.
- Madden, N.A., Slavin, R.E., Karweit, N.L., Dolan, L.J., & Wasik, B.A. (1993, Spring). Success for all: Longitudinal effects of a restructuring program for inner-city elementary schools. American Educational Research Journal, 30(1), 123-148.
- McMasters, L. (1991). What is the importance of a prekindergarten program to achievement and school success for students in rural Chatham County, NC. Unpublished Ed.D. Dissertation, University of North Carolina at Greensboro.
- Mitchell, D.E., Beach, S.A., & Badarak, G. (1989, Fall). Modeling the relationship between achievement and class size: A re-analysis of the Tennessee Project STAR data. Peabody Journal of Education, 67(1), 3 4-74. (Published in 1992).
- Nye, B.A., Achilles, C.M., & Bain, H.P. (1994-95). The test-score "value" of kindergarten for pupils from three class conditions, at grades 1, 2 and 3. National Forum of Educational Administration and Supervision Journal, 12(1), 3-15.
- Nye, K.\* (1993, In Process). Do small classes help ameliorate the generally negative effects of large schools on pupil achievement, K-3?
- Robinson, G.E. (1990, My). Synthesis of research on the effects of class size. Educational Leadership, 47(7), 80-90.
- Sarason, S. (1993). Letters to a SERIOUS education president. Newbury Park, CA: Corwin Press.
- Slavin, R.E., Madden, N.A., Karweit, N.J., Livermon, B.J., & Dolan, L. (1990, Summer). Success for all: First-year outcomes of a comprehensive plan for reforming urban education. American Educational Research Journal, 27(2), 255-278.
- Tomlinson, T.M. (1988). Class size and public policy: Politics and Panaceas. Washington, DC: US Department of Education, Office of Educational Research and Improvement.
- Tomlinson, T.M. (1990, Fall). Class size and public policy: The plot thickens. Contemporary Education, LXII(1), 17-23.
- Weikart, D.P. (1989, June). Quality preschool programs: A long-term social investment. Occasional paper Number 5. Ford Foundation Project on Social Welfare and the American Future. New York: The Ford Foundation. (28 pp.).
- Word, E., Johnston, J., Bain, H., Fulton, B., Zaharias, J., Lintz, N., Achilles, C.M., Folger, J., & Breda, C. (1990). Student/teacher achievement ratio (STAR): Tennessee's K-3 class size study. Final report and final report summary. Nashville, TN: Tennessee State Department of Education.
- Zaharias, J.B.\* (1993, December). The effects of random class assignment on elementary students' reading and mathematics achievement. Unpublished Ed.D. dissertation, Tennessee State University, Nashville.
- Zigler, E. (1992, June 27). Head Start falls behind. New York Times, p. 15.

Table 2. Distribution of STAR classes by grade (K-3) by designation S (Small), R (Regular), and RA (Regular and Aide).

	<u>K (n classes)</u>			<u>1 (n classes)</u>			<u>2 (n classes)</u>			<u>3 (n classes)</u>		
	S	R	RA	S	R	RA	S	R	RA	S	R	RA
11										2		
12	8			2			3			2		
13	19			14			16			15		
A 14	22			18			27			17		
15	23		1	31			32			31		
16	31	1		16	1		29	1		31		1
17	24	4	1	33	1		19			27		
18		1	2	6	2		6			10	1	
B 19		7	6	3	4	3	1	3	3	5		4
20		6	6	1	10	6		2	1		9	13
21		14	12		18	18		7	11		11	12
22		20	20		27	15		23	21		13	16
23		16	21		19	20		20	21		10	14
24		19	14		16	11		22	25		15	14
25		6	6		7	9		9	15		16	15
C 26		4	3		5	9		6	7		5	12
27		1	6		2	4		4	1		5	8
28			1		1	2		1	0		2	6
29					1	2		2	2		2	2
30					1	1						
TOT	127	99	99	124	115	100	133	100	107	140	90	107
	325			339			340			337		

A = range for (S); B = "out of range"; C = range for both (R) and (RA) classes.

Table 3. Correlations of class size with reading and math (SAT) scores, K-3, Project STAR (1986-1989) using (S) and (R) classes only (A=all S and R) (B is S=13-15; R=23-26).

	Grades							
	K		1		2		3	
	A	B	A	B	A	B	A	B
Reading	-.19**	-.28**	-.27**	-.24**	-.23**	-.35**	-.23**	-.30**
Math	-.14*	-.20*	-.26**	-.23**	-.18**	-.30**	-.18**	-.26**

Sig: \* = .05, \*\* = .01

Table 4. Correlations of class size with reading and math (SAT) scores, K-3, Project STAR (1986-1989) using (RA) classes only (A=all RA) (B is 23=27 pupils in RA).

	Grades			
	K	1	2	3
	A	A	A	A
Reading	-.11	-.05	.13	.08
Math	-.04	.03	.08	.04

Sig: \* = .05, \*\* = .01

Table 5. Interviews of 50 "Most Effective" STAR Teachers (Based on Gain Scores) (Grade 3).

Of those 50:	<u>N</u>	<u>Class Size</u>	<u>%</u>
	8	22-25* (R)	16
	23	13-17 (S)	46
	7	18-21	14
	12	full-time aides (RA)	24
	<u>50</u>		<u>100%</u>

\*22-25 pupils is probably smaller than many regular classes nationwide.

Table 6. Means, Standard Deviations, etc. for ANOVAs, grades K-3 for Small (n=13-15) and Regular (n=23-27) classes, STAR, showing significance and effect sizes (ES).

	GRADES															
	K			1			2			3						
	READ	MATH		READ	MATH		READ	MATH		READ	MATH					
	X	SD	n=64/46	X	SD	n=63/49	X	SD	n=75/61	X	SD	n=63/51				
S=13-15	442.02	19.10	491.76	30.68	528.58	32.97	538.14	27.50	592.80	25.60	588.82	21.13	623.80	19.45	626.38	21.62
R=23-27	432.06	17.54	481.64	30.53	513.64	31.35	525.79	25.01	576.72	22.51	575.19	21.74	611.24	18.70	613.95	22.35
Diff	9.96		10.12		14.94		12.35		16.08		13.63		12.56		12.43	
Total	437.86	19.03	487.53	30.89	522.04	32.98	532.74	27.04	585.65	25.48	582.76	25.71	618.30	20.04	620.89	22.71
Effect Size <sup>1</sup>	.52		.33		.45		.46		.63		.53		.63		.55	
Effect Size <sup>2</sup>	.57		.33		.48		.49		.71		.63		.67		.56	
Significance	.01		.09		.02		.02		.000		.002		.000		.003	

<sup>1</sup>ES using SD of pooled variance for total; <sup>2</sup>ES using SD of control only.

NOTE: Classes include all assigned students regardless of years of STAR participation.

Table 7

Distribution (%) of Small and Regular Classes into Top (25%), Middle (50%) and Bottom (25%) of Class  
Average Scores (Total Reading) Unadjusted for "Out of Range"

---

	K		1		2		3	
	S	R	S	R	S	R	S	R
Bottom 25%								
Actual	21	28	16	33	21	31	18	31
Diff	-4	+3	-9	+8	-4	+6	-7	+6
Middle 50%								
Actual	46	49	49	50	46	51	46	45
Diff	-4	-1	-1	0	-4	+1	-4	-5
Top 25%								
Actual	34	23	35	18	33	18	36	24
Diff	+9	-2	+10	-7	+8	-7	+11	-1

---

Table 8

Percent of High (Top 25%) and Low (Bottom 25%) Scoring Classes (Reading) that are Small (S) or Regular (R) by Grade

	K		1		2		3									
	S	R	S	R	S	R	S	R								
	A	Dif	A	Dif	A	Dif	A	Dif								
Bottom 25%	35	-4	35	+4	23	-14	45	+11	33	-6	37	+7	30	-12	33	+7
Top 25%	51	+12	27	-4	49	+12	23	-11	50	+11	21	-9	56	+14	24	-2

A = Actual Percent

Dif = Difference (+,-) from Expected distribution (see "Note").

Note: S classes are 39% of classes; R are 31% of the classes (K).

S classes are 37% of classes; R are 34% of the classes (grade 1).

S classes are 39% of classes; R are 30% of the classes (grade 2).

S classes are 42% of classes; R are 26% of the classes (grade 3).

APPENDIX A

PRIMARY AND EXTENDED ANALYSES DESIGNS:  
STAR (1985-1989); LBS 1990-1991.

Sample Design:

- 4 Locations (urban, rural, etc) (fixed effect)
- Schools nested in locations (random effect)
- Class types (S,R,RA) crossed with locations and school types (fixed effect)
- 2 training categories\* (fixed)

Source Table

Error Term:

- Location (L)
- Schools
- Training\* (TR)
- Type (T)
- School x type
- L x T
- School x type
- T x TR
- School x type
- L x T x TR
- School x type

618



Degrees of Freedom (df)

	Ach. Meas.	Noncog. Meas.	
Schools	75	69	
School x Type	149	137	
Classes within School-Types (etc.)			Etc.

Primary Model; Measures  
Achievement (Ach.);  
Noncognitive (Noncog.);

SESAT, SAT, BSF  
SCAMIN, atten-  
dance, behav-  
ior,  
etc.

Matched  
t-tests

Extended Model;

Measures:

Sex (or race, or SES)  
Sex (or race, or SES)  
Training\*

Ave. diff. scores  
on Ach.  
Ave. diff. scores  
on Noncog.

Multi-  
variate  
models

Two planned contrasts S class mean vs. means of all R and RA, S vs. (R + RA + 2) - RA class mean vs. R class mean  
Each effect tested holding constant earlier effects in order of elimination TR and T each tested as last main effect, L x TR and L x T each tested as last two way interaction  
Analysis of BSF done with "long odds index"

Appendix A and Appendix B provide basic STAR data. They appear in various published articles and conference papers by the same authors

Tables 2 and 3 a random subset of schools was chosen to study the effects of any of the other training (TR) on pupil outcomes. Although not discussed in detail here, the training used had no significant effect

51 of School Leadership 3(6), Nov. 1993

11 617-619

APPENDIX B

ANALYSIS OF VARIANCE FOR COGNITIVE OUTCOMES, STAR,  
GRADES K-3 (SIG. LEVELS  $p \leq .05$  OR GREATER ARE TABLED).

Effect/ Grade <sup>a</sup>	Location (L)	Reading			Mathematics		
		Multi- variate <sup>b</sup>	SAT <sup>c</sup> Read	BSF Read	Multi- variate <sup>b</sup>	SAT Math	BSF Math
			.02			.05	
	1	.01	.06		.05		
	2	.001	.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		.001			.02	
	1	.001	.001	.001	.001	.001	.05
	2	.001	.001	.001	.001	.001	.05
	3	.001	.001	.001	.001	.001	.001
SES	K		.001			.02	
Loc x Race	1		.05		.05		
Loc x Type	K-3						

NOTE: Only statistically significant ( $\leq .05$ ) results are shown.

Appendix A and Appendix B provide basic STAR data. They appear in various published materials and conference reports by the same authors.

The nonorthogonal 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, 1990).

\*Obtained from F-approximation from Wilks' likelihood ratio. Essentially, no statistically significant differences were obtained on the self-concept and/or motivation (SCAMIN) measures. No training main effect, or training-by-type interaction. Trained and untrained teachers did equally well across all class types and the (S) advantage (and absence of A) effect is found equally in all four locations for trained and untrained teachers.

(S) advantage and all effects found for total class generally apply equally to white and nonwhite pupils. (S) significantly better than (R,RA) on all tests; no R vs. RA tests significant.

Appendix C

Rankings of Challenge Districts (n=17) of 138 TN School Systems Based on Grade 2 TCAP Scores (Reading and Math). (State TN ave. rank is 69.)

	READING		MATHEMATICS	
	<u>89-90</u>	<u>90-91</u>	<u>89-90</u>	<u>90-91</u>
Sum of Ranks	1681	1591	1448	1336
÷ 17	98.9	93.6	85.2	78.6
Difference	(+90)	(+114)	(+112)	(+325)
÷ 17	+5.3 RK	+6.7 RK	+6.6 RK	+19.1 RK
90-92	+204, or 12 RK (Ave)		+437, or 25.7 RK (Ave)	

Average system (1990-1992) moved up 33.1 Ranks (Reading) and 22.4 Ranks (Math).

NOTE: Challenge systems began well below the state average. After 3 years of treatment (to involve pupils who had 1:15 in K, 1 and 2) Challenge systems are some above the state average in Math and increasing each year in reading.

Table 1

Samples of Studies Derived From and Building Upon the STAR Initiative Classed as "Subsidiary" (directly from STAR), "Ancillary" (building on and using STAR database) and "Related" (triggered by STAR results and usually involving STAR researchers) Studies.

<u>CATEGORY, TITLE &amp; PURPOSE</u>	<u>DATE(S)</u>	<u>AUTHOR(S) OR PUBLICATION</u>
<u>Subsidiary Studies</u>		
•Lasting Benefits Study to follow STAR pupils	1989-Present	Nye et al., 1994, 1993, 1992, 1991
•Project Challenge 1:15 in 17 poor counties	1989-Present	Nye et al., 1994, 1993, 1992, 1991
<u>Ancillary Studies</u> (Use or extend STAR data. Some of these are dissertations.)		
•Retention in Grade	1994	Harvey
•Achievement Gap	1994	Bingham
•Participation in School Grades 5, 7	1990, 1994	Finn et al., 1989
•Value of K in Classes of Varying Sizes (test scores)	1985-89	Achilles et al., 1994
•School-Size and Class Size Issues	1985-89	Nye, in process
•Random v. Non-Random Pupil Assignment and Achievement	1985-89	Zaharias, 1993
•Class Size and Discipline in Grades 3, 5, 7	1988, 1990, 1992	In Process
•Outstanding Teacher Analyses (top 10% of STAR teachers)	1985-89	Bain, Bain et al.
<u>Related Studies</u>		
•Success Starts Small: Grade 1 in Chapter 1 (1:15, 1:25) Schools	1993-94	Achilles, et al. (1994)
•Burke County (NC) Schools and Their 1:15 Experiment	1991-94	SERVE, Achilles, et al.

\*Note: This list is not complete. It provides samples of the types of studies. Not all "authors" appear in the references in the exact way that they are listed here.

25

Draft