QUALITY SCHOOLS BUILD ON A QUALITY START♦

by

Jayne Boyd-Zaharias
Center of Excellence: Basic Skills
Tennessee State University

C.M. Achilles
Eastern Michigan University

Helen Pate-Bain
Education Consultant

Barbara A. Nye, Director
Tennessee State University
Center of Excellence: Basic Skills

B.D. Fulton
Tennessee State University
Center of Excellence: Basic Skills

ABSTRACT

Quality schools build on a quality start. Results of Project STAR, Tennessee's randomized, longitudinal (1985-1990) experiment of class-size effects (K-3) are definitive. Results of ancillary studies using the STAR database (almost 10,000 pupils) are becoming available. Studies include class-size effects and: grade retention, test scores" gap" reduction (white/non-white) and "value" of kindergarten, lasting benefits (to grade 7) of early small-class participation, and school size/achievement. Researchers will also review STAR design and outcomes. The paper is presented as a letter responding to questions about the studies.

♦ This paper was presented at the third annual National Conference on Creating the Quality School, Oklahoma City, OK, April 1994. It also appeared as a book chapter in *Creating The Quality School*, edited by Edward W. Chance and published by Magna Publications, Inc. of Madison Wisconsin.

April 1, 1994

Concerned Reviewer
Quality Schools Conference
c/o CSSRS, University of Oklahoma
Norman, OK 73037-0005

Dear Reviewer:

Your letter to the Center of Excellence for Research for the Basic Skills, Tennessee State University, is much appreciated and deserves careful response -- probably in more detail than we can provide in this limited space. We'll answer your questions and provide you an update on some of the original and continuing STAR research, as well as on the Lasting Benefits Study (LBS) and Tennessee' Project Challenge findings. Many of these results appear in Achilles, Nye, Zaharias, and Fulton (1993), Finn and Achilles (1990), and Word, Johnston, Bain, Fulton, Zaharias, Lintz, Achilles, Folger, and Breda (1990).

Some of the questions that you raise cannot easily be clarified in a brief letter or with a "yes" or "no" response; they really need a symposium-like atmosphere where we could access the STAR/LBS database to explore all the loose ends. This is a useful database with nearly 10,000 pupils assigned at random to one of the three conditions (Small or S, Regular or R, and Regular with a full-time aide or RA), and with teachers assigned at random in a diverse group of 79 schools. The STAR "within-school" design [wherever there was the (S) condition there were at least one each of (R) and (RA)] controlled for building, district, etc. issues. We also "hedged our bets" by including a set of 20 comparison schools in STAR districts that were as well matched to the STAR sites as possible. All of this plus the LBS data are "on line" (STAR kids are in grade 8 in 1993-94). This provides an opportunity to do what Cooley and Bickel have called "decision-oriented research."

Using the STAR database we have been considering questions related to teacher aides, pupil retention in grade, school size and class size, student participation (Finn & Cox, 1992), homogeneous/heterogeneous grouping [analyzing STAR regular (R) results with comparison school results], achievement issues related to race and class size/class condition, the conundrum of the Black/White achievement gap on norm-referenced tests (NRT), analyses of "quality" teachers as determined by classes achieving in the top 10% each year on gain scores, etc. Some of this work should produce interesting results that you may want to review.

Before we move to your specific concerns, we'd like to point out that the initial "charge" for Project STAR was to find out if classes of about 15 made a difference in pupil achievement as measured by the instruments used in Tennessee [e.g., the NRT was the Stanford Achievement Test battery or SAT and the Criterion Referenced Test (CRT) was the Basic Skills First or BSF]. Results were to guide the Tennessee Legislature and State Board of Education in setting the state teacher-pupil ratios based not on a guess or on tradition but on carefully designed research. As such, we manipulated only one variable, class size, and also checked on the efficacy of using teacher aides in larger classes as a way to help increase achievement.

Smaller classes achieve better than larger classes with little or no adjustments by teachers; at least now we know that. If you use about 15 pupils as a base and do other things, you get even greater effects as Slavin, Madden, Karweit, Livermon and Dolan (1990) and Madden, Slavin, Karweit, Dolan and Wasik (1993) have shown in "Success for All." They have not demonstrated large gains with large classes in "Success for All" -- at least that we know of. But, this point deserves added discussion.

Now to some of the points in your letter. (Any data included are still "draft" as they have not been released for peer review and some are still exploratory. We'll share the final papers when they are available. Please treat these data as "draft" or "tentative.") Numbers at left refer to questions in your letter.

1. You asked if students "learned more each year." This 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 (Evertson and Folger, 1989) than pupils in (R) and (RA).

> 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 believe.

You asked about the additiveness of learning each year. The lack of an increase in effect size is disturbing. We are exploring retention in grade, special education placement and "passing" scores. For example, a smaller percentage of pupils are 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.) As the (S) class moved along as a cohort it had lower scoring pupils than did the (R) class. (Also higher. The ratio may be key.) As cohorts moved through the grades they would "pick up" pupils who had been retained the previous years. Also, as the years of the study went by, the line between (S) and (R) became blurred. (More on this when we discuss Table 2).

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

Decision	Class Condition		
	Small	Regular	Regular+Aide
Promote	441	435	436
Retain	422	427	421
Difference	19	8	15

We may not have reanalyzed the database if you (and others) had not raised the questions you have about the size of the differences between groups. Our responsibility was to answer the question in the legislation. The answer was clearly 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 (etc.) we returned to the data. <u>It is very clear that the reported results were conservative</u>. Your question about, , the 1:15 ratio spurred us to check the frequency distributions of class sizes in the S,R,RA conditions over the four years (K,1,2,3) of the study. Several points are important here. Table 2 provides the frequency distribution information from which we developed the answers below.

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

Grade and Classes (n) by Condition

		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	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.

a. A class was designated S,R,RA based upon the K distribution. If a (S) class grew "out of range," we still considered it (S) for the analyses as the class was set up in K. A class designated (R) was still treated as (R) even though it may have shrunk "out of range." Now that we are being asked for more precise analyses, we'll continue to refine and use the database. This will mean that we'll lose some classes from the original analyses.

b. The preponderance of in-range (pupil n=12-17) small classes have 16 or 17 pupils (except grade 2); the preponderance of in-range (n=22-26) regular classes have 22-23 pupils (except grade 3). Thus, many (S) classes have more than 15 pupils and many (R) classes have fewer pupils than 24 as the study proceeds from year to year.

c. The distribution of "out of range" classes (n=18-21) seems to favor (R) or (RA), but this remains to be tested. We believe this for several reasons: Table 3 shows basic Pearson (r) for class size with average scores for each class size using the composite of (S) and (R) with all classes in the study and only with "in-range" classes. The correlations generally increase with only "in-range" classes. Also, most of the "Most Effective" teachers based on gain scores in Grade 3 were in (S) (Table 4).

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

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

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

2. You asked if small classes by themselves caused achievement gains. We <u>are</u> saying that small classes <u>by themselves</u> lead to improved achievement and that, to get that achievement, teachers needn't do anything special. Finn's (1993) and Finn and Cox's (1992) notion of increased student "participation" may be important in this discussion. Slavin et al. (1990) show that if teachers do other things, achievement goes up. Our research suggests that some of Slavin et al. (1990) and Madden et al. (1993) Effect Size (ES) may well be due <u>only</u> to their starting with a base of n=15 pupils to one certified teacher. Please refer to results provided in Table 5. This analysis was based on class average achievement for (S) at 13-15 and (R) at 23-27, for all pupils in these classes regardless of the number of years in STAR. This analysis removes the "drift" to "out-of-range" classes and begins to develop a clear picture of (S) vs (R). We like these effect sizes as a base for other education interventions.

3. You suggest that the findings are "mundane." We disagree. The finding that students in small classes learn more than students in larger classes (K-3) is not "mundane." We are gratified to find among the poor research (or data) out there at least one fairly secure finding supported by a strong design, longitudinal and conservative analysis, etc. 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 STAR/LBS/Challenge and ancillary studies (Bingham, 1993; Harvey, 1993; Boyd-Zaharias, 1993, etc.) meet these tests.

4. You ask if teacher "anticipation" could cause the results. Randomization, regrettably, does not rule out anticipation. Perhaps if we inject teachers of larger classes with anticipation, we can increase class sizes. We did find in our interviews (we conducted over 1000) that teachers of larger classes spoke of not letting teachers in smaller classes beat them out. We referred to this as the "John Henry" effect. However, all teachers <u>should</u> anticipate that they will make a difference with their students. Clearly, as you suggest, teachers in (S) would anticipate that they would do better. We do not know of a research design to alleviate this issue.

There are some puzzling findings compared to prior studies. We have continued to work on the data to try to see "why." One reason that we are finding a continuing benefit of (S) on academic achievement when pupils from (S) go back to regular classes in grade 4 and later (the LBS) may be the <u>size</u> of this study, <u>or</u> the fact that pupils were in the (S) condition for several years, rather than the short, one-shot approaches reported before. (We're doing a small observational study of K-1 classes this year and reviewing other class-size studies to try to expand our ideas on this.)

5. Does class-size reduction work outside the "experiment"? We are also monitoring Project Challenge in Tennessee to see if we can get any clues from it. After reducing class sizes (K-3) in 17 of Tennessee's poorest counties (Project Challenge), the systems have moved substantially upward in the ranking of the Tennessee systems -- even to well above the average rank in math. Strange! These usual doormats now are part of the "above average" systems. This was one policy application of STAR results. See Table 6.

6. As to your question why brighter students do not seem to profit more than the disadvantaged, we can only at this point offer conjecture. You will note that Robinson said the same thing in his ERS (1978, 1980) and <u>Educational Leadership</u> (1990) reviews of the class-size research. The instrumentation may have had a ceiling effect for bright kids. The less bright kids had a lot further to go and the (S) condition may have been a facilitating situation for the growth. But, as we said earlier, some of these points need scholarly review, not a passing comment in an exchange of letters.

7. What are we doing now with the database? 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 some 9,000 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 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 initiated the series of Ancillary Studies. The following are 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 hope to repeat this in grades 8 and 12. Finn has linked student participation to not dropping out of school (1993).

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 retainment; (S) is not a treatment or remediation later after retention has occurred (Harvey, 1993).

C. "Gap reduction" between White and minority pupils. The (S) condition seems to prevent a large gap opening (K-3) between test 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, initial (S) placement seems to help keep this gap from opening.

D. Homogeneous or heterogeneous? Using the randomly-assigned STAR (R) classes 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 (Boyd-Zaharias, 1993).

E. Class size/school size. Numerous studies (e.g., Barker & Gump, 1964; 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. The school-size effect was present in the analysis of test scores of (R) pupils in the STAR schools (negative correlation, p .05).

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" is in process. Researchers hope to explore the range of questions inherent in "Is the (S) treatment a preventive or a remedial event?"

The ancillary studies are leading more and more to the idea that a small-class start in school is a preventive, not a remedial, event. This seems consistent with the positive Reading Recovery results and the generally dismal results of much of Chapter I. We're doing some reading outside of education to help with this line of reasoning. Strange as it may seem, a work by C. Pfaffenberger (1963), The New Knowledge of Dog Behavior, New York: Macmillan (esp. Chapter IX) may have support for the prevention vs remediation idea. So, too, might emergent work in Paleobiology, such as reported by Rush Dozier, Jr., (1993) Codes of Evolution, New York: Crown, or R. Ornstein (1991), The Evolution of Consciousness, New York: Prentice Hall. Where the window of learning opportunity is open, we need to take advantage of it before it slams shut.

We believe that this continuing work on seeking appropriate class sizes for pupil learning and on the effects of pupil class sizes is important, especially in light of the changing demographics of children entering our schools (e.g., Hamburg, 1992; Hodgkinson, 1992). We know that quality schools begin with quality preschool (e.g., Weikert, 1989; Zigler, 1992) and continue with quality early primary programs.

Some of these results may help unravel common effectiveness and efficiency mix-ups. Often policy folks judge schools by effectiveness measures (e.g., test scores) and then hamper the possibility of meeting the standards by imposing efficiency shackles (e.g., larger classes, larger schools, more grade retention, tracking) that, by themselves, negatively impact the achievement of the standards.

We continue to analyze the STAR and LBS and Challenge data. Critical review helps us refine our questions and approaches. We are heartened by increasing numbers of positive comments on STAR design, analyses and findings. We hope that this clarifies the issues you raised in your letter, that the new information may be useful, and that you'll consider some of this in your future deliberations. We'll provide added research data and rationale for any claims or assertions that we make.

Sincerely,

The STAR/LBS/Challenge Team

References/Bibliography

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.

Barker, G., & Gump, P.V. (1964). Big school, small school. Stanford, CA: Stanford University Press.

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.

Boyd-Zaharias, J.* (1993, December). The effects of random class assignment on elementary students' reading and mathematics achievement. Unpublished Ed.D. dissertation, Tennessee State University, Nashville.

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). Multivariance VII user's guide. Chicago, IL: Scientific Software, Inc.

Finn, J.D., & Cox, D. (1992, Spring). Participation and withdrawal among fouth-grade pupils. American Educational Research Journal, 29(1), 141-162.

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

Hamburg, D.A. (1992). Today's children. New York: Time Books, Random house.

Harvey, B.* (1993, December). An analysis of grade retention fn pupils in K-3. Unpublished Ed.D. Study, University of North Carolina at Greensboro.

Hodgkinson, H. (1992, June). A demographic look at tomorrow. Washington, DC: Institute for Educational Leadership.

* 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.

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.

Nye, K.* (1994, In Process). Do small classes help ameliorate the generally negative effects of large schools on pupil achievement, K-3?

Robinson, G.E. (1990, May). 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.

U.S. Department of Education. (1993). Background document on proposed funding: CSEARS. Washington, DC: OERI.

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.

Zigler, E. (1992, June 27). Head Start falls behind. New York Times, p. 15.

* 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.