DOCUMENT RESUME

ED 363 956 EA 025 441

AUTHOR Achilles, C. M.
TITLE The Teacher Aide Puzzle: Student Achievement Issues. An Exploratory Study.
PUB DATE Nov 93
NOTE 33p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 1993).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Class Size; Elementary School Students; Grade Repetition; Instructional Effectiveness; Primary Education; Scores; *Small Classes; *Teacher Aides; *Teacher Student Ratio
IDENTIFIERS *Tennessee

ABSTRACT

This paper presents findings of a study that examined the relationship between the performance (as measured by test results) of pupils in classrooms with full-time teacher aides. The Student Teacher Achievement Ratio project (STAR) was implemented in Tennessee from 1985 through 1989 to study the effects of class size on student achievement and development in grades K-3. Data for this paper were derived from an analysis of data from Project STAR and its derivative studies. Project STAR compared the results of three situations: (1) small classes with a 1:15 teacher-pupil ratio (S condition); (2) regular classes with a 1:24 teacher-pupil ratio (R condition); and (3) regular classes with a full-time teacher aide (RA condition). In nearly all comparisons, the small classes had the. highest student test outcomes, followed by regular-sized classes with aides and regular-sized classes. However, pupils who were retained a grade before entering the STAR program benefitted most in their test scores from the RA condition and often least from the S condition. Implications are that retention fails to help students improve test outcomes and that small-class intervention does not remedy already-defined test-score deficits after students have experienced regular classes. Thirteen tables and an appendix containing statistical data are included. (LMI)

THE TEACHER AIDE PUZZLE: STUDENT ACHIEVEMENT ISSUES. AN EXPLORATORY STUDY.

C.M. Achilles
B.A. Nye
H.P. Bain
J. Zaharias
D. Fulton

Paper Presented at
Mid-South Educational Research Association
(MSERA)

New Orleans, LA
11/12/93
R

Running Head: Teacher Aide Puzzle

TEACHER AIDE PUZZLE: STUDENT

ACHIEVEMENT ISSUES

Introduction

Although as part of the original Project STAR (1985-1989) analysis there was some consideration of the effects on pupil achievement and development of a pupil being in a "regular" class (a class with 22-26 pupils, with an average of 24) with a full-time teacher aide (designated RA), the literature review identified that, really, there was little research on the topic of teacher aides and pupil achievement. In fact, there was surprising little research on the full range of teacher aides (use, training, effects, etc.). Since STAR was specifically a study of the effects of class size on pupil achievement and development in K-3, the major analyses were of class size, with the Teacher Aides questions treated secondarily. The basic STAR analysis and subsequent ancillary studies using the STAR database raised questions that have prompted additional attention to the "Teacher Aide Puzzle." This present paper reviews some of the issues but focuses almost exclusively on performance (test results) of pupils in classrooms with full-time teacher aides in Project STAR.

Since the mid-1960s the concept and use of teacher aides -- paraprofessionals, teacher assistants -- have grown. Federal support for teacher aides was found in education legislation and especially in ways that Local Education Agency (LEA) leaders chose to implement (then) Title I of the Elementary and Secondary Education Act, or ESEA (PL89-10). The various amendments to PL89-10 have produced new PL numbers, slightly reduced rules and regulations for implementation, and even a new designation (Chapter 1), but a mainstay of programs supported by the original Title I has remained the use of teacher aides or "parapros" in classrooms.

For a program in such common usage, there is little serious research/evaluation information available about teacher aides: demographics, preparation and training, use in classrooms and use school-wide, their influence on pupil achievement (this seems like a major

issue), etc. In some states teacher aides have strong associations (unions) that bargain for them for salary, fringes and working conditions. Teacher aides have been a part of some state education reform legislation. (E.g., in Tennessee the state provided at least one aide for each three teachers, K-3). Indeed, since much state legislation regarding education from the mid 1980s to now has been aimed at improving student test scores, one would infer that the emphasis on aides implies that state education reform advocates believed that there was positive benefit for student achievement (translate as increased test score outcomes) related to the use of teacher aides. This 1980s idea has a long legacy: a major purpose of Title I -- and still of Chapter 1 -- was to improve the education outcomes of "educationally deprived" students. Some 1965 debates in Congress regarding the passage of PL89-10 centered around whether this was education or poverty legislation. Given the increase in the use of teacher aides that can be traced to PL89-10, one might rightly ask if its continuation has been education or employment legislation. Expenditures for teacher aides adds credence to the claim that ESEA is employment related. What about the education issues?

Purpose

The purpose of The Teacher Aide Puzzle study is to provide a review and summary of theory, evaluation and research on the phenomenon of teacher aides, especially on the use of teacher aides in classrooms. Because of the dearth of research and evaluation data regarding teacher aides, and especially teacher aides related to student achievement, the paper relies heavily on data from Project STAR (Student Teacher Achievement Ratio) and its derivative studies and analyses, particularly the Lasting Benefits Study or LBS. Indeed, Slavin (nd) notes that the STAR database should be very useful in trying to understand the uses of teacher aides: "The longest and best-designed study on the use of aides was part of the Tennessee class size study" (p. 9).

Project STAR (1985-1989) is best known as a major study of the effects of class size (1:13 - 1:17 with an average teacher-student ratio of 1:15 vs 1:22 - 1:26 with an average ratio of 1:24) on pupil achievement and development (Finn & Achilles, 1990, Word et al., 1990; Finn et al., 1989;

Nye et al., 1992; and numerous other reports and articles by authors of the above publications). [Note: The 1:24 ratio in STAR may have been smaller than the average regular class size in other districts throughout America during 1985-89, the time of Project STAR. This was partly due to Tennessee's' average class size in early primary grades (K-3) at the time, and to the fact that when STAR researchers set up the STAR project, the design created conditions where the regular class in STAR schools had slightly fewer students than the average regular classes statewide.]

In establishing Project STAR, the Tennessee legislature "hedged its bet" that small (S) classes (1:15) would do better than R classes. Since (S) classes would be expensive to implement statewide, the legislation required that the design contain the condition of a regular class with an aide (RA) dedicated to work full-time in the regular class. If this condition should provide improvements similar to those expected from the small-class condition, then the (RA) condition might be employed at a considerable savings. So, STAR really had three options [Small or (S), Regular or (R), and Regular with a Full-time Aide (RA)] and could as well have been considered a study of the effects on pupil achievement and development of using a teacher and a full-time teacher aide in a class of approximately 24 as being considered a study of the effects of a small teacher-pupil ratio (1:15).

The LBS has followed STAR pupils from all three conditions as they exited STAR and entered regular classes in grade 4. Although the STAR pupils are in grade 8 (1993-94) the researchers have only completed the analysis of LBS data through grade 6. The present material includes a synthesis of research on teacher aides and their use, and it relies heavily on STAR and LBS, particularly due to the size, scope and design of STAR

The Basic STAR Analyses (Summary)

In STAR the unit of data collection was the pupil, but the unit of analysis was the class (the class average) performance on a norm-referenced test (NRT), the Stanford Achievement Test (SAT) series (K-3) and on Tennessee's criterion-referenced test (CRT), the Basic Skills First (BSF) for grades 1-3). [The design for the STAR analyses and the summary of results have been

reported elsewhere and appear here only as support in Appendices A and B (Finn & Achilles, 1990; Nye et al., 1991, 1992, 1993; Word et al., 1990).]

Essentially these analyses favored the (S) condition (K-3) in all situations and in all years over the (R) and (RA) conditions in regard to pupil achievement. In K there was essentially "no difference" between (R) and (RA), and in grades 1-3 the pupils in (R) performed some better than the pupils in (RA), but the differences between (R) and (RA) results were less than the differences between (S) and other conditions. The LBS analyses followed pupils who had been in (S,R,RA) and for grades 4-6 the basic findings were the same as for K-3, but not quite as pronounced (Nye et al., 1991, 1992, 1993). Analyses are not complete for grade 7. Folger and Breda (1992) and Johnston (1992) did some follow-up analyses on the Teacher Aide Puzzle [a puzzle because intuitively one would expect pupils in the (RA) condition to outperform pupils in the (R) condition], and also reported results of a review of some teacher and aide interviews.

Some Clues from Other Analyses

Johnston (1992) reporting on an analysis of over 1000 interviews collected at year-end (1986, 1987, 1988, and 1989) from STAR teachers noted that "three generalizations emerged and served to organize our findings" (p. 5). Teachers who had a full-time aide or a small class reported "perceptions of: (1) increased time available for teaching and learning, (2) increased opportunity to individualize instruction, and (3) positive differences in the physical, social, and emotional classroom environment" (p. 5). However, in his treatment of the interview data Johnston (1992) did not make clear distinctions between the findings for the (S) and for the (RA) conditions. Yet, it is clear that the (RA) and the (S) conditions both improved the classroom environment and teaching opportunities.

Researchers in STAR interviewed the 50 "Most Effective" teachers based on gain scores (pupil average gain on reading scores from the end of one grade to the end of another grade). Table 1 shows that the largest group of teachers in the "Top 50" (n=23 or 46%) was the (S) condition in the pupil range of 13-17. The next largest group (n=12 or 24%) was the RA group,

suggesting that, for this analysis at least, the (RA) condition resulted in good student achievement.

Table 1. Interviews of 50 "Most Effective" STAR Teachers (Based on Gain Scores)

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, 1985-89.

Teacher Aides and Pupil Achievement

Slavin (nd) summarized clearly the initial findings in Project STAR related to pupil achievement, noting that the results seldom reached statistical significance when comparing (R) and (RA) class results, and that the effect size (ES) was never large enough to consider the results "educationally significant."* Several STAR analyses (e.g., Folger & Breda, 1992; Johnston, 1992; Word et al., 1990) reported information related to the (RA) condition. Generally, results suggested that there was little benefit to student achievement when (RA) and (R) results were compared, but that there seemed to be some benefits to teachers and in terms of classroom "climate." (More on this later.) Slavin (nd) noted that "While research does not generally support the use of aides as they are most often used, there is evidence that aides implementing structured one-to-one tutoring. . .can. . .increase the achievement of. . .first graders" (p. 10). With this in mind, plus the nagging, common-sense belief that the use of teacher aides should be accompanied by increases in pupil achievement, the researchers returned to Project STAR, to the LBS, and to a review of other studies on the use of teacher aides. Because in K there was essentially "no difference"

* Effect size can be computed as $\bar{x}$ of X minus $\bar{x}$ of 0 ÷ SD (pooled). Alternatively, if one expects that the treatment will influence the SD of X (and thus the pooled SD), one might consider ES = $\bar{x}$ of X minus $\bar{x}$ of 0 ÷ SD of 0. While statistical significance is greatly influenced by (n), ES is not; some people discuss ES as "educational signficiance."

between results of (R) and (RA) and due to some procedural issues, STAR researchers reassigned some (R) and (RA) pupils at random in Grade 1. Most STAR analyses are (S) v. (R+RA) or (S) v. "other" in the Final Report. The analyses here look "inside" the basic and prevously-reported analyses.

Review of STAR Findings. In STAR, pupils and teachers were randomly assigned to S,R,RA conditions. Replacement for students who moved or additions of new students were also random but classes designated as S,R,RA remained as designated for purposes of analysis for K-3. This provided a potential for classes to get "out of range." That is, in time a (S) class might have more than 17 (the stated maximum for S) or (R) and (RA) classes might shrink to fewer than 22 (the stated minimum for R and RA classes). A post hoc review of the frequency distributions of classes by numbers of pupils for STAR, K-1-2-3, is in Table 2. This shows that the (S) class might have more than n=17 and continue to be treated as (S) for analysis purposes and that the (R) or (RA) classes could have fewer than n=22 and still be counted as (R) or (RA) for analyses. In moving "out of range" (S) would get larger and (R) and (RA) would get smaller.

Table 2 about here

Note that all (S) classes in K were "in range" and that a few (R) and (RA) classes were "out of range," but that most out-of-range K classes were a teacher-pupil ratio of 1:21. For the (S) condition in grades 1, 2 and 3 there was some "drift" so the (S) condition included classes with 18, 19 and even 20 pupils. For a quick check to see if the "out of range" phenomenon might influence the achievement results, a correlation (Pearson product moment) was run for class size and for the class average reading and math achievement scores for the (S) and the (R) classes only. A negative correlation would show that as class size increased, test scores decreased (Table 3). Given the STAR findings (Word et al., 1990) the negative correlations are not surprising. These correlations were not corrected either for "out-of-range' classes or by removing students who may have entered STAR by retention in grade or who may have had only one year of STAR exposure.[1]

Table 3. Correlations of class size with reading and math (SAT) scores, K-3, Project STAR (1986-1989) using (S) and (R) classes only.

	Grades			
	K	1	2	3
Reading	-.19**	-.27**	-.23**	-.23**
Math	-.14*	-.26**	-.18**	-.18**

Significance *p ≤ .05, **p ≤ .01

This consistent negative correlation suggests that the true picture of differences between (S) and (R) or (S) and (R+RA) or (R) and (RA) were probably understated in the original STAR analyses, or may have even been slightly skewed to the advantage of (R) over (RA). Thus, for the present study certain STAR achievement analyses were recomputed, often without the "out of range" classes. Additional analyses were employed to try to determine more about the teacher aide puzzle.

Table 4 shows the correlation coefficients for the (RA) classes by size (ranges are in Table 2) for pupil achievement in math and reading. These correlation coefficients are neither as large (or significant) nor as consistently negative as those for the classes without the teacher aides.

Table 4. Correlations of class size with reading and math (SAT) scores, K-3, Project STAR (1986-1989) using (RA) classes only.

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

Significance *p ≤ .05, **p ≤ .01

Table 5 presents the STAR class average reading and math SAT scores and the percentile ranks for classes (S,R,RA) for grades K-3. These are the scores of all eligible pupils each year and there was no attempt to analyze further to determine if such things as number of years in STAR influenced the results or if one group or another (e.g., White or minority) seemed to benefit more

than another. Data in Table 5 generally show the (S) class benefit. The (R)/(RA) differences are less pronounced and often "mixed," showing almost no difference in K, some small advantage to (RA) over (R) in grades 1 and 2, but a small advantage to (R) over (RA) in grade 3.

Table 5 about here

In Table 6 data are displayed not by class average, but by pupil average based on class-type placement in (S,R,RA) for grades 1-3 for SAT reading and math and by whether the pupil entered STAR in K or in grade 1. In 1985-86 when STAR began Tennessee did not require K, so a rather large new cohort (n=2278) entered STAR in 1986-87 in first grade. On the surface this analysis shows several trends. Scores in grades 1-3 are 8 to 19 points higher for pupils who had the K experience for all class types (S,R,RA) except for Black students in the Regular with Aide class condition (RA) in grade 1 (no difference) and in grade 3 (only 4 point difference). The RA condition seems to operate some differently for Black than for White pupils. Table 7 shows differences in summary form between pupil achievement in (S) and in (RA) by Race (BL/WH) in grades 1, 2 and 3 for those pupils who entered STAR in K and stayed from 1985-1988.

Table 6 about here

Table 7. Differences in White (WH) and Black (BL) pupil scores in (S) and (RA) classes, grades 1-3, for pupils who entered STAR in 1985-86 in K. (See Table 6 also.)

	Grade								
	1			2			3		
	WH	BL	Diff	WH	BL	Diff	WH	BL	Diff
S	543	514	29	604	575	29	633	613	20
RA	539	493	46	602	566	36	628	604	24
Diff	4	21		2	9		5	9	

While there seems to be little difference to a White pupil (entering STAR in K) by being in (S) or (RA) -- scores favor (S) from 2-5 points from grades 1-3 -- the Black pupil is more disadvantaged by being in (RA) when compared to (S) and especially so in grade 1 where the pupil in (RA) in K and 1 is 21 points below the pupil in (S) in K and 1.

Tables 8 and 9 show the impact of class size (S) and (RA) on students who were reatined in grade and those who were not retained. The importance of these comparisons resides in the ract that some people seem to believe that retention in grade may benefit pupils; others seem to believe that one path to remediating pupil "deficiencies" on standardized or other tests is to put them into small classes (e.g., into Chapter 1 pull-out classes) where pupils are grouped homogeneously more by inability than by anything else. The widespread use of Teacher Aides (TA) in Chapter 1 suggests that, at least in the minds of those who advocate and use TAs, their use should help students achieve better and, perhaps, reduce their deficiencies as measured by the tests.

Data in Table 8 show that those retained and entering STAR in K as retainees (although a year older than those not retained) score considerably lower than the non-retained group at all grades (K-3) in both reading and math in all three class conditions (S,R,RA). Results within the groups are not as consistent. Among the retained grop (n=253) the ANOVAs never reach significance among scores of (S)(R)(RA) and, unlike the general results for STAR, pupils in (S) ao not do better than pupils in (R) and (RA) -- in fact, although statistically N/S, the (R) group scores highest, and the (RA) group seems to be second. Among the non-retained group (n=6041) pupils in (S) do best in all situatios except reading, Grade 3. All ANOVAs reach significance ($p \leq .01$), but this is due to the large n in each comparison. Generally, the (S) condition has highest scores, followed by (R) and then (RA), a finding observed in the STAR multivariate analyses (Word et al., 1990). The "fade" in grade 3 is similar to the findings for some Headstart evaluations.

Data in TAble 9 provide the CRT results for the BSF (a test tied closely to the Tennessee curriculum objecitves) for the pupils who enter STAR in K (1985) as retainees from K (1984-85) and as not-retained pupils. Results parallel those provided for SAT in Table 8 except that for

retained pupils the (RA) pupils score highest, followed by (R) and then (S), although the differences do not reach significance ($p \leq .05$). The trend is different for those not previously retained, with (S) consistently highest and (RA) and (R) following. Pupils previously retained seem to do better in the (RA) condition than in (R) or even (S). However, those who were previously retained score lower than non-retained pupils, and they never narrow the gap by being in the (S) condition. The (S) condition does not serve a remedial function for retained pupils. If begun in K, a full-time aide does seem to help the achievement of once-retained pupils.

Data in Table 10 review the achievement of retained or non-retained entrants to K by race (black or white) by placement into (S), (R), or (RA) in grades 1 and 2 on the CRT for reading and math. Some key points seem consistent: the retained pupils are constantly below the non-retained pupils, white pupils usually outscore black but these differences are much less in the (S) condition for non-retained pupils and in the (RA) condition for retained pupils than for the other conditions in each comparison; for non-retained pupils the (S) condition is superior to the other conditions, but for retained pupils this is not always true.

Tables 11, 12 and 13 present the same data for grade 1 entrants as Tables 8, 9 and 10 present for K entrants to STAR. Generally, the results are quite similar for grades K and 1:

- Non-retained puipls outperform retained pupils on all tests.
- (S) is best for non-retained, followed by (RA); (RA) often is best for retained, followed by (S).
- (S) does not serve to remediate the performance of retained pupils.

The BSF (CRT) data presented in Table 12 show (with few exceptions) the (S) advantage generally for both retained and non-retained pupils in all grades in both reading and math, the constant advantage of non-retained over retained pupils, and the lack of any remediation effect of retention and of (S) for retained pupils. In most comparisons for both retained and non-retained pupils the (S) condition provided the highest outcomes, followed by (RA) and then (R).

Table 13 provides the picture of class type (S,R,RA) impact for black and for white pupils in grades 1-3. There are mixed results for (S) and for (RA) -- in some cases pupils in (S) did better than pupils in (RA) and in other cases this trend was reduced. Two trends are clear: retained

pupils do progressively worse each year compared both to their prior scores and to non-retained pupils each year and the (S) condition does not serve a remedial purpose.

Some Tentative Conclusions

Achievement data from groups of pupils in STAR have been disaggregated by a variety of pupil facts (e.g., race, retained or not retained, grade of entry to STAR as either K or 1) and by the class-type (S,R,RA) palcement experienced by the pupil while in STAR as an initial step to explore the relative benefit of a full-time teacher aide in the "regular" class. (Note that while this analysis looked at groups or categories of pupils, the basic STAR multivariate analyses were based on class averages. In the present analyses, the results are primarily treated as descriptive and heuristic, not as answers. The present data will be used for future and expanded analyses.)

Although this material focused on the value of a teacher aide in a regular classroom (RA) in Project STAR -- or with the STAR pupil database -- the comparisons identified additional issues that may need future exploration. First, however, let's review the (RA) condition and fit this piece into the Teacher Aide Puzzle.

In nearly all comparison explored in the present study the (S) condition resulted in the highest pupil test outcomes followed by the (RA) and then the (R) condition. There was one notable exception: pupils retained in grade before entering STAR (either in K or in 1) benefitted most in their test scores from the (RA) condition and often least from the (S) condition. (Note: This finding supports other work that shows a) that retention isn't a useful way to help students improve on test outcomes and b) that the (S) intervention is not a way to remedy already-defined test-score deficits after pupils have begun schooling in "regular" classes, but that the (S) condition may be a valuable tool in preventing large deficits.)

More of the Teacher Aide Puzzle is to study aspects of teacher aides in the classroom other than just pupil performance. Some work has been begun (e..g, Chase & Mueller, 1993; Folger & Breda, 1992; Green, 1985; Johnston, 1992; Knox County, 1991; Slavin, n.d.), and the researchers are using this database as a starting place for future work with the STAR data relating to teacher aides.

Notes

[1]As STAR moved from K to 3, pupils who had been retained in grade in a prior year entered the STAR database. STAR began in 1985-86 in K and in 1986-87 the 1985-86 pupils (in not retained or mobile) were in grade 1. Note, however, that 1985-86 grade 1 pupils who were retained in grade 1 now (1986-87) entered STAR and were distributed at random to S,R,RA. Was this important?

In 1985-86 Tennessee did not have mandatory K. In 1986-87 there were 2277 new entrants to STAR. Of these new entrants, 1152 were 6.9 years old or older as of 10/1/86 and were presumed to be retainees as they were at least one year older than "normal" dud to Tennessee rules for school entry. This seemed confirmed based on known retainees (n=253) entering STAR in K. Of this group 242 or 96% were 5.9 or older at 10/1/85.

Author Notes

We would like to thank those who have conducted studies on teacher aides and provided part of the database for this review. Dr. Robert Slavin made available prepublication materials. From the STAR project, Dr. John Folger and Dr. John Johnson conducted analyses of STAR data with particular attention to teacher aide results. We thank all those who contributed to and worked on Project STAR and on the LBS effort. Craig Leviner, Reidsville, North Carolina shared an initial literature review and about teacher aides. A special thanks to Dr. Jeremy Finn, consultant to STAR, for design and analysis help, to DeWayne Fulton and Dr. Jayne Zaharias for massaging the STAR and LBS databases, to Dr. John Folger and his colleagues who conducted the training study for STAR, and to others who offered suggestions on the drafts of the material.

Authors

C.M. Achilles, Professor, Education Leadership & Cultural Foundations, School of Education, UNC Greensboro, 27412-5001; B.A. Nye, Director, Center of Excellence in Basic Skills, College of Education, Tennessee State University, Nashville, TN 37203-3401; Helen Pate-Bain, former Principal Investigator, Project STAR, and Jayne Zaharias, and DeWayne Fulton, Research Associates, Center of Excellence, Tennessee State University.

References

Chase, C.I., & Mueller, D.J. (1993, Spring). Teacher aides: What do they really do in instruction? ERS Spectrum, 16-20.

Finn, J.D. (1989, Summer). Withdrawing from school. Review of Educational Research, 59(5), 117-142.

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., Achilles, C.M., Bain, H.P., Folger, J., Johnston, J., Lintz, M.M., & Word, E. (1990). Three years in a small class. Teaching and Teacher Education, 6(2), 127-136.

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

Finn, J., Zaharias, J., Fulton, D., & Nye, B. (1989, Fall). Carry-over effects of small classes. Peabody Journal of Education, 67(1), 75-84. (Published in 1992).

Folger, J. (1989, Fall). Project STAR and class size policy. Peabody Journal of Education, 67(1). (Published in 1992).

Folger, J., & Breda, C. (1990, April). Do teacher aides improve student performance? Lessons from Project STAR. Paper presented at AERA, Boston. (A revision appeared in The Peabody Journal of Education, 6(1), 17-33. Fall, 1989 but published in 1992).

Green, L.L. (1985, October). Studies on the effectiveness of teachers' aides. Mimeo. Working paper for Project STAR. Nashville, TN: Vanderbilt University.

Johnston, J.M. (1989, Fall). Teacher perceptions of changes in teaching when they have a small class or an aide. Peabody Journal of Education, 6(1), 106-122. (Published in 1992).

Knox County (TN) Schools. (1991, January). Using instructional aides as an alternative to reducing class size. Knoxville, TN: Author. Mimeo. 19 pp.

Nye, B., Zaharias, J., Fulton, D., Achilles, C.M., & Hooper, R. (1991). The lasting benefits study: Technical report. Nashville, TN: Tennessee State University Center for Excellence. (Also reports for 1992 and 1993 for grades 5 and 6).

Robinson, E.G., & Wittebols, J.H. (1986). Class size research: A related cluster analysis for decision making. Arlington, VA: Educational Research Service, Inc.

Slavin, R.E. (N.D.). School and classroom organization in beginning reading: Class size, aides, and instructional grouping. Draft material prepared as book chapter.

Word, E., Johnston, J., Bain, H., Fulton, D., Zaharias, J., Lintz, M.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.

Table 2. 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	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

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

Table 5. Class average scores and percentile ranks on SAT math and reading by class type (S,R,RA) for STAR, Grades K-3.

Grade and Class Type	SAT Scaled Score (Rounded) READ	MATH	Percentile Rank READ	MATH
K (1986)				
S	441	489	59	66
R	435	482	53	61
RA	436	482	54	61
1 (1987)				
S	530	539	64	59
R	513	526	53	48
RA	520	530	58	51
2 (1988)				
S	590	585	61	76
R	579	576	52	68
RA	581	577	54	69
3 (1989)				
S	622	624	62	76
R	613	616	55	69
RA	612	614	54	68

Source: Word et al., 1990, pp. 183-184. Results computed by class average and include all pupils/year, regardless of number of years in STAR.

Table 6. Rounded SAT average reading scores (Grades 1-3) of pupils* entering STAR by race in Grade 1** by class-type placement (S,R,RA) compared to those entering STAR in K.

	Grade 1 Score Enter in			Grade 2 Score Enter in			Grade 3 Score Enter in		
	K	1	(Diff)	K	1	(Diff)	K	1	(Diff)
White (S)	533	517	16	599	580	19	627	610	17
White (R)	543	524	19	604	585	19	633	619	14
White (RA)	539	523	16	602	585	17	628	612	16
Black (S)	495	487	8	566	553	13	608	591	17
Black (R)	514	498	16	575	566	9	613	596	17
Black (RA)	493	493	0	566	556	10	604	600	4

*Data are by pupil averages (total) and not by class average.

**Of 2276 grade 1 STAR entrants, 1152were 6.9 years of age at 10/1/86. At this age they were likely (.96) to be retainees in grade 1 from 1985-86 school year. This estimate is based on Tennessee school entry rules and the fact that of the known K retainees entering STAR in 1985-86 (n=253) 242 or 96% were 5.9 or older at 10/1/85.

TAble 8. K-3 reading and math SAT scores (rounded) of pupils retained into STAR (n=253) in K (1985-86) and pupils not retained but entering STAR in 1985-86 (n=6041) by class type (S,R,RA).

	K				1				2				3				GAIN, K-3		
	READ		MATH		READ		MATH		READ		MATH		READ		MATH		READ	MATH	Survive**
	n	x	n	x	n	x	n	x	n	x	n	x	n	x	n	x			*K 3
Retained (n=253)																			
S	59	422	61	475	45	485	49	503	34	549	33	543	18	587	17	594	163	119	30
R	93	427	93	472	63	496	76	508	50	557	50	556	37	607	36	607	180	135	40
RA	76	421	77	466	41	487	47	503	35	552	36	546	20	604	20	603	183	137	26
TOT	228		231		149		172		119		119		75		73				32
ANOVA (F)	1.87		1.04		.74		.33		.34		1.20		1.52		.79				
PROB ≤	.16		.35		.48		.72		.71		.30		.23		.46				
Not Retained (n=6041)																			
S	1673	441	1694	492	1292	536	1319	543	1027	599	1023	594	886	603	898	631	162	139	53
R	1906	435	1932	484	1393	525	1415	533	1112	594	1111	590	964	623	971	626	188	142	50
RA	1959	436	1991	483	1460	524	1502	532	1106	591	1104	587	960	623	976	626	187	153	49
TOT	5538		5617		4145		4236		3245		3238		2810		2845				51
ANOVA (F)	18.9		16.6		18.9		24.6		7.5		7.9		11.3		5.5				
PROB.	*		*		*		*		*		*		*		*				
Diff.	1 vs 2&3		1 vs 2&3		1 vs 2&3		1 vs 2&3		1 vs 2&3		1 vs 2&3		1 vs 2&3		1 vs 2&3				

All comparisons for retainees are NS. Patterns of scores show that (S) does not seem to offer a remediation. The n=253 students were known retainees in K (1984-85) and entered STAR (1985-86) as retainees. Data track the 253 who took tests in STAR (K-3). Reduction in total taking test could be retention, a move, special educaion placement, absence, etc. Continuation rate in STAR by class type using largest number/year of test takers is (S) [61 to 18 or 30%], (R) [93 to 37 or 40%], and (RA) [77 to 20 or 26%] from K to 3 of those entering STAR in K as retainees from 1984-85. Other K entrants were not retainees (6041).

*p ≤ .01
**Computed on the largest n tested at grades K and 3.

Table 9. BSF percent (rounded) of items correct by grade (1-3) for condition (S,R,RA) by prior retention and no retention for entry into K, STAR, 1989.

	GRADE 1				GRADE 2				GRADE 3			
	READ		MATH		READ		MATH		READ		MATH	
Retained (n=253)	n	% Pass	n	% Pass	n	% Pass	n	% Pass	n	% Pass	n	% Pass
S	39	70	39	76	38	65	38	74	29	67	29	70
R	48	69	49	79	38	68	39	78	26	71	25	80
RA	44	73	45	83	43	67	44	81	32	74	33	75
TOT	131	70	133	80	119	67	121	78	87	71	87	75
Sig.		.58		.33		.84		.28		.42		.24
Not Retained (n=6041)												
S	1208	88	1202	92	1228	87	1247	90	1103	85	1101	88
R	1161	84	1153	89	974	85	987	89	736	84	735	87
RA	1094	85	1091	90	1072	86	1093	90	987	84	986	86
TOT	3463	86	3446	90	3274	86	3327	90	2826	84	2822	87
Sig.		.00*		.00*		.00*		.00*		.05*		.08*

Difference (Retained vs Not Retained) by Class Size)

	%	%	%	%	%	%
S	18	16	22	16	18	18
R	15	10	17	11	13	7
RA	12	7	19	9	10	11
TOT	16	10	19	12	13	12

*Probably heavily influenced by the large n.

Table 10. Percent items correct on BSF criterion-referenced tests in Grades 1 and 2 of pupils entering STAR as retainees in K (n=253) and of pupils entering STAR in K as new (n=6041) by Race (BL/WH) by class type (S,R,RA).

	GRADE 1						GRADE 2					
	READ			MATH			READ			MATH		
	S	R	RA	S	R	RA	S	R	RA	S	R	RA
Retained (n=253)												
BL	71	63	71	62	67	79	62	54	67	72	62	75
WH	70	71	73	79	84	84	82	72	67	86	82	83
DIFF	1	8	2	17	17	5	82	18	0	87	20	8
Not Retained (n=6041)												
BL	87	77	80	89	84	85	87	79	79	92	85	86
WH	88	87	87	93	92	92	20	87	88	15	90	91
DIFF	1	10	7	4	8	7	7	8	9	6	5	5
DIFFERENCES OF RETAINED VS NOT RETAINED BY RACE*												
BL	16	14	11	27	17	6	-	25	12	1	23	11
WH	18	16	14	14	8	8	25	15	21	20	8	8

*All differences favor the "not retained" group. The n's are possible; actual numbers who took tests and/or remained with the cohort decrease each year.

Table 11. Reading and math SAT scores (rounded) for grades 1-3 of pupils entering STAR at grade 1 either as retainees (n=1152) or as new and without STAR K (n=1124)

	GR1				GR2				GR3				Survival**	Gain	
	Read		Math		Read		Math		Read		Math		% 1-3	Read	Math
	n	x	n	x	n	x	n	x	n	x	n	x			
Retained (n=1152)															
S	146	501	153	523	96	563	95	567	65	596	66	599	43	95	76
R	472	499	505	518	336	555	339	558	234	591	238	596	47	92	78
RA	405	506	438	523	297	561	296	561	228	596	236	599	54	90	76
TOT	1023	502	1096	521	729	558	730	560	527	594	540	597	49	92	76
(F)	2.70		2.66		2.67		1.64		1.75		.50				
Prob.	.07		.07		.07		.19		.18		.60				
Sig.*	N/A		R/RA		R/RA		N/A		N/A		N/A				
Not Retained (n=1124)															
S	199	523	202	531	113	597	113	586	94	627	96	626	48	104	95
R	459	507	466	520	251	583	249	575	186	622	188	624	40	45	104
RA	400	518	408	526	243	594	242	585	184	623	187	625	46	105	99
TOT	1058		1076	524	607	590	604	581	464	623	471	625	44		101
(F)	8.11		5.64		5.19		4.20		.77		.03				
Prob.	.00		.00		.00		.01		.46		.97				
Sig.*	RA/R S/R		RA/R S/R		RA/R S/R		RA/R		N/A		N/A				

*Pairs of groups sig. at $p \leq .05$. Essentially, comparisons for retainees are not significant and tend to show no particular benefit of (S) for retainees. Comparisons for new, non-retained students tend to favor (S) and/or (RA) over (R).

**Computed on the largest n tested at grades 1 and 3.

Table. 12 BSF percent (rounded) of items correct by grade 1-3 for condition (S,R,RA) of pupils retained into 1st grade and entering grade 1 with no retention, STAR, 1989.

	GRADE 1				GRADE 2				GRADE 3				SURVIVAL*
	READ		MATH		READ		MATH		READ		MATH		% 1-3
Retained (n=1153)	n	% Pass	n	% Pass	n	% Pass	n	% Pass	n	% Pass	n	% Pass	
S	154	79	151	88	136	75	138	85	123	70	123	76	49
R	481	76	480	86	307	73	313	82	198	71	199	74	41
RA	438	78	435	86	314	72	324	81	255	71	258	74	59
TOT	1073	77	1066	86	757	73	775	82	576	71	580	74	54
Sig.**	.08		.23		.33		.01		.75		.51		
Not Retained (n=1124)													
S	194	85	190	91	144	86	145	90	138	84	137	86	71
R	455	80	454	86	213	82	212	87	136	83	136	85	30
RA	389	83	388	88	259	85	264	89	202	83	200	87	52
TOT	1038	82	1032	88	616	84	621	87	476	83	473	86	46
Sig.**	.00		.00		.01		.03		.51		.70		

Difference of Retained vs Not Retained by Class Size

	%	%	%	%	%	%
S	6	3	11	5	14	10
R	6	0	9	5	12	11
RA	5	2	13	8	12	13
TOT	5	2	11	5	12	12

*Computed on the largest n tested at grades 1 and 3.
** ANOVAs.

Table. 13 Percent items correct on BSF grades 1-3 criterion-referenced test of pupils entering STAR as retainees (n=1153) in grade 1 and of entering STAR, grade 1 without STAR K (n=1124) by Race (BL/WH) and by class type (S,R,RA).

	GRADE 1						GRADE 2						GRADE 3					
	READ			MATH			READ			MATH			READ			MATH		
	S	R	RA	S	R	RA	S	R	RA	S	R	RA	S	R	RA	S	R	RA
Retained (n=1153)																		
BL	76	72	74	85	82	81	73	68	63	85	80	74	63	68	64	69	70	67
WH	81	78	80	90	88	89	76	77	77	85	83	84	74	73	73	81	77	77
DIFF	5	6	6	5	6	8	3	9	14	0	3	10	11	5	9	12	7	10
Not Retained (n=1124)																		
BL	82	76	80	87	82	84	82	76	81	87	83	86	79	78	80	81	80	83
WH	88	83	85	93	89	91	89	87	88	91	90	91	87	85	86	88	89	89
DIFF	6	7	5	6	7	7	7	11	7	4	7	5	9	7	6	7	9	6
Differences of Retained vs. Not Retained by Race																		
B	6	4	6	2	0	3	9	8	18	2	3	12	16	10	16	12	10	16
W	7	5	5	3	1	2	13	10	11	6	7	7	13	12	13	7	12	12

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

Source of Variation:	*Error Term:*
Location (L)	Schools
Training* (TR)	Schools
Type (T)	School × type
L × T	School × type
L × TR	School
T × TR	School × type
L × T × TR	School × type

		Degrees of Freedom (*df*)		
		Ach. Meas.		Noncog. Meas.
Schools	e.g. (1986)	75		69
School × Type	e.g. (1987)	149		137
Classes within School-Types (etc.)			Etc.	
Primary Model; Measures				Matched *t*-tests
Achievement (Ach.):	SESAT, SAT, BSF			
Noncognitive (Noncog.):	SCAMIN, attendance, behavior, etc.			
Extended Model; Measures:				
Sex (or race, or SES)	Ave. diff. scores on Ach.			Multivariate models
Sex (or race, or SES)	Ave. diff. scores on Noncog.			
Training*				

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 × TR and L × 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 materials and conference papers by the same authors.

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

Achilles, Nye, Zaharias, Fulton (1993, Nov).
Creating Successful Schools for All Children: A Proven Step 619
Sch. of School Leadership, 3(6), 617-619

APPENDIX B

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

		Reading		Mathematics			
Effect/ Grade[a]		Multivariate[b]	SAT[c] Read	BSF Read	Multivariate[b]	SAT Math	BSF Math
Location (L)			.02			.05	
	1	.01	.06		.05		
	2	.001	.001	.001		.001	.001
	3	.001	.001	.001	.001	.001	.001
Race (R)	1	.001	.001	.001	.001	.001	.001
	2	.001	.001	.001	.001	.001	.001
Type (T)	K		.001			.02	
	1	.001	.001	.001	.001	.001	.05
	2	.001	.001	.05	.001	.001	.05
	3	.001	.001	.001	.001	.001	.001
SES	K		.001			.02	
Loc × Race	1	.05		.05			
Loc × Type	K-3	All N/S. The class-size effect is found equally in all locations—inner city, suburban, urban and rural schools (tabled as important).					
Race × Type	1	.05	.05	.01			
L × R × T	1			.05			.01
L × TR × T	2	.05	.01	.05	.05	.05	.01

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.

[a]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 × race tested eliminating main effects and loc × type; loc × type tested eliminating main effects and loc × race; race × type tested eliminating main effects and other two-way interactions, and loc × race × type tested eliminating all else (Finn and Achilles, 1990).

[b]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 Aide effect) is found equally in all four locations for trained and untrained teachers.

[c](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.