DOCUMENT RESUME

ED 406 407 TM 026 294

AUTHOR	Achilles, C. M.; And Others
TITLE	"This," as Paul Harvey Was Wont To Say, "Is the Rest of the Story."
PUB DATE	Nov 96
NOTE	20p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Tuscaloosa, AL, November 1996).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Class Size; Databases; Decision Making; Educational Research; Elementary Secondary Education; Research Design; Research Methodology; Research Problems; Research Utilization; Scores; State Legislation; *Test Results; Test Use
IDENTIFIERS	*Project STAR; Tennessee

ABSTRACT

A project conducted in Tennessee from 1984 through 1989, Student Teacher Achievement Ratio (Project STAR), serves as a context for a discussion of educational research. The decisions required in major research projects and the problems in carrying out research are seldom discussed in conferences that present research results as completed efforts. Project STAR illustrates the long-term consequences of early decisions and implementation and the additional value research may have. The research question was established through state legislation, but the researchers had to operationalize or limit key variables. The study was conducted to explore the influence of class size on student achievement. How to measure achievement and how to analyze the test results that were chosen as measures of achievement became questions of importance as the project evolved. Relying solely on student outcomes and certain behavioral indicators was not considered adequate for the study, and researchers began to gather other information about schools, students, and teachers. About 100 classes of each type (small, regular, and regular with an aide) were used each year. The size of the database eventually developed, the care researchers had taken, and the in-school design allowed STAR information to be used in subsidiary and ancillary studies. As each inquiry moved further from the initial research question, however, the power of the study and confidence in the results diminished. (Contains 5 tables and 28 references.) (SLD)

"THIS," as Paul Harvey Was Wont to Say,

"IS THE REST of THE STORY"

Paper Presented at the Mid-South

Educational Research Association (MSERA)

11/96. Tuscaloosa, AL

C. M. Achilles, Professor, Eastern Michigan University:
Principal Investigator, STAR and
Consultant to STAR-Related Studies

H. P. Bain, Retired.
Principal Investigator, STAR

B. D. Fulton, Research Specialist, STAR Data Manager
TN State University (TSU) Center for Research and Policy (RPC)
330 10th Avenue, North
Nashville, TN 37203-3401

The research reported herein was partially supported by the Center for Research and Policy (RPC), Tennessee State University, Nashville, TN. Dr. B. A. Nye is director of the Center.

Training Session: Research

"THE REST OF THE STORY"

C. M. Achilles, Eastern Michigan University

B. D. Fulton, Tennessee State University

H. P. Bain, (Retired)

At conferences, research is presented as a "done deed," with attention to design, method, and findings. Yet, major research has decisions and problems that seldom get discussed in public.

Purposes are: (1) To review key decisions and issues in designing, conducting, and reporting on research that has had national implications. Early decisions had long-term implications. The research questions were established in legislation, but the researchers had to operationalize or limit key variables; (2) To discuss serendipitous outcomes of the research processes and results, and (3) To discuss transference of these ideas to other research.

The format is a conversation, with project researchers as presenters and discussants, and audience as interlocutors and critics. Discussion will emphasize points such as: power, sample, "goodness of fit," primary analysis, secondary analyses, ethical questions of the research and results, and dilemmas of releasing new results. Researchers will address continuing use of a database designed for one purpose, but which is uniquely suited for answering new questions that add knowledge about student achievement and school improvement. As each inquiry moved further from the initial research question, the power of the study and confidence in the results diminished.

"THIS," As Paul Harvey Was Wont to Say: "IS THE REST of the STORY"![1]
(A Retrospective Narrative)

Preface

Education research seldom gets "rave reviews" for its depth and rigor. Some commentaries, such as Kaestle's (1993) "The Awful Reputation of Educational Research" are pointed; professional journals decry the absence of educational research in policy discussions or in classroom practice. Critiques such as those by Donmoyer (1993), Wagner (1993), and Achilles (1990) exemplify the many criticisms and illuminate weaknesses in both the conduct of educational research and the use of results. Thus, positive critiques of one study that adds knowledge to help policy maker and practitioner alike in improving schools and schooling are good news. Noted Harvard professor of mathematical statistics (emeritus), Frederick Mosteller (1995), reviewed Project STAR (Student Teacher Achievement Ratio) conducted 1984 through 1989 in Tennessee and said:

> This article briefly summarizes the Tennessee class size project, a controlled experiment which is one of the most important investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools (p. 113).

Orlich (1992) recognized STAR's value as a base for school improvement especially suited to equity. Orlich said:

> The study lasted for four years and, in my opinion, is the most significant educational research done in the US during the past 25 years (p. 632).

Given the general low estate of education research's reputation and the fairly positive critiques of Project STAR, it seems appropriate to discuss STAR's background as a path for understanding how early research decisions influence many things, including the questions that can be answered with the data, and the uses of study results. Critiques of STAR are quite positive, but there is, as commentator Paul Harvey says, "The Rest of the Story."

Introduction and Purpose

People usually see only research results presented as well-polished publications. Seldom do researchers report post hoc on details and problems that influenced the development and design of a research project. To the uninitiated, it is as if the research sprang pure and in full-bloom, like Athena from the head of Zeus. An old philosopher noted that any shingle, no matter how thin, has two sides. So, the purpose of this paper is to provide a retrospective narrative of the general operation of a major research undertaking. Project STAR, a study of class size and student outcomes began in 1984 and continues (1996) in subsidiary and ancillary studies and in re-analyses of original data. Longitudinal experimental studies of this magnitude are not common in education (or in other fields). STAR began with a legislative mandate in 1984-85 known as House Bill 516. The study was to answer the following general question:

---

[1] C.M. Achilles, Professor, Educational Leadership, College of Education, Eastern Michigan University, Ypsilanti, MI, 48197 was one Principal Investigator (PI) for the studies used as the basis for this narrative. This paper draws on his memory of events and upon archival and published documents from the research. A special thanks to all STAR PIs, participants, consultants, and others associated with STAR. Any errors are the author's. Proximity or familiarity may breed contempt, but who knows better?

What is the effect of reduced class size on student achievement and development in early primary grades?

Many teachers felt that the study was unnecessary; as practitioners they knew that they could teach better an that children learned better in reasonable-sized classes. The study was conducted, however, to answer definitively the question of class-size effects on learning. If the findings were favorable to small classes, educators would have data to use in convincing policy people that small classes, especially in K-3, are educationally and practically important.

The research team included principal investigators (PI) from four of Tennessee's major universities: The University of Memphis, Tennessee State University, Vanderbilt University, and The University of Tennessee, Knoxville. The PI team was supported by advisory groups of a) practicing educators and b) research consultants. Also, there were provisions for securing part-time help as needed and for employing a person who would have primary responsibility for research design and for final data analysis. A full-time project director was appointed from the Tennessee State Department of Education (SDE); staff (n=2) were hired.

The Challenge

The research team was careful conducting this study for several reasons other than the normal care involved in a large study. When STAR began, there were debates on this topic, including work by Glass & Smith (1978), and the Education Research Service or ERS (1978, 1980). The Tennessee legislature sought a definitive answer to the question of the impact of class size on student learning and given differences of research and opinion, the legislature needed solid data upon which to make statewide class-size decisions for TN.

Initial Considerations and Guiding Principles

With little lead time, the PI team made initial decisions to assure as much flexibility as possible later. Two principles guided all decisions:

1. No student would receive less by being in a Project STAR school than if STAR were not conducted in that school.

2. There would be no adjustments to any portion of schooling other than the manipulation of class size. (This was an experimental study of class-size effects.)

The first condition was important, for example, because no student should be in a class that would exceed the state-mandated class size because of STAR. (No student was placed in this condition.) In fact, "regular" classes in STAR schools were smaller than the state average. The state could seek definitive answers about class size as long as it did not violate its own class-size rules.

The second condition was important because the experimental study was of the effects of class size, so class size should be the only variable manipulated. The addition of a full-time instructional aide to a regular-size class was also a variable, but it established the second experimental condition in STAR.

The full-time teacher aide condition was included as a financial "hedge." If the aide condition proved equal to or better than the small class, it would provide a less costly alternative than hiring a second teacher.

Operationalizing the Study

Next, researchers had to "operationalize" the general research question for the study. The PI team developed the following decision rules:

1. Early primary grades included grades K-3. STAR began with students in K. In TN in 1984, K was not required, so new students would enter STAR in grade 1 who had not been in the experiment in K.
2. Student achievement was scores on the regular state tests; there would be a minimum of extra testing. State personnel were developing curriculum-driven criterion-referenced tests (CRT). As they became available, the Basic Skills First (BSF) tests keyed to the state-mandated objectives became part of the STAR outcome measure for student achievement. The Norm-Referenced Test (NRT) measures for STAR were the appropriate Stanford Achievement Tests (SAT) for the students' grades. Knowing the problems of testing very young children (K-1) and of competing test limitations (ceiling and floor effects), researchers considered (a) the state curriculum and (b) the need to get as many students as possible "on board" for a baseline measure and chose the SESAT I over the SESAT II while recognizing that the ceiling effect could understate K-1 gains.
3. Development was taken to mean behavioral elements that could be captured early and later in the study: attendance, discipline, grade retention, etc. These measures were as objective as possible, and available for longitudinal review. The PI team worked with external researchers to re-validate a measure of self-concept called the SCAMIN (Self Concept and Achievement Motivation) useful in K-3. Studies reporting the reliability and validity of SCAMIN as used in STAR are available. (Davis & Johnson, 1987; Davis, Sellars & Johnson, 1988).

Evolution of the Design

Given the preceding decisions, a major challenge now became designing the study, keeping in mind both rigor and parsimony. "Effect" demanded a carefully controlled experiment. At a minimum, this would require randomization accompanied by careful checks to see that any elements accepted into the study were not unlike those elements throughout the state ("goodness of fit"). At the very basic level, districts had to have school board approval to participate, and principals had to agree that their schools would be in STAR. An Attorney General's ruling supported the state's right to do the experimentation necessary to establish class size, since class-size decisions were within the purview of legislative responsibility.

Each superintendent of schools received a letter inviting participation. Each district volunteered by the deadline was placed into the applicant pool. While this initial screening process was going on, the PI team determined other design elements that would influence the final selection of schools to be in the study. Each decision should maximize the potential for obtaining results that would withstand criticisms of design weaknesses. The researchers benefited from many reviews, critiques and discussions of prior class-size research.

One immediate decision had to be the class-size conditions. The legislation suggested a 1:15 teacher-pupil ratio derived from other research (e.g., Glass & Smith, 1978). The PI team chose 1:15 as the average for "small classes" (S), with a range of 13-17. The regular classes (R) had to have enough pupils to differentiate them from (S), but could not exceed state class-size regulations. The average size for (R) classes was 1:24 (range 22-26). In the real world of pupil mobility, class-size ranges were important to allow flexibility.

Besides class size, the legislation contained one other question. Would an (R) class with a full-time aide (RA) perform as well as, better than, or worse than, the (S) or the (R) class on the variables of interest? The teacher-aide question arose from PrimeTime, a statewide class-size project initiated in Indiana in 1982. (Mueller, Chase, & Waldon, 1988). PrimeTime was a project and STAR was an experiment. Although PrimeTime was evaluated, its results on the efficacy of small classes were mixed. "Clearly, the PrimeTime program affected 1st grade more than 2nd grade, and reading more than mathematics"(p. 50).

Thus, STAR really had two "treatments" and one control condition. The treatments were (S) and (RA), and the control condition was the (R) class of approximately 25 youngsters. Although considered a study of class-size effects, STAR could as legitimately be an experimental study of the effects of a teacher and teacher aide working in the same classroom.

After considerable discussion, the PI team decided on an in-school design. Any school with (S), also would have both other class conditions, (R) and (RA). Designating the same school as the experimental and control site helped control for such things as building-level variables and district-specific items (leadership, curriculum, texts, expenditures, scheduling, social class, etc.). The in-school design was also an attempt to obviate such concerns as the "Hawthorne Effect", and other things that may influence a study when the experimental condition is in one school and the control condition is in another. The design reduced the drop-out problem of control schools that gain little by remaining active in a study. The in-school design was parsimonious. A visit to one school allowed PI's to monitor all conditions equally. This design reduced the need for collecting additional district and building-level descriptive data.

The in-school design decision did, however, influence school selection. To be in the study, a school had to have enough youngsters to accommodate all three class types (S, R, RA). Any school with fewer than 57 youngsters in K (the minimum class sizes were 13, 22, and 22) was systematically excluded. Small schools were not in STAR, but perhaps this was positive given findings of school size and achievement (e.g., Nye, 1996: Fowler & Walberg, 1991).

Random Selection

Randomization (each student was randomly assigned to one of three class conditions, then teachers were randomly assigned to classes and classrooms) meant that a post test only design would be appropriate. This avoided the messy issue of a pretest of K youngsters; youngsters would have a year in school before taking the required tests. Knowing that there would be an influx of pupils in grade 1, researchers established processes for random replacement and random establishment of new classes in participating schools.

Variable Groupings and Unit of Analysis

Researchers grouped the variables into cognitive (test results) and non-cognitive (e.g., attendance, discipline, and self-concept). The unit of analysis was the class average as this was a study of class-size effects. Researchers believed that each student was not an independent measure due to teacher and peer influence. This decision reduced dramatically the degrees of freedom and the probability that a small difference would be statistically significant.

A Stumbling Block

The classroom or class average as a unit of analysis provided a future stumbling block. Once a class was designated as (S), it had to remain designated as (S) throughout the study. A class that started out as (S) could, due to mobility and district growth, take on enough students to move outside of the small-class range. Also, a class at the small end of the (R) or (RA) could, conceivably, lose enough students to become (S) or classes might drift into the out-of-range area which would be classes of 18-21 -- class sizes not established in the study. Given the research question, the PI team agreed on a conservative analysis and a conservative design. (Later analyses are employing other methods such as Hierarchial Linear Modeling or HLM).

The Testing Issue

Young children are not well experienced in test taking, and test scores can fluctuate if testing conditions are not the same for all. To assure equivalent test conditions, the PIs hired monitors to be in all schools during testing. Testing occurred in groups of approximately 10 youngsters under the direction of the teacher, with a monitor present. Testing occurred within a five-day period in the same week at all sites. Test data were processed by the State Testing Bureau in the same manner that all other state test data were handled.

The Cohort Issue

Researchers established class types in 1985-1986 and essentially once a class was designated S, R, or RA, it remained that type until 1988-1989 when the pupils exited grade 3. Pupils moved as a cohort through K-3 (or 1-3 if a pupil entered in grade 1) so a "cohort effect" is possible, but the in-school and random design meant that all class types had the cohort experience equally.

The Data Analysis Decisions

The PI team contracted out the primary data analysis to assure neutrality. Team members conducted other analyses, but these were considered confirmatory or exploratory while the external analysis was considered the primary and "official" analysis. Differences between primary and subsidiary analyses were discussed in the final report, but the primary analysis is the official documentation of STAR.

Other Data Questions

Analyzing only student outcomes (test results) and certain behavioral indicators would not seem adequate for a study of the magnitude and duration of STAR, and which cost in excess of 13 million dollars during its four years. Researchers, therefore, determined other data needs and established protocols and questionnaires for obtaining these data: demographics of pupils, teachers, and administrators; information about schools and school districts; interviews and questionnaires; logs of time usage, etc. Teachers were asked about grouping practices, use of volunteers, student participation, etc. All teachers and some aides were interviewed at the end of each year. Demographics included such things as the pupil's date of birth, race, sex, free and reduced lunch, special education placement. Teacher/aide/administrator information also included experience and training.

Data Cleaning

If a student were not placed in a class by a certain date (November 1 of each school year), that student's test scores were not counted in the aggregate for that classroom. A student's test scores were included whenever a student took the test as scheduled, but no attempt was made to have a student do make-up tests.

Project Monitoring

The team divided the state into quarters with one PI responsible for each portion. Each PI had funds to employ graduate assistants to help with details, research, monitoring, and to assure that project protocols were followed. This included monitoring class assignments, assuring that students were in their assigned classroom, collecting data, interviewing, writing reports, and in other ways maintaining the fidelity of the project design. Research assistants helped disseminate results through presentations and publications; some developed their own research interests and used STAR as a base for their own studies.

Advisory Structure

Advisory committees helped the researchers. One committee included "experts" who had research and design experience and familiarity with other class-size initiatives. A second advisory committee included representatives of state educator groups who could help the researchers understand the political and cultural questions that might influence the interpretation, dissemination, and use of research results.

Some Actions Necessary to Assure a Sound Research Project

Since the initial selection of districts and schools required permission and thus could not be random, researchers checked the resulting sample against the state averages to assure comparability. Table 1 shows results of this comparison. STAR schools and districts were like the state averages in all respects, except that they did deviate slightly in district size (.05). The STAR districts were somewhat larger than the state average since the state's three largest districts were part of STAR (Memphis, Nashville, Knoxville).

TABLE 1 ABOUT HERE

At the end of K, some external critics commented that proportionally more "smart" kids were in (S) than in (R) or (RA). (This was part of the treatment effect.) STAR researchers checked the K demographics to confirm a "normal" distribution. With 6325 pupils assigned in K and 1900 of these (30%) in (S), one would expect that about 30% of the males, females, black, white, free lunch, (etc.) pupils would be in the (S) condition if assignment were random. Table 2 shows that this was true, validating the random distribution. (Serendipitously, the difference in percentages of pupils in special education is not a question of demographics – a person is not "special ed." until identified as such.) Apparently teachers in (S) are more adept at assessing pupil learning difficulties (A class-size benefit?) than are teachers in larger classes.

TABLE 2 ABOUT HERE

Besides checks early in the study, researchers periodically compared STAR results with state averages. It made sense that (R) classes in STAR should approximate the state average scores, since an (R) class was randomly established. Table 3 shows selected results of one second- grade analysis to check on the appropriateness of the STAR design and its random-selection process. This check, using only participants in all class types who had a full K-2 treatment, shows the state average percentile ranks on the SAT scores for reading and math. STAR schools in general score a bit ahead of state average and the (R) classes are very close to the average. The STAR (R) classes, with an average of 24 students, were also a bit smaller than the state average classes since no student could receive less by being in a STAR school.

TABLE 3 ABOUT HERE

Even though the in-school design was self contained, researchers still sought alternatives. Another way to watch STAR's progress was to have comparison points. There were two logical comparisons: (a) the state average as a benchmark, and (b) the selection of comparison schools.

A comparison group was established by asking 21 superintendents who had schools in Project STAR to identify another school in the district that was as alike as possible to the STAR school on a variety of variables (i.e., demographics), and to let STAR researchers collect test data from those schools. STAR personnel made no contact with the schools; they simply obtained the student test data from the State Testing Bureau and necessary demographic information about the teaching and administrative personnel from one brief questionnaire. Because of the fidelity of the in-school design, researchers made little use of external comparisons other than as benchmarks while the study was being conducted. One researcher analyzed test results from STAR (R) and from comparison-school classes to explore the issue of random vs. non-random pupil assignment and subsequent achievement, K-3. STAR (R) classes exceeded comparison schools in math and reading test results. (Zaharias, 1994; Zarharias, Achilles, Bain & Cain, 1995).

Sample Size and Power Analysis

The decision to use the class average as the primary unit of analysis influenced the sample size. While STAR had between 6300 and 7200 pupils per year, there were only about 100 classes of each type (S, R, RA). Researchers determined the minimum sample size through a power analysis and "over-sampled" as a precaution against pupil mobility and other real-life schooling factors that could confound the study results. Researchers needed enough schools and classrooms in 1985 to guarantee that there would be enough schools and classrooms left in 1989 to make the results consistent and believable. Approximately 90 classes of each type were required to meet the criterion of .95 confidence, so about 100 classes of each type were used each year.

Table 4 shows the actual distribution of classes by class types throughout the four years of STAR treatment, [(S), (R), and (RA)]. Only in K were no classes in the "out-of-range" area designated in the table as (B). [The designation in Table 4 for (S) is (A), and (C) shows the (R) and (RA) class range.] This frequency distribution shows that as STAR proceeded (that is, as the study followed pupils K-3) there was drift of classes into the out-of -range section (B), some drift of classes toward the large end of the distribution for (S), and toward the small end of the distribution for (R) and (RA). This shift had the potential to understate class-size differences, but once designated as (S), (R), or (RA) a class remained in that type for the entire study.

TABLE 4 ABOUT HERE

Summary of Strengths of the Study

Numerous design and methodological strengths were present in STAR, and its closely-related studies. A summary would include:

- random assignment of students and teachers;
- in-school design, study size, and longitudinal nature of the study;
- conservative analysis (the class unit), and comparison group;
- care and monitoring throughout the study, and advisory groups;
- external analysis for primary results.

Design Weaknesses

Although STAR was carefully designed, some limitations in STAR are:

- no very small schools (minimum K size=57 pupils);
- maintenance of class designation despite out-of-range drift;
- cohort effect as part of design;

- overall Hawthorne or Pygmalion (or "John Henry") effects;
- using the class as unit of analysis hindered reaggregating students for other analyses, such as race, poverty, or gender relationships, etc., as part of the primary results.

The database is intact and growing, however, so some weaknesses can be overcome through new analysis procedures such as the HLM analyses in progress.

Assuring Objectivity

Persons intimately involved in research may become advocates for the study, yet good research requires objectivity. The STAR PI team took steps to try to assure objectivity, including sharing the data with other researchers to have new analyses done. Steps to obviate bias include:

- External primary data analysis with internal confirmatory analyses.
- Discussions in staff meetings and with advisory boards.
- Making results public (peer review) annually and continually.
- Re-analyses of some results with different procedures.
- Reconfiguring the data and conducting studies on special topics, such as retention in grade, test-score gap reduction, school-size and class size, etc.

Serendipitous Outcomes

The magnitude of the STAR database, the experimental and in-school design, and the care in conducting STAR allowed researchers the luxury of doing subsidiary and ancillary studies. These studies may use the actual STAR data for detailed analyses; they may depend on STAR as the base and extend STAR [e.g., the Lasting Benefits Study (LBS) is following STAR students to see if early (S) benefits continue]; they may relate conceptually to STAR but be separate, such as Success Starts Small (Achilles, Kiser-Kling, Owens & Aust, 1994) or the Burke County (NC) study (Achilles, Harman, & Egelson, 1995). Several students have completed dissertations using the STAR database to answer questions that were not in the original study. Table 5 shows selected studies with brief designations of the authors(s) and purposes of each study. Some studies mentioned in Table 5 are being extended using different analysis procedures. STAR personnel have also arranged with researchers at the Institute of Education at the University of London to re-analyze the basic STAR data.

TABLE 5 ABOUT HERE

Dissemination of Results and Policy Use of Results

Because researchers have been so busy "researching," part of the dissemination function may have been neglected. External persons who reviewed STAR have published information about it (e.g., Mosteller, 1995). STAR researchers have widely disseminated the class-size results in journals, at conferences [AERA, AASA, NCPEA, MSERA, NAESP, Quality Schools Conferences, NEA (Board of Directors, Regional, State and local conferences; etc.) and at State Conference] internationally (England, Sweden), through ERIC, at workshops, etc. A 50-page bibliography of class-size work lists most of the STAR materials and other class-size studies (Nye et al., 1996). STAR findings have generated considerable policy debate and positive action. For example, by 1996 leaders in 18 states have considered some class-size legislation or policy, and other state leaders continue the policy discussions. The Arizona legislature published a class-size report (Shaw & Sheane, 1995) in response to a fairly negative and speculative report from the Goldwater Institute (Flake, et al., 1995).

Summation

The strengths of the STAR design add confidence to researchers in their discussions about the potential of class-size reduction in primary grades not only to influence student achievement, but also to serve as the base for meaningful "restructuring" activities. Results of STAR and STAR-related studies provide strong research evidence to support common-sense ideas that smaller classes are beneficial. Plus, STAR results "square" with studies of early childhood education, with studies of Kindergarten, and with common sense.

Prospective Proposals

The general coronary health of Americans has been dramatically changed by the longitudinal Framingham Heart Study (Dawber, 1980; Kannel, Dawber, Kagan, Revotskie & Stokes, 1961). Results of this study have influenced actuarial tables; diets; labels on foods; risk factors such as weight, cholesterol, smoking, fat; life-style changes in exercise and stress, etc. The Framingham Heart Study changed the arena of Coronary Heart Disease (CHD) and its treatment based on careful examinations of approximately 5200 people beginning in 1948. Project STAR has class-size K-3 treatment data on over 11,000 students and researchers continue to examine the results. Studies by Calhoun (1972) of the "behavioral sink" caused by crowding among Norway rats, by Tinbergen (1952) of destructive behavior generated among stickleback fish, and studies of asocial behavior caused by crowding in huge public housing projects such as Pruitt-Igoe in St. Louis (Hall, 1972) offer fodder for considering the impact of large classes (crowding) on the very young – especially children from crowded living quarters. STAR, unfortunately, did not have very "crowded" classes [the average (R) class was only 24], but there is evidence of behavior differences favoring pupils who had the (S) treatment over those who had (R) and (RA) starts in school. Here, is one example of the STAR's heuristic potential.

Notes on Project Challenge and The Lasting Benefits Study (LBS)

Two studies closely related to STAR are continuing STAR's legacy due, in part, to STAR's design and findings. In Project Challenge 16 of TN's poorest districts received funding for class-size reduction. Over time, the 16 districts' average rankings in math and reading moved from way below the state mean to slightly above it (in grades 2 and 3). Project Challenge results show that small classes (1:15) influence pupil achievement positively in grades K-3 and support findings of the LBS. The LBS showed that, in general, STAR small-class benefits were continued at least into grade 8, although the amount of the benefit declines or fades [from a grade-3 effect size (ES) of about .6 to a grade-8 ES of about .15] as students move through the grades. The downward change in the rankings of systems in Challenge between grade 3 (the last grade of class-size reduction) and grade 4 (after students return to the regular-sized classes) suggests that in poverty situations (e.g., Challenge) the class-size effect fades more quickly than in conditions of less poverty in STAR's random sample. This idea "squares" with other research, and with Hodgkinson (1992, 1995) and Cooley (1993), who point out that poverty is the most important factor that educators must deal with in terms of pupil achievement. STAR, Challenge, LBS, and most of the subsidiary and ancillary studies show that 1:15 is a good "treatment" for K-3 pupils for across-the-board achievement increases.

Dissemination

Dissemination of research results -- use of research results -- has been a weak link in education improvement. Mosteller's (a995) critique had a great impact on dissemination and acceptance of STAR/LBS results. Mosteller's work, through its connection with the American Academy of Arts and Sciences (AASA) added structure to STAR results and opened new questions of dissemination.

Dissemination and use of important research findings must get greater attention; results need to get into the hands of those who are committed to improve public education. Given the competing forces for limited resources, educators need the research results, the political savvy and the professional cohesiveness to speak strongly for what research supports. (Note that other successful "programs" really employ a small-class base: Reading Recovery, Success For All, Peer Tutoring, etc.).

The Rest of the Story

The rest of the story waits. Researchers plan to follow STAR students at least until they exit grade 12 (1998), and to report Challenge and LBS results annually. Decisions made early in STAR generally support the later studies and have made the researchers' work easier. Most "glitches" did not influence STAR negatively. Lucky? Serendipitous? Perhaps.

References

Achilles, C. M. (1990, Summer). Research in education administration (EA): One position. The AASA Professor, 13 (1), 1-4.

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

Achilles, C. M., Kiser-Kling, K, K., Owens, J., and Aust, A. (1994). Success Starts Small: Life in a Small Class. Greensboro, NC: University of North Carolina at Greensboro, Final Report. Small-grant/School-based research grant.

Achilles, C. M., Nye, B. A., Zaharias, J. B., Fulton, B. D. (1995, April). Policy use of research results: Tennessee's Project Challenge. Paper at Annual Meeting of American Educational Research Association, San Francisco, CA.

Achilles, C. M., Nye, B. A., Zaharias, J. B., Fulton, B. D. & Cain, V. (1996, March), Education's equivalent of medicine's Framingham heart study. Paper at Fifth annual Conference on Creating the Quality School. Oklahoma City, OK. (March, 1996).

Calhoun, J. B. (1962, February). Population density and social pathology. Scientific American, 206 (2).

Cooley, W. W. (1993, May). The difficulty of the educational task. Pittsburgh, PA: University of Pittsburgh. Pennsylvania Educational Policy Studies. Paper #16.

Davis, T. M. & Johnston, J. M. (1987). On the stability and internal consistency of the self-concept and Achievement motivating Inventory: Preschool/Kindergarten form. Psychological Reports, 61 871-874.

Davis, T. M. , Sellers, P. A., & Johnston, J. M. (1988). The factor structure and internal consistency of the self-concept and Achievement Motivation Inventory: What face would you wear? Preschool/Kindergarten form. Educational and Psychological Measurement, 48 237-246.

Donmoyer, R. (1993, April). Yes, but is it research? Educational Researcher, 22 (3), 41.

Flake, J. (1995). Class size and Student Achievement. Is there a link? Phoenix AZ: Goldwater Institute.

Fowler, W. J. Jr. & Walberg, H. J. (1991, Summer). School size, characteristics, and outcomes. Educational Evaluation and Policy Analysis, 13 (2), 189-202.

Hall, E. T. (1972). The Hidden Dimension. New York: Doubleday.

Hodgkinson, H. (1992) A Demographic Look at Tomorrow. Washington, DC: Institute for Educational Leadership.

Hodgkinson, H. (1995, October) What shall we call people? Phi Delta Kappan, 77 (2), 193-179.

Kaestle, C. F. (1993, January-February). Research News and Comment: The awful reputation of educational research. Educational Researcher, 22 (1), 23-31.

Mosteller, F. (1995 Summer/Fall) The Tennessee study of class size in the early school grades. The Future of Children; Critical Issues for Children and Youth, 5 (2), 113-127.

Mueller, D. J., Chase, C. I., & Walden, J. D. (1988). Effects of reduced class sizes in primary classes. Educational Leadership, 45 (5) 48-50.

Nye, K .E. (1995). Do small classes help ameliorate the generally negative effects of large schools on pupils achievement, K-3?. Unpublished doctoral dissertation. University of Tennessee, Knoxville.

Nye, B. A., Achilles, C. M., Sotoohi, G., Cain, V. A., Fulton, B. D. & Zaharias, J. B. (1996, March). Draft Bibliography of Class-Size Research. Nashville, TN: Center of Excellence for Research and Policy on Basic Skills. TN State University.

Nye, B. A., Zaharias, J. B., Cain, V., Fulton, B. D., Achilles, C. M., (1995a). Project Challenge: Fifth-year Summary Report. (Also 1994, 1993, 1992, 1991 annual reports). Nashville: Center of Excellence for Research and Policy on Basic Skills. TN State University.

Nye, B. A., Zaharias, J. B., Cain, V., Fulton, B. D., Achilles, C. M., (1995b). The Lasting Benefits Study: Technical Report, Eighth Grade. (Also 1994, 1993, 1992, 1991 annual reports). Nashville: Center of Excellence for Research and Policy on Basic Skills. TN State University.

Orlich, D. C. (1991, April) Brown v. Board of Education: Time for a reassessment. Phi Delta Kappan, 72 (8), 631-632.

Shaw, R. R. & Sheane, K. E. (1995). RESCUES '96: Research in Education on Small Classes Uncovers Educational Success. Phoenix, AZ: Arizona House of Representatives.

Tinbergen, N. (1952, December). The curious behavior of the Stickleback. Scientific American, 187 (6).

Wagner, J. (1993, June-July). Ignorance in educational research, or How can you NOT know that? Educational Researcher, 22 (5), 15-23.

Word, E., Johnston, J., Bain, H., Fulton, D. B., Boyd-Zaharias, J., Lintz, M. N., Achilles, C. M., Folger, J., and Breda, C. (1990) Student/Teacher Achievement Ratio (STAR): Tennessee's K-3 Class-size Study. Nashville: TN State Department of Education.

Zaharias, J. B., Achilles, C. M., and Cain, V. A. (1995, Fall) The effect of random class assignment on elementary students' reading and mathematics achievement. Research in the Schools, 2 (2), 7-14.

Table 1
Discussion and Indicators that STAR Sample Was Equivalent to TN Systems on Various Measures.

Item	STAR Average	State Average
Per-Pupil Expenditure (1986-87)	$2,724	$2,561
Average Teacher Salary	$23,168	$22,627
Average System Size	8,462	4,202**
Teacher-Pupil Ratio Kindergarten (1986-86)	22.7*	22.3
Percent of Teachers with Master's Degree or Higher (System Figures)	42	40

* Based on regular-sized STAR classes.
** $p \geq .05$.

Note: Project STAR systems are weighted by the number of pupils or teachers from each system who are participating in the project.

A comparison of test scores for grade-two students in project schools, the comparison schools, and the statewide average indicated that project schools had scores lower than the state average and the average of the comparison schools. These differences reflect the higher proportion of inner-city schools in STAR; students in inner-city schools scored 10 to 12 points lower on the average than students in suburban schools. Differences in scores among urban, rural, and suburban schools were smaller. The comparison schools did not include any inner-city schools. STAR schools in the same systems with comparison schools scored slightly (not significant) higher than the comparison schools.

Spring, 1986	Math	Reading
State Average for 2nd Grade	572	582
All Project STAR Schools	566	578
Comparison Schools	577	587
STAR Schools (Same Systems as Comparison Schools)	579	590

From Word et al. (1990).

Table 2.
STAR Kindergarten (1985) Pupils Shown by Their Distribution (%) on Selected Demographic variables into Three Class Types (S, R, RA).

	CLASS TYPE						
	S		R		RA		Total
Total N	1900		2194		2231		6325
% by Type (Tot)	30.0	Dif*	34.7	Dif*	35.3	Dif*	100
% Male	30.1	+.1	34.4	-.3	35.5	+.2	100
% Female	30.0	0	35.0	+.3	35.0	-.3	100
% Nonwhite	29.0	-1.0	34.5	-.2	36.5	+1.2	100
% White	30.6	+.6	34.8	+.1	34.7	-.6	101**
% Free Lunch	29.2	-.8	34.2	-.5	36.6	+1.3	100
% No Free Lunch	30.8	+.8	35.2	+.5	34.0	-1.3	100
% Sp Ed	35.6	+5.6	33.2	-1.5	31.2	-4.1	100
% No Sp Ed	29.9	-.1	34.7	0	35.4	+.1	100

* Difference (+, -) from "expected" distribution based on the proportion in Total. If 30.0% of students are in S, 30.1% of males would be in +.1%.

** Rounding. This reflects the .1% error internally.

Table 3
Grade Two Comparisons of STAR Results with State Indicators

	SAT Scaled Score		Percentile Rank	
	Rounded			
2nd Grade and Class Type	Reading	Math	Reading	Math
State Norm	----	----	59	73
Total STAR (1988)	594 (N=1,426)	588 (N=1.422)	65	78
Small	599 (N=817)	593 (N=813)	68	81
Regular	587 (N=286)	584 (N=286	59	75
Regular & Aide	588 (N=323)	582 (N=323)	61	74

Small= 13-15; Regular and Regular & Aide = 23-27. Sample uses only students in STAR for K + 1+ 2.

Table 4

Distribution of STAR Classes by Grade (K-3) by Designation S (Small), R (Regular), and RA (Regular and Aide)

		K (n classes)			1 (n classes)			2 (n classes)			3 (n classes)		
		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	4		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		116	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 5.
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).

CATEGORY, TITLE & PURPOSE *	DATE(S)	AUTHOR(S) OR PUBLICATION
Subsidiary Studies		
• Lasting Benefits Study to follow STAR pupils	1989-Present	Nye et al., 1994
• Project Challenge (TN)	1989-Present	Nye et al., 1994, Voelkl, 1995
• Participation on Grades 4, 8	1990, 1994	Finn, 1989 Finn and Cox, 1992
Ancillary Studies (Use or extend STAR data. Some of these are dissertations.)		
• Retention in Grade	1994	Harvey, 1994
• Achievement Gap	1994	Bingham, 1993
• Value of K in Classes of Varying Sizes (tests scores)	1985-1989	Nye et al., 1994-1995
• School-Size and Class Size Issues	1985-1989	Nye, K., 1995
• Random v. Non-Random Pupil Assignment and Achievement	1985-1989	Zaharias, 1995
• Class Size and Discipline in Grades 3,5,7	1989,1991,1993	In Process, Hibbs.
• Effective Teacher Analysis (top and bottom 10% of STAR teachers)	1985-1989	Bain et al., 1992
Related Studies		
• Success Starts Small: Grade 1 in Chapter 1 (1:14, 1:23) Schools, Burke Co., NC	1993-1995	Achilles et al., 1994

* This list is not complete. It provides samples of the types of studies done. Not all authors appear in the references in the exact way listed here. This table appears in several STAR reports in substantially this same form.

Date: Fri, 25 Oct 1996 17:29:24 +0200 (MET DST)
X-Sender: wbruin@solair1.inter.nl.net
To: CM Achilles <sheckle@vivanet.com>
From: "W. de Bruin" <wbruin@redactie.volkskrant.nl>
Subject: Re: STAR

Dear dr. Achilles,

an udate on class[size reduction in the Netherlands

tuesday 18 october the commission on class size reduction presented its report. Based om american experiences, dutch investigation in the Prima cohort, and good sense as the chairmen said, the commission concluded that class size matters. They advise a class size of an avarege of 20 in the first four years, age 4 tot 7, and 28 in the classes 5 till 8 until age 12. Cost are estimated on 1.000.000.000 guilders, about 400.000.000 dollar. The commision expects from the schools that they improve their methods and have to publish more data about what schools accomplish. At this moment schools dont have to publish their results. Also the commission wants to give parents council power in the distriution of teachers at the classes. There is an implementation scheme from 1997 till 2001 in three steps.

The secretary of state reacted unexpectetly in favour of the plan. She announced a first step at the beginnening of next schoolyear and has the support of the minister on finance to make an investment in primary education

The situation of class size changed radically the last two months, in which publication of the star[reort in our newspaper played a role.

Thank you for your help, and I am planning a visit tot the USA later this or next year when the details of the class size reduction in the Netherlands are fillid in.

Robert Sikkes

best e mail adress; forum@volkskrant.nl

At 10:46 PM 10/13/96 -0400, you wrote:
>The STAR technical report is available for $20 (US) from P. Egelson at
>SERVE, POB 5367, Greensboro, NC 27435. (A check should be made out to
>SERVE). If I get a couple, I"ll mail one to you if you don't buy one.
>CM Achilles
>
>
>

TM026294

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: "THIS," as Paul Harvey Was wont to say, "IS THE REST OF THE STORY"

Author(s): C.M. Achilles, H.P. Bain, B.D Fulton

Corporate Source: RPC, TN. State UNIV., 330 10th Ave N. Nashville, TN 37203-1341

Publication Date: 11/7/96

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

[X] Check here
For Level 1 Release:
Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) ***and*** paper copy.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

The sample sticker shown below will be affixed to all Level 2 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

[ ] Check here
For Level 2 Release:
Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but ***not*** in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature: C M Achilles	Printed Name/Position/Title: Achilles Professor Ed. Adm	
Organization/Address: E. Mich. UNIV, College of Educ. Ypsilanti, MI 48197	Telephone: 313 487 0255	FAX: 313 487 4608
	E-Mail Address:	Date: 11/7/96

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:

Address: N/A

Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:

Address: N/A

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Acquisitions
ERIC Clearinghouse on Assessment and Eva;uation
210 O'Boyle Hall
The Catholic University of America
Washington, DC 20064

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2d Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com