Class Size and Students At Risk

What is Known? What Is Next?

A Commissioned Paper
prepared by
Jeremy D. Finn

National Institute on the Education of At-Risk Students
Office of Educational Research and Improvement
U.S. Department of Education

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Institute on the Education of At-Risk Students
Edward J. Fuentes
Director

Media and Information Services
Cynthia Hearn Dorfman
Director

April 1998

This report was prepared for the National Institute on the Education of At-Risk Students, Office of Educational Research and Improvement (OERI), U.S. Department of Education. The conclusions and interpretations in this report are those of the author alone and do not necessarily reflect the findings or policies of the Institute, OERI, or the U.S. Department of Education.

Foreword

This report is an overview of recent research on the effects of class size on the academic performance and behavior of students at risk. In several ways, it is not a conventional literature review. It emphasizes one recent large-scale investigation, Tennessee's Project STAR (Student-Teacher Achievement Ratio). It is more evaluative than most reviews of research, emphasizing the strengths and weaknesses of the studies cited. It stresses the need for future research more than the typical literature review.

All of these features were requested by the U.S. Department of Education's Office of Educational Research and Improvement (OERI) in the work statement for the preparation of this report. The purpose of the review is to advise OERI, particularly the National Institute on the Education of At-Risk Students, on the implications of recent research on class size for the

design of early educational interventions for at-risk students, for education policy, and for further research on class size. The first of four stated priorities was to address the question "How conclusive is this research?" Beyond the review of research on small classes, the work statement called for a discussion of "approaches that can be taken to assess the costs and benefits of reducing class size" and "the implications of small class size for classroom management and instructional strategies." The final task was to address "the implications of these findings for future research on class size" and to identify "some key questions that should be investigated."

The request to evaluate the conclusiveness of the findings was right on target. Many school districts and states are currently undertaking some form of small-class initiative, with substantial expenditures of money and effort. In other words, the implications of this research for guiding school policy are profound and there must be certainty that the costs are warranted. At the same time, Project STAR is unique in its design and magnitude. Unlike most educational research, it has the ability to provide tight cause-and-effect conclusions. We are in the unusual position of being able to evaluate a practice that appeals to many educators and which may have a tangible impact on the academic performance of students.

Although STAR provides some answers about the effectiveness of small classes, to date it provides only hints about other related questions. The key questions that remain include the long-term consequences of attending a small class, the interactions of instructional processes with class size, and the particular impact of small classes on students at risk. Other past and current studies provide some answers to these questions and more than a few hypotheses. Yet there remains a tremendous amount of work to be done. The extensive research agenda given in this report was developed, not only because it was requested by OERI, but because it identifies a large number of unanswered but pressing educational concerns.

About the Author

Dr. Jeremy Finn has conducted extensive research on students at risk and engagement in school, and served as an external evaluator for Project STAR and the STAR follow-up studies. He received his Ph.D. in Educational Measurement, Evaluation, and Statistical Analysis from the University of Chicago and currently is a Visiting Scholar at the Center for Research in Human Development and Education at Temple University. Dr. Finn's regular position is Professor of Education at State University of New York at Buffalo.

Acknowledgments

The author is grateful to Patrick Harman and Jayne-Boyd Zaharias for reaction to an earlier version of this report, and especially to Charles Achilles for ideas, reviews and comments throughout the development of the report. In addition, many current and former National Institute on the Education of At-Risk Students staff members contributed to its inception and development, including Judith Anderson, David Mack, Susan Talley, Oliver Moles, Patricia Dabbs, Ron Pedone and especially D. Hollinger Martinez. Gerry Devlin, Bob LeGrand, Phil Carr and particularly Chris Sheedy of OERI's Media and Information Services also provided invaluable editorial and production assistance.

Contents

Introduction

It is a propitious time with regard to questions about class size. After years of debate, speculation, and research that yielded only partial and less-than-definitive answers, a major longitudinal study provides answers to the question "Do small classes result in greater academic achievement in the elementary grades?" Begun in 1985, Tennessee's Project STAR (Student-Teacher Achievement Ratio) set the stage for asking and answering a number of policy questions that could not be addressed before.

The first chapter of this paper reviews the status of research on class size with particular attention to the STAR investigation and to the research spawned by STAR. The conclusiveness of the findings are discussed as well as implications for students at risk and for education policy in general. The second chapter discusses approaches that have been taken to

assess the costs and benefits of reducing class size and proposes additional dimensions that need to be considered; and the third explores the implications of small class size for classroom management and instructional strategies, with particular attention to the need to increase the academic engagement of students at risk. Issues requiring further research are identified throughout the paper. In the last chapter, however, these are summarized as a "research agenda" with priorities for further work.

This review and research agenda focus largely on the effects of small classes in the early grades. There are two reasons for this. First, the most current (and best) research to date has been conducted in kindergarten through grade 3. The state of research with respect to small classes in the upper grades is fragmented and even contradictory, leaving little to say that is based on substantive research results. Second, there are good reasons for starting research and intervention projects in the early grades. The assumption that the early years lay the foundation for much that follows is explicit throughout this review and has been substantiated repeatedly by research in the social sciences.

Research on the Academic Effects of Small Class Size

The question "Are smaller classes better than larger classes?" continues to be debated among teachers (and their unions), administrators, and parents as well as in the research community. The issue persists because of the powerful common-sense appeal of small classes to alleviate problems indigenous to our classrooms. Small classes are an integral component of nationally subsidized programs including special education classes for disruptive or learning-disabled students and Title I interventions for children living in poverty. Small classes or small groups working with one teacher or tutor also are a key element of programs targeted most often at students at risk, for example, Success for All (Slavin, et al., 1990; Slavin & Madden, 1995) and Reading Recovery (Pinnell, deFord, & Lyons, 1988).

The issue persists because of the tension between the research findings and the cost of implementation. A great deal of empirical data have been collected. However, they have so far been less than convincing and not consistent enough to justify the expense of the additional classrooms and teachers that would be required. Targeted remedial programs are generally less costly and easier to deploy. They tend to be adopted for a portion of the school day to address learning problems in one or a small number of subject areas. In contrast, maintaining small classes throughout a grade level or school requires pervasive organizational changes. Of course, proponents would argue that the benefits are also pervasive being realized throughout the school day and affecting the entire range of school subjects unlike the band-aid approach of experimenting with one targeted program after another.

Overviews of Research on Small Classes

Over the past 2 decades there have been many summaries of research on the relationship of class size to academic achievement. Three are particularly worthy of note because of their comprehensiveness, and because they planted the seeds for much of the research that followed.

Without doubt the most widely cited review is the classic Meta-analysis of

research on the relationship of class size and achievement (Glass & Smith, 1978). The authors collected and summarized nearly 80 studies of the relationship of class size with academic performance that yielded over 700 class-size comparisons on data from nearly 900,000 pupils. The two primary conclusions drawn from this material are:

* reduced class size can be expected to produce increased academic achievement (p. iv); and

* [t]he major benefits from reduced class size are obtained as the size is reduced below 20 pupils (p. v).

Although the extensiveness of the Glass-Smith meta-analysis was commendable, the selection of studies to include was subject to justifiable criticism. A number of studies were of short duration; many compared normal-sized classes to one-on-one tutoring; other studies did not include "realistic" class sizes as their comparison groups; and at least one study related to instruction in non-academic subjects (i.e., tennis). In spite of these deficiencies, however, the two conclusions drawn by Glass and Smith have endured and have received further support.

A compilation of studies examined by Educational Research Service (Robinson & Wittebols, 1986; Robinson, 1990) is noteworthy because of its extensiveness more than 100 separate studies were reviewed. Robinson's (1990) conclusions added an important set of qualifications to the findings of Glass and Smith:

> [R]esearch does not support the expectation that smaller classes will of themselves result in greater academic gains for students. The effects of class size on student learning varies (sic) by grade level, pupil characteristics, subject areas, teaching methods, and other learning interventions. (p. 90)

In particular, the review concludes that small classes are most beneficial in reading and mathematics in the early primary grades and that: "[t]he research rather consistently finds that students who are economically disadvantaged or from some ethnic minorities perform better academically in smaller classes" (p. 85). Unfortunately, the wide-ranging review failed to distinguish even the best designed studies from those using the poorest methodology, and thus the conclusions must be viewed as tentative.

A third review is noteworthy because of its focus on high-quality research conducted in accordance with accepted scientific standards. Using a procedure termed "best evidence synthesis," Slavin (1989) reviewed only those studies that lasted a minimum of 1 year; involved a substantial reduction in class size, that is, larger classes were compared to classes that were at least 30 percent smaller and had 20 students or fewer; and involved either random assignment of youngsters to class sizes or matching to assure that the groups were initially equivalent.1

Of the research summarized by Glass and Smith (1978) and others, Slavin identified only eight studies that met all three criteria. From these eight studies, Slavin concluded that substantial reductions in class size have a small positive effect on students (the median effect size for the eight studies was only 0.13[sigma]); and the effect was not cumulative and even disappears in later years.2 Slavin's reinterpretation of the Glass-Smith findings is that large effects are not likely to be seen until the class

size is reduced to one (e.g., one-on-one tutoring).

Other research syntheses. In a brief overview of research, Finn and Voelkl (1994) identified three approaches to studying the issue of class size: the classroom-focus approach, the cost-related approach, and the ecological approach.

The reviews by Glass and Smith (1978), Robinson and Wittebols (1986), and Slavin (1989) summarize classroom-focus studies; this research examined the number of pupils in each classroom, the interactions between the teacher(s) in that classroom, and the outcomes that were realized by the pupils in that classroom. It provides the most direct and intensive view of the effects of a small class setting.

The cost-related approach examines the actual or potential costs of implementing small classes and weighs them against the benefits that may accrue. This approach is discussed in considerable detail in the next chapter of this paper.

The ecological approach views class size in historical or geopolitical perspectives. For example, Tomlinson (1988, 1989) examined the changes in median class size in the United States over several decades and related them to changes in standardized test scores. The analysis does not show performance benefits for smaller classes, and it ignores a multitude of intervening factors, including population shifts and both cultural and institutional changes over the same time period. Likewise, the comparison of class sizes between countries introduces a number of confounding variables including national differences in educational expenditures, educational goals, teacher preparation, and student characteristics, to name a few. Class sizes also may vary dramatically within a country over time or among schools at one point in time (see Finn & Voelkl, 1994). Thus, ecological associations with pupil performance only obscure the effects of having a smaller or larger number of individuals in a particular class setting.

Class size is not pupil/teacher ratio. The analysis of pupil/teacher ratios is characteristic of the ecological approach and shares some of the same difficulties. Although the number of pupils can be compared to the number of teaching staff in a single school, the ratio obfuscates the workload faced by a teacher in one classroom, the amount of attention the teacher gives to any one pupil, and dynamics of a small or large class that may impact on pupil participation;3 these interactions may be especially important for students at risk. At the same time, pupil/teacher ratios are often smaller in urban districts (because of Title I programs, special education programs and remedial teachers), while actual class sizes may be larger. One significant study (Boozer & Rouse, 1995) found that average class size a more direct measure of classroom organization was more important to academic achievement than the pupil/teacher ratio. Although several studies discussed in this paper did examine pupil/teacher ratios, the emphasis is on classroom-focus research.

Statewide Class-size Studies: PRIME TIME and STAR

Indiana's PRIME TIME. In 1984 the state of Indiana funded an initiative to reduce class sizes in grades 1 through 3 to an average of 18 pupils, or to 24 pupils if an instructional assistant was in the classroom. During the initial year, 286 of 303 districts participated to a greater or lesser

extent. The main PRIME TIME intervention took place over 3 years, beginning with grade 1 in 1984, adding grade 2 in 1985, and grade 3 (or kindergarten, on option) in 1986.

The outcomes of PRIME TIME are summarized in numerous publications (e.g., Center for School Assessment, 1986; Chase, Mueller & Walden, 1986; Malloy & Gilman, 1989; McGiverin, Gilman, & Tillitski, 1989; Mueller, Chase, & Walden, 1988). In brief:

* Positive outcomes were found for small classes on such factors as time on task, individualized instruction, well-behaved classes, and teacher satisfaction; but

* The results for academic achievement were mixed at times, small classes were found to have superior outcomes and, at times, the large classes performed better.

Project PRIME TIME is noteworthy because it demonstrates important principles for the research that followed, namely, the feasibility of a statewide class-size initiative and the need to conduct an intervention of this type over a period of years. Virtually all class-size research that preceded PRIME TIME was cross-sectional in nature. However, PRIME TIME was designed as a demonstration project and did not follow rigorous procedures needed for a thorough evaluation in that: no control was implemented to equalize or match smaller and larger classes at the outset; small classes may not have been kept small for the entire school day; different achievement tests were administered in different schools; and other local, state, and federal programs were functioning in some schools but not others simultaneously with the class-size intervention.

More unfortunately, PRIME TIME did not implement a single, well-defined, small-class intervention. While the average class size of 18 pupils was viewed as a target, actual class sizes ranged from 12 to 31; classes of 24 pupils with a teacher aide were considered to be small despite the number of pupils in the classroom. As a result, the evaluations of PRIME TIME cannot be interpreted as confirming or refuting a class-size effect.

Tennessee's Project STAR. Project STAR, the only large-scale, controlled study of the effects of reduced class size, was conducted in 79 elementary schools in the state of Tennessee from 1985 to 1989. The design drew heavily upon previous research findings, namely, that any benefits of small classes are likely to be realized in the primary grades, that there may be different outcomes for students based on race or economic disadvantage, and that only substantial reductions in class size are likely to have noteworthy impact.

Within each participating school, children entering kindergarten were assigned at random to one of three class types: small (S) with an enrollment range of 13 to 17 pupils; regular (R) with an enrollment range of 22 to 26 pupils; or regular with a full-time teacher aide (RA) with 22 to 26 pupils. Teachers also were assigned at random to the class groups. Teachers in the STAR classrooms received no special instructions of any sort, and the duties of teacher aides were not prescribed but were left to the teacher's discretion.4

Classes remained the same type (S, R, or RA) for 4 years, until the pupils were in grade 3. A new teacher was assigned at random to the class each

year. Standardized achievement tests (Stanford Achievement Tests, or SATs) were administered to all participating students at the end of each school year. Also, curriculum-based tests (Basic Skills First, or BSF) reflecting the state's instructional objectives in reading and mathematics were administered at the end of grades 1, 2, and 3. Finally, a measure of motivation and self-concept intended for young children also was administered to each pupil (Milchus, Farrah, & Reitz, 1968). In all, about 7,500 pupils in more than 300 classrooms participated in the 4-year longitudinal study.

Comments on the design. Before reviewing the outcomes of Project STAR, the particular strengths of this initiative should be underscored. The within-school design was an effective way to control for differences among school settings including, but not limited to, the economic status of the student body, per-pupil expenditures, and the manner in which schools were administered. The value of this type of design cannot be underestimated. The random assignment was monitored carefully by state- level evaluators. A large and diverse population of students was longitudinally tracked over the 4 year period, and the data were collected, cleaned, and collated with a high degree of care. Both norm-referenced and criterion-referenced achievement data were collected. The norm-referenced tests, based on item-response theory, permitted comparisons of achievement levels from one grade to the next. The design of STAR, together with its magnitude and the follow-up research conducted after the 4-year period, led Harvard's Frederick Mosteller to term Project STAR "[a] controlled experiment which is one of the most important educational investigations ever carried out" (1995, p. 113).

The primary results. The main analysis of STAR outcomes consisted of four cross-sectional analyses, one at the end of each school year.5 The statistical methods were variations of common confirmatory procedures for evaluating experimental outcomes, for example, analysis of variance, multivariate analysis of variance, and analysis-of-covariance procedures (see Finn & Achilles, 1990). In addition to tests of significance, "effect size" measures were derived each year for all students and for white and minority students separately. The results were compiled into a Tennessee State Department of Education report (Word,et al. 1990).

Four primary results were reported consistently across the 4 years of analysis:

* Differences among the three class types were highly statistically significant for all sets of achievement measures and for every measure individually. In every case, the significance was attributable to the superior performance of children in small classes, and not to classes with full-time teacher aides.

* With only minor exception, there was no significant interaction with school location 6 or sex of the pupil. A significant small-class advantage was found in inner-city, urban, suburban, and rural schools alike and the advantage of small classes was found both for males and females.

  In each year of the study, some of the benefits of small classes were found to be greater for minority students than for nonminorities, or greater for students attending inner-city schools.

* No differences were found among class types on the motivational scales.7

The results are given in the form of small-class effect sizes in Table 1.8 Each effect size is the mean score for small classes minus the mean score of regular and teacher-aide classes [S - (R+A)/2] in standard deviation units. Since they all favor small classes, the researchers referred to the difference as the "small-class advantage." For the criterion-referenced Basic Skills First (BSF) tests, the difference is computed for the percentage of students exceeding the state's mastery criterion.

Table 1.
Small-class effect sizes, grades kindergarten (K) through 3, by skills, motivation, and self-concept data

Scale	Group	Grade Level			
		K	1	2	3
Word Study Skills	W	0.15	0.16	0.11	
	M	0.17	0.32	0.34	N/A
	ALL	0.15	0.22	0.20	
Reading	W	0.15	0.16	0.11	0.16a
	M	0.15	0.35	0.26	0.35a
	ALL	0.18	0.22	0.19	0.25a
Total Reading	W	-	0.17	0.13	0.17
	M	-	0.37	0.33	0.40
	ALL	0.18	0.24	0.23	0.26
Basic Skills First (BSF)	W		4.8%	1.6%	4.0%
Reading	M	N/A	17.3%	12.7%	9.3%
	ALL		9.6%	6.9%	7.2%
Total Mathematics	W	0.17	0.22	0.12	0.16
	M	0.08	0.31	0.35	0.30
	ALL	0.15	0.27	0.20	0.23
Basic Skills First (BSF)	W		3.1%	1.2%	4.4%
Mathematics	M	N/a	7.0%	9.9%	8.3%
	ALL		5.9%	4.7%	6.7%
Motivation	W	0.00	-0.02	-0.03	-0.01
	M	0.03	-0.01	0.07	0.11
	ALL	0.01	0.00	0.01	0.00
Self-Concept	W	0.10	0.07	0.00	-0.05
	M	0.10	0.05	0.03	0.04
	ALL	0.11	0.7	0.02	0.02

NOTE: The values for BSF Reading and BSF Mathematics represent differences in the percent passing (no standard deviation). All other values are mean differences: Small - (Regular + Aide)/2, divided by the standard deviation of the scale. Standard deviations computed for all students in regular classes, and all white (W) and minority (M) students separately.

[sigma]Total Language scale in grade 3 (not Reading).

In every instance, small classes outperformed the other class types; effect sizes for the total sample (All range from about 0.15 [sigma] in kindergarten to about 0.25[sigma] in grades 1, 2, and 3.9 And like the research that preceded STAR, the small-class advantage was consistently greater for minority students (most of whom were black) than for whites. In

most comparisons, the impact on minorities was about twice as large as it was for white students. This resulted in a considerably reduced achievement gap. In reporting this effect, Finn and Achilles (1990) noted that the difference between minorities and whites in mastery rates on the grade 1 reading test was "reduced from 14.3 percent in regular classes to 4.1 percent in small classes" (p. 568).

Two additional points should be noted. First, the effect sizes in Table 1 show that small classes present up to a 1/4[sigma] advantage compared to larger classes in every subject tested.10 Although the researchers did not devise methods for computing the total impact on achievement, it is greater than any single difference would indicate. Second, the effect sizes in Table 1 actually underestimate the true small-class advantage. An unavoidable phenomenon during the 4-year project was the "drifting" of some classes out of the target size range, as students transferred into or out of a class or school. Preliminary indications are that the effect sizes would be substantially greater if out-of-range classes were removed from the data.11

In sum, due to the magnitude of the Project STAR longitudinal experiment, the design, and the care with which it was executed, the results are clear:

* This research leaves no doubt that small classes have an advantage over larger classes in student performance in the early primary grades.

At the same time, the research leaves behind a wealth of data that have only begun to be analyzed for what they can tell us.

The follow-up: the Lasting Benefits Study. After the positive STAR findings, Tennessee authorized a study to see how long the initial benefits of small classes would persist. Although all children were returned to regular-size classes in grade 4, the Lasting Benefits Study (LBS) continued to follow a significant portion of these pupils.12 In the 1995 1996 school year, the majority of STAR students were in grade 10 and were still being tracked.

The grade 4 evaluation included standardized and criterion-referenced achievement tests plus a new measure of student engagement in learning activities, the Student Participation Questionnaire (SPQ) (Finn, Folger, & Cox, 1991). The SPQ is a 28-item scale on which each pupil is rated by his or her teacher. It yields reliable, valid measures of student "effort" that the student allots to learning, "initiative-taking" in the classroom, and "nonparticipatory" behavior (disruptive or inattentive- withdrawn behavior). The grade 4 results (Finn,et al. 1989) showed that, even after the small-class intervention was disbanded:

* Students who had been in smaller classes had higher achievement in all academic areas compared to students in regular or teacher-aide classes;

* The small-class effect size (small to regular) ranged from 0.11[sigma] in social studies to 0.16[sigma] on the criterion-referenced mathematics test; and

* Pupils who had been in small classes were rated as expending more effort in the classroom, taking greater initiative with regard to

learning activities, and displaying less disruptive or inattentive behavior compared to their peers who had been in regular-size classes.

Positive achievement results continued to be obtained in later grades. The median small to regular difference in grade 5 for the total sample was approximately 0.18[sigma] ; in grade 6 it was approximately 0.16[sigma] ; in grade 7 it was approximately 0.14[sigma] 0. As in earlier grades, the differences were statistically significant on all norm-referenced and curriculum-based tests.13

The carry over effects are consistent with findings from other early interventions, for example, the Perry Preschool Project (Berrueta-Clement,et al. 1984). They raise the possibility that small classes in the early grades have significant long-term consequences for all students generally and that they may begin students at risk of educational failure on a positive trajectory that will increase their chances of school success through the years.

As of this writing, resources are not available to explore these data in any but the most cursory ways. The data base continues to grow, however. In grade 8, two teachers rated each student on the SPQ and each student completed a self-report "Identification with School" scale (Voelkl, 1996). Achievement test scores have been obtained for grades 8 and 9. In sum, STAR and the LBS have laid the groundwork for building an important data base for examining educational effects longitudinally. Its potential to address both basic and policy-relevant research issues is elaborated in a later section of this report.

Other STAR-related studies. Based on the positive findings of STAR and the LBS, Tennessee implemented Project Challenge in 17 of the state's poorest school districts, that is, districts with the lowest per capita income and highest percentages of pupils in the subsidized lunch program. Beginning in 1990, small classes (pupil to teacher ratio of 15:1) were introduced in all schools in these counties in the primary grades; grades 2 and 3 in 1990, grades 1 through 3 in 1991, and grades kindergarten through 3 in 1992 and later years. Project Challenge was not a controlled experiment as was Project STAR, but was a thorough effort to implement small classes in particular targeted districts.

The project was assessed through an analysis of district rankings on statewide achievement tests (Achilles, Nye, & Zaharias, 1995). Since Tennessee has 138 districts, a rank of 69 would be considered average. In terms of the mean rankings of the 17 Challenge districts, the results were:

* In grade 2 reading, the mean ranking improved from 99 in 1990 (among the lowest) to 94 in 1991, 87 in 1992, and 78 in 1993; and

* In grade 2 mathematics, the mean ranking improved from 85 in 1990, to 79 in 1991, to 60 in 1992, and 57 in 1993 that is, from performance below the state average in 1990 to performance above the average in 1992 and 1993.

It is also interesting to note that because of the staggered introduction of small classes, grade 2 students in 1991 had been in small classes for just 1 year, whereas the grade 2 students in 1991 had been in small classes for 2 years (grades 1 and 2), and the 1992 and 1993 grade 2 students had been in small classes for 3 years (kindergarten through grade 2). That is:

* Each additional year in the small-class setting was accompanied by further improvement in reading and mathematics.

This study adds non-experimental evidence that small classes are beneficial in the primary grades. The data also indicated that in-grade retentions were reduced when small classes were implemented (Achilles, n.d.).

Two smaller studies of class size were conducted in North Carolina pursuant to STAR. In 1991 educators, citizens, and the school board in Burke County, North Carolina began a project to reduce the class size to 15 in grade 1, followed by grades 2 and 3 in subsequent years (Achilles, Harman, & Egelson, 1995; Egelson, Harman, & Achilles, 1996). And in a related effort, the principal of the Oak Hill elementary school in the Guilford County, North Carolina system restructured classes in grades kindergarten through 3 into a small-class format (15 students). The initiative was termed Success Starts Small (Achilles,et al. 1994; Kiser-Kling, 1995). Oak Hill school was fully Chapter 1 eligible, with 78 percent of its students in the subsidized lunch program. Matched comparison groups were used in both studies.

The results of both projects favored small classes in academic achievement small-class effect sizes were in the range 0.4[sigma] to 0.6[sigma] (Achilles,et al. 1994; Achilles, Harman, & Egelson, 1995) 0. Significantly, Success Starts Small included systematic comparisons of teaching behavior in small and regular classes:

* Teachers of small classes spent significantly more time on task and significantly less time on discipline or organizational matters compared with teachers of regular-size classes.14

Conclusions. Both Project STAR and the LBS provide compelling evidence that small classes in the primary grades are academically superior to regular-size classes. The findings were confirmed for every school subject tested. Teachers of small classes received no special instructions or training; the outcomes result from class size and from whatever perceptions and advantages accompany having substantially fewer students in a room with one teacher. This is not to say, of course, that the effects could not be accentuated if additional teacher preparation initiatives were provided.

A clear small-class advantage was found for inner-city, urban, suburban, and rural schools; for males and females; and for white and minority students alike. The few significant interactions found each year indicated greater small-class advantages for minority or inner-city students. Targeting small classes in particular schools or districts may provide the greatest benefits at a cost that is contained, although it may also mean denying the benefits to other students or schools.

These studies were based on research suggesting that small-class benefits are most likely to occur in the primary grades. The findings of Project STAR are limited to grades kindergarten through 3 no reasonable extrapolation beyond those grades can be made from these data. At the same time, the LBS results indicate clearly that the effects carry over into later years. The large, diverse database created through STAR, the LBS, and ongoing data collections offers the opportunity to answer a number of significant questions about the long-term effects of small classes on achievement, pupil engagement in school, and student behavior.

Assessing the Costs and Benefits of Smaller Classes

Without exception, the greatest obstacle to widespread implementation of smaller classes is the expense of additional teachers and classrooms. The cost issue is raised by researchers (Tomlinson, 1990) and by state and local policy makers who control the purse strings. They are to be commended for being cautious with tax dollars until the expenditure is of proven worth. At the same time, we do not have any widely-accepted procedures for determining the dollar value of particular increments in school achievement. And most economic analyses of class size to date have been severely limited. However, several approaches to the problem have been taken.

Educational Production Functions

The production function approach relies heavily on multiple regression analysis to relate a series of inputs (such as cost factors) to an output (such as student achievement). Hanushek (1986) reviewed 112 studies that used educational production functions to examine the effects of instructional expenditures on student achievement, using indicators such as teacher experience, teacher education, and pupil/teacher ratio.15 Pupil/teacher ratio was statistically significant in only 23 of the 112 studies, only 9 of which were significant in the expected direction. This "vote counting" procedure led Hanushek to conclude that pupil/teacher ratio is not an important correlate of student performance. More sophisticated analyses of the same data, however, have led others to conclude that low pupil/teacher ratios (and other cost-related inputs) are associated with increased pupil performance (Hedges, Laine, & Greenwald, 1994; Laine, Greenwald, & Hedges, 1995).

Unfortunately, the production-function approach often fails to consider findings of earlier research on class size. For example, most production-function analyses do not focus on the elementary grades, although two recent exceptions are noteworthy. In an analysis of national survey data at the district level, Wenglinsky (1997) concluded that expenditures to reduce pupil/teacher ratios impact positively on academic achievement at grade 4 but not at grade 8. Ferguson and Ladd (1996) analyzed achievement scores for students in grades 4, 8 and 9 of 131 districts in Alabama. These researchers used average class size in their multi-level regression models instead pupil/teacher ratio, concluding that class size does matter in both the earlier and later grades.

Other important differences remain. Most production function analyses include schools and districts with classes within "normal" ranges 22 to 40 students or so and the results do not answer the question of what the impact would be if classes were reduced substantially.16

Of greater concern, most production function analyses focus on school-wide or district-wide pupil/teacher ratios rather than actual class size.17 For a host of reasons, pupil/teacher ratios do not indicate how many students are enrolled in any given class or interacting with the assigned teachers (see Boozer & Rouse, 1995, for a comparison). Project STAR demonstrated the benefits of a small- class setting and provided some insight into why they occurred. It did not demonstrate that reducing the pupil/teacher ratio for a school or district would have the same impact, unless actual class sizes decreased at the same time.

Cost Analyses

Cost-effectiveness analysis examines both costs and consequences in considering alternatives for decision making. In educational applications, outcomes are typically assessed in terms of school achievement. Levin (1988) illustrated this approach to compare four strategies for educational improvement: cross-age tutoring, computer-assisted instruction, lengthening the school day, and reducing class size. Data on class size were taken from 14 evaluations collected in previous research; effect sizes were expressed as "estimated months of achievement gain" in reading and mathematics. Costs were estimated using an "ingredients approach" which involved the identification of ingredients of each intervention and their respective values, and determination of the overall cost of implementation. For example, the ingredients needed to reduce class size include personnel, facilities, and equipment.

Although the projected annual cost per student of reducing class size by five students was not found to be as great as either lengthening the school day or use of computer-assisted instruction, larger reductions in class size become quite expensive:

* With respect to an additional month of mathematics achievement, reducing class size was the most cost effective of all interventions except for peer tutoring; and

* With respect to reading achievement, reducing class size was estimated to be the least cost effective except for tutoring by adults.

The principles of cost-effectiveness analysis are sound, if fraught with methodological difficulties. The cost of an intervention can often be determined with some degree of accuracy, but the effectiveness side of the equation is more complex. Even in the simplistic applications given by Levin (1988) a small change in an effect size can have a large impact on the cost-effect ratio. When an intervention has numerous or diverse outcomes (only as different as mathematics and reading), or effects that differ from one population to another, the method provides no clear-cut way to determine cost effectiveness in toto. Introducing small classes into a school or district is at least this complex, precluding any easy answers to the cost-effectiveness question. One analysis of costs is noteworthy even though it did not consider small classes directly. King (1994) compared costs time and money associated with three educational interventions: Henry Levin's Accelerated Schools, Robert Slavin's Success for All, and James Comer's School Development Program. Although the Accelerated Schools and the School Development Program have costs that are similar, Success for All is more expensive to implement. The major expense of Success for All with demonstrated efficacy is the cost of additional staff members, particularly tutors. It would be useful to compare the costs, benefits, and feasibility of implementing this program with those of reducing class size. The main effective ingredient of Success for All may be the smaller number of students working with a particular teacher or tutor, that is, a small-class arrangement.

Further Work

The question posed by cost-effectiveness analysis is entirely appropriate, namely: What benefits are associated with what levels of investment? The current state of knowledge dictates that we evaluate the effectiveness of

small classes more completely by documenting the full spectrum of outcomes that are realized, and ask whether the extra investment can be put to best use by directing it to schools where it is needed most, for example, those serving students with poor educational prognoses.

The database created for STAR and the LBS can provide a fuller picture of short- and long-term outcomes. There is a real possibility that attending a small class in the primary grades can begin students on a path that reduces the need for special education, grade retentions or disciplinary measures, and increases the likelihood of high school graduation. Even if there is no further payoff after a student graduates, the cost savings would be appreciable.

Economist Alan Odden (1990) explored whether the effects of reducing class size on student achievement could be achieved with other lower-cost interventions, or whether larger effects could be obtained through other interventions at the same cost. He concluded that particular uses of small classes are worthwhile, especially in kindergarten through grade 3. Odden recommended reducing class size for students achieving below grade level and combining individual tutoring with classes reduced to 15 students for language arts-reading instruction. He also proposed that small classes be coupled with a "larger comprehensive set of strategies" shown to be effective for low-income, ethnic and language minority students. Early childhood education is one example.

Unfortunately, at this point in time there are no well-established procedures for summarizing diverse effects of any major intervention or, further, for comparing one intervention with another. The effect sizes in Table 1 only begin to indicate the range of outcomes, and even these are not well represented by one or two figures. If outcomes are attained that are conceptually different (e.g., improved behavior) the problem of comparison becomes even more complex.

To obtain valid comparisons with other specific instructional strategies, the duration of the intervention also needs to be considered. For example, individualized instruction (tutoring and computer-assisted instruction) and cooperative learning (see Slavin & Madden, 1989; Wasik & Slavin, 1990) are often utilized for a portion of the day to provide support in one or a few school subjects. To compare, the costs and effects should be prorated to ask what the cost-benefit ratio would be if the strategy were implemented all day for all school subjects. Making comparisons with full-scale intervention programs in which small classes, small groups, or tutors are a component (e.g., Success for All) is a slightly different matter. It may be feasible to estimate the effect of this component alone and compare that to overall program effectiveness. In either case, a number of methodological issues need to be resolved before meaningful cost comparisons can be obtained.

Instructional Practice snd Student Behavior

Two questions may be posed with respect to teaching practices in small classes:

1. How does teacher behavior actually change when there are fewer students in the classroom (and are these changes beneficial)? and

2. What sorts of teaching practices should be implemented to take maximum

advantage of a small-class setting?

Some answers to the first question are available and are summarized below but, to date, the second question can only be answered with additional research.

This discussion focuses on the construct "student engagement" on the assumption that a primary objective of instructional practice should be to maximize the engagement of individual students in the learning process. Research is reviewed that addresses three propositions, each of which is discussed in detail:

* Student engagement having both behavioral and affective elements is essential to learning;

* Disengagement from learning in both behavioral and affective forms is especially problematic among students at risk; and

* Small classes, by their nature, promote student engagement in learning and provide the conditions for teachers to encourage student engagement further, if they wish.

Student Engagement

The phrase "engagement in school" is often cited as an essential component of dropout prevention programs or other interventions for students at risk. However, there have been very few attempts to define engagement behaviorally or to study it as part of the learning process. Finn (1989) presented a model of student engagement with two central components, participation and identification.

Participation, the behavioral component, includes basic behaviors such as the student's acquiescence to school and class rules, arriving at school and class on time, attending to the teacher, and responding to teacher-initiated directions and questions. Noncompliant behavior for example, inattentiveness, disruptive behavior, or refusing to complete assigned work represents a student's failure to meet these basic requisites. Other levels of participation include initiative-taking on the part of the student (initiating questions or dialogue with the teacher, engaging in help-seeking behavior), and participation in the social, extracurricular, and athletic aspects of school life.

Identification, the affective component, refers to the student's feelings of belonging in the school setting (sometimes called school membership) and valuing the outcomes that school will provide, for example, access to post-school opportunities.

To the extent that it has been studied, the relationship of specific engagement behaviors with academic performance is strong and consistent across populations defined by background characteristics and grade level (see Finn, 1989; Finn, 1993; Finn & Rock, 1997; for summaries). These studies also have shown that positive engagement behaviors explain why some students perform well in school in spite of the adversities they face as members of high-risk populations; that is, they are "academically resilient."

Engagement and Students At Risk

Behavioral and affective disengagement from class and school is a particular problem among minority students from low-income homes (Steele, 1992). It may be difficult or impossible for some students to see any advantage to school participation when the immediate rewards are few and relationships with school staff are adversarial. And there is a substantial body of evidence that poor engagement behaviors are more common among students at risk. For example, minority students participate less fully in learning-related activities in class (Finn, Folger, & Cox, 1991; Lamborn, et al., 1992; Treuba, 1983) and exhibit more behavior problems in school (Farkas,et al. 1990; McFadden,et al. 1992; Velez, 1989) in comparison to their non-minority peers.

One form of disengagement inattentive-withdrawn behavior is worthy of special note because of educators' failure to recognize the severity of the problem, even though it has been shown to be related to depressed academic performance in the elementary grades (Finn, Pannozzo, & Voelkl, 1995). Exhibited more commonly among minority students, inattentive- withdrawn behavior has been characterized as a "loss of contact with what is going on in class" (Swift & Spivack, 1968, p. 141). Such students generally avoid calling attention to themselves;18 may seem distracted or preoccupied; and, if required to participate in classroom interactions, may give responses that are off-topic. They are even less likely than disruptive students to be directed to constructive learning activities. Finn, Pannozzo, and Voelkl (1995) found that, although the academic performance of both groups was below par, inattentive-withdrawn students performed significantly lower than disruptive students on all achievement measures.

It is established that small classes have a positive impact on academic achievement, at least in the early grades. If small classes also have a positive effect on student engagement, then the effects are likely to be especially profound for minority students and for other students at risk of educational failure. Further, a small class setting may make it difficult for a youngster to withdraw from participating, and make it difficult for a teacher to overlook the needs of particular students.

These relationships can be summarized in the form of a diagram:

Figure 1. Relationship between class size and academic performance
[Class]

Although the diagram is intended only to indicate where class size and engagement fit into a larger picture, it serves as a rudimentary model for explaining pupil achievement. The arrow from academic performance to student engagement represents the assumption that positive outcomes tend to reinforce productive behaviors; if this cycle is established, the likelihood that a student will persist in school is also increased.19

Teacher and Pupil Behavior in Small Classes

Until recently, the classroom processes that distinguish small from large classes have proven remarkably elusive. For example, a well-designed study of process was conducted in Toronto, Canada (Shapson,et al. 1980), Teachers and students in grade 4 classes were assigned at random to one of four class sizes: 16, 23, 30, or 37 pupils. Students were randomly reassigned in grade 5 and followed for another year. In addition to achievement measures, ratings were made by trained observers that included measures of

teacher-pupil interaction, pupil participation, pupil satisfaction, method of instruction, subject emphasis, physical conditions, use of instructional aids, classroom atmosphere, and the quality of classroom activity. Additional questionnaires were administered to participating teachers and pupils.

In spite of the plethora of measures, most of the findings were negative. Teachers expected smaller classes to facilitate more individualized programs and stated later that their expectations were confirmed. They generally had more positive attitudes in the smaller classes and were pleased with the ease of managing and teaching in a small-class setting. They felt that they had made changes to adapt to the different class sizes. However:

> The observation of classroom process variables revealed very few effects of class size. Class size did not affect the amount of time teachers spent talking about course content or classroom routines. Nor did it affect the choice of audience for teachers' verbal interactions; that is, when they changed class sizes, teachers did not alter the proportion of their time spent interacting with the whole class, with groups, or with individual pupils. (pp. 149-150)

No differences were found in pupil satisfaction or affective measures, and no differences were found for most teacher activities, subject emphasis, classroom atmosphere, or the quality measures.

We can only speculate about the reasons for the negative findings in such a thorough investigation. One possibility, raised in Project STAR and Project Challenge, is that a small class intervention in later grades (grade 2 and up) is not as effective as an earlier intervention. However, even today, the question of classroom process remains a top priority for further work. Some recent research has begun to reveal differences associated with class size.

A study of teaching practices in year 5 mathematics classes conducted in Melbourne, Australia (Bourke, 1986) produced a list of factors related to class size. The 63 classes studied ranged from 12 to 33 students, with more than 10 percent of the classes having 20 students or fewer. Significant positive correlates of class size included amount of noise tolerated, non-academic management, and teacher lectured or explained. The significant negative correlates were more numerous: use of whole class teaching, amount of homework assigned and graded, teacher probes after a question, teacher directly interacting with students, and positive teacher response to answer from student.

The non-experimental nature of the study leaves us with a number of possible explanations for these correlations, and the results may be specific to mathematics. However, the pattern of results suggests that in smaller classes:

* Less time is spent on classroom management; and

* There is more interaction between teachers and individual students, with the interaction more protracted.

Both of these are conducive to increasing the academic engagement of pupils.

Several STAR-related studies also support these conclusions. For example, observations were made of mathematics and reading lessons in 52 of STAR's grade 2 classrooms (Evertson & Folger, 1989). Although the amount of observation time was limited, the positive findings included the following.

* "Teachers in the small classes devoted an average of an hour to reading instruction, while teachers in regular classes spent an hour and twenty-four minutes" (p. 7). That is, higher average levels of performance were obtained with less time expenditure.

* In mathematics, students in small classes initiated more contacts with the teacher, for purposes of clarification, giving answers to questions that were open to the whole class, and contacting the teacher privately for help.

* In reading, small classes had more students on-task and fewer students off-task and spent less time waiting for the next assignment, compared with students in regular classes.

* Teachers in small classes were rated as better monitors of students' understanding of class material and as more consistent in their management of student behavior.

Interviews conducted with STAR teachers were consistent with the observations. Teachers preferred the small-class setting and felt they were able to provide more individual attention, make greater use of supplemental texts and enrichment activities, and provide more frequent opportunities for pupils to engage in firsthand learning activities (Bain,et al. 1992). In total, it appears that classroom management was more efficient and the quality of teacher-student interaction was improved in smaller classes.

North Carolina's Success Starts Small (Achilles,et al. 1994; Kiser-Kling, 1995) provided further support. In this study, trained observers collected over 7,100 "communication events" in the small and matched regular-size classes. Events were classified as personal, institutional, or task oriented. In brief, the study found a greater percentage of on-task events in small classes and a smaller percentage of institutional events (e.g., discipline or organizational) in comparison to regular-size classes. On-task behaviors increased as a percentage of all behaviors between October and April in small classes, and decreased over the same time span in the larger classes. Further, discipline referrals among grade 1 pupils declined in small classes from 38 to 28 to 14 over the 3-year period.

The studies described here indicate that student engagement and the conditions that facilitate engagement are affected positively in a small-class setting. In general, management problems were reduced and instructional interactions were enhanced.

Other Outcomes

Short- and long-term benefits in addition to enhanced performance and academic engagement may accrue from small-class participation. Research to date suggests a number of practices that may be impacted, as described below.

Discipline. The STAR grade 4 follow-up (e.g., the LBS) demonstrated that

students who had been in small classes were less disruptive than their peers in regular classes. The Success Starts Smallproject documented that grade 1 disciplinary referrals dropped over successive years in small classes. We have yet to learn whether this pattern persists through the grades.

Grade retentions. A dissertation study was conducted from STAR data that focused on pupils who entered kindergarten and grade 1 as retainees (Harvey, 1993). The study concluded that proportionately fewer students were retained in small classes and that pupils in small classes were passed to the next grade with a wider range of scores. The possibility of using small class placement as an alternative to grade retention was raised. To date, no analysis of student retentions through later grades has been performed.

Special education. With both academic and behavioral advantages, it is possible that small classes could reduce the need for special education placements. This would, of course, represent an important cost savings.

Attendance. The STAR analysis of attendance did not reveal any differences in grades kindergarten through 3. However, younger pupils do not have the autonomy that would permit skipping classes or school. Attendance needs to be monitored through later grades.

Summary

Project STAR demonstrated that small classes benefit students in grades kindergarten through 3 academically. That pupil behaviors are affected was shown clearly in the STAR grade 4 follow-up (i.e., the LBS). Ratings of specific engagement dimensions revealed improvements in the expenditure of effort, initiative taking, and reduced disruptive and inattentive behavior in comparison to students in regular classes. Both of these outcomes enhanced performance and academic engagement are likely to be beneficial especially to students at risk. Yet results for this population have not been examined closely enough to reveal the extent to which this is so.

Substantially more research is needed to tell us about the connections among teaching practices, engagement behaviors, and academic achievement particularly for students at risk, and particularly through the later grades.

Research Priorities: Five Issues in Need of Further Research

The past decade of research on class size has opened exciting possibilities for improving the performance of students in the elementary grades and, hopefully, in later grades as well. Recognizing the potential of small classes, a number of states are beginning initiatives to reduce class size in some or all districts. At the same time, important questions remain unanswered, especially with respect to students at risk; the most pressing of these are outlined here.

Many of the issues raised in this research agenda can be partially addressed through the use of existing data. Using extant data bases offers unique opportunities and considerable economy. The process is not intrusive and shortens the time required to provide information substantially (Cooley & Bickel, 1986). In this instance, the data base that can be assembled from STAR and related studies is of unusual scope and quality. Mosteller (1995)

noted:

> Because a controlled education experiment (as distinct from a sample survey) of this quality, magnitude, and duration is a rarity, it is important that both educators and policy makers have access to its statistical information and understand its implications. Thought should be given to making sure that this information is preserved and well documented and that access to it is encouraged. (p. 126)

As of this writing, the STAR/LBS data base consists of the original kindergarten through grade 3 data on approximately 7,100 children each year,20 including class placement, demographic information, and achievement scores obtained annually. The LBS follow-up data include achievement tests through grade 9, ratings on the Student Participation Questionnaire in grades 4 and 8, and student responses to the "Identification with School" scale, administered in grade 8. Other STAR data have been collected but are not computerized; these include teacher exit interviews, teacher and teacher-aide time logs, data on a matched sample of comparison schools that did not participate in the small-class experiment, and observations of a sample of grade 1 teachers conducted the year prior to teaching a small class and again during the small-class year. If STAR/LBS data were made available to the research community, their analysis could prove invaluable.

At the same time, new data may be preferable for answering some questions and may be the only way to obtain definitive answers to others. In order to obtain answers, it is important that districts and states undertaking small-class initiatives systematically collect information before, during and after implementation. Not only will problems and successes associated with small-class initiatives be documented, but there is much to be learned of interest to educators generally.

Issue 1: Short-Term and Long-Range Effects of Small Classes for Students At Risk

Additional research on the effects of small class size, focused on at-risk students is needed to answer questions such as those discussed below.

> What are the "true" immediate and continued effects of small classes on the achievement of students at risk?

Immediate outcomes. Project STAR found significant academic benefits for pupils enrolled in small classes. However, as many as 18 percent of the classes drifted out of the ranges defined as "small" or "regular" during the 4-year study when students transferred into or out of participating schools. The Burke County study and Success Starts Small found small-class effect sizes as large as 0.4[sigma] and greater. An examination of just those STAR classes that remained in-range may yield effects substantially larger than those in Table 1.

A focused analysis of STAR data also could ask whether small classes reduce the achievement gap between minority students or students from low-income homes, and their non-minority peers. Since some students were only in small classes for 1, 2, or 3 years, the reanalysis could also reveal the benefits of small classes to students who are more mobile than others an issue of particular importance to students at risk.

Medium-term outcomes. The LBS documented a continued but diminishing impact

of small classes over subsequent years (grades 4 through 9). Again, analyses have not focused on the at-risk population and did not examine the achievement gap between white and minority students.

Long-term outcomes. Further data are needed to address the effects on pupil performance through high school.

> What are the effects of small classes on non-achievement outcomes among students at risk?

Several studies (e.g., Rand study, Head Start, Project High Scope) suggest that the benefits of some early interventions persevere through and beyond the school years. The LBS documented improved classroom behavior in grade 4 but went no further. Of the negative events experienced disproportionately by students at risk, it is important to ask whether small classes reduce the need for disciplinary action, for special education placement, for in-grade retention, and increase the likelihood of a student graduating from high school.

Issue 2: Teaching Practices to Maximize the Effectiveness of Small Classes

Studies to date suggest that small classes create a more personalized environment for teacher and students and that small classes produce a time efficiency by reducing the need for discipline and classroom management and delivering effective instruction in a shorter amount of time. Additional research is needed to answer subsequent questions such as those discussed below. How do the most effective teachers take advantage of a small class setting to deliver more individualized instruction to pupils? and How can other teachers be taught to use these strategies?

Some teachers may use techniques designed to increase the participation of each individual student in classroom interactions. This is important in light of some youngsters' tendency to withdraw from participation a particularly debilitating strategy.21 Some teachers may be able to increase parents' involvement in their youngsters' schooling. And some may be available to provide extraordinary support (e.g., extra attention; after-school help) for students having difficulty with class material; these "extras" are often lacking in schools serving students at risk (Ralph, 1989).

> How do the most effective teachers take advantage of the time-efficiency provided by small-class instruction?

How do teachers in small classes allocate their time to working with individual students, small groups, or the whole class? What kinds of activities can be undertaken when instruction is more efficient? For example, if course material is reinforced, are additional activities implemented to push the students beyond the usual content? Is more focused evaluation and feedback provided?22

Issue 3: School and Classroom Conditions That Interact With Class Size

STAR findings showed a disproportionate impact on minority students in some achievement areas each year (kindergarten through grade 3). Further research should examine other characteristics of schools and programs that may interact with class size and address questions such as those discussed below.

Can small classes offset some of the disadvantages of attending a large school?

Past research has documented that attendance and participation in academic extracurricular activities are inversely related to school size, that is, larger schools have decreased student participation (Lindsay, 1982; Cockman, Bryson, & Achilles, 1989; Fowler, 1992). There is also a carryover effect: high participants in high school tend to participate actively in post-schooling cultural and community activities (Lindsay, 1984). Most of this research involved high-school students. The mechanisms that explain the association of school size with student participation have not been uncovered, but results indicate that smaller schools are seen as "warmer" and more supportive settings (Finn & Voelkl, 1993);23 that is, they provide a more personalized environment.

Given that large schools are ubiquitous, this research raises questions about the potential benefits of small classes. One study using STAR data (Nye, 1995) concluded that the negative correlation between school size and achievement disappears for students attending small classes. Other questions yet to be addressed include: Does attending a small class even in the earlier grades produce higher student attendance and involvement in later grades independently of the size of the school? If so, is this associated with improved student performance and increased likelihood of graduating from high school? Is there an interaction of class size and school size in the elementary grades as well? Is the increased engagement associated with small classes beneficial particularly to students at risk attending large, perhaps more impersonal, schools?

Do small classes accentuate and extend the benefits of other early childhood programs and practices?

To date, no analyses have examined the combined impact of small classes with federal, state, or local programs directed at students living in poverty or who are otherwise at risk for school failure (e.g., Title I). Preschool participation and attending full-day kindergarten may also promote the development of children generally and students at risk in particular. Some states do not have state-mandated kindergarten and, in others, half-day kindergarten is common. An analysis of some of the STAR data (Achilles, Nye, & Bain, 1994 1995) indicated a significant "test score value" for children who attended kindergarten. Further work is needed to document the combined impacts of preschool participation, attending kindergarten, and being enrolled in a small class on students at risk. Both short-term and long-range outcomes should be examined.

Do small classes accentuate and extend the benefits of other classroom practices?

Several examples illustrate this research question:

Cooperative learning has been used to promote the achievement of all students but students at risk in particular. Are cooperative learning techniques less effective or equally effective if the class size is small, or are the benefits accentuated?

Heterogeneous small groups and heterogeneous classes have been found to be academically beneficial to at least some students. In a review of the

problem of "stratification" in heterogeneous classrooms, Cohen and Lotan (1995) noted that, with appropriate intervention, higher rates of participation can be encouraged among low-status students. No investigation has examined the interaction of class size with class heterogeneity by racial-ethnic, socioeconomic, or primary language characteristics. There are many possible avenues to explore.

Teacher aides are a major education intervention (e.g., Title I; special education; some remedial programs). The academic value of teacher aides depends both on their qualifications to provide instruction and on how they are deployed (e.g., for order-keeping, for bookkeeping, or as a true teaching resource). Research should ask whether teacher aides can be utilized to further enhance the benefits of small classes, or whether judicious use of well-prepared teacher aides in regular-size classrooms can produce some of the same benefits as small classes, but at lower cost.

Issue 4: Small Classes and Positive "Long-Term Trajectories" for Students At Risk

Educational risk may be described in terms of group status characteristics or in terms of a set of behaviors. If these "risk" behaviors manifest in negative ways, such as not attending to the teacher, not completing required work, and skipping school, they create impediments to learning. On the other hand, if a student exhibits positive behaviors, such as attending and participating (e.g., "engagement behaviors"), the behaviors may serve as "protective" mechanisms to improve the chances of school success in spite of group risk status. While status and behavioral risk factors are often found in the same individuals, risk behaviors may be amenable to influence by parents, school personnel, and school programs.

There is evidence that risk behavior in school and the classroom and its obverse, engagement, is developmental and begins in the early school grades (see Finn, 1989, 1993). Active participation in the early grades, accompanied by some degree of academic success, serves to perpetuate continued participation throughout the school years; this would be a "positive trajectory." When a young student does not participate in the classroom, this may begin a cycle that results in adverse consequences over time. Barriers to success multiply. Risk factors "cluster;" that is, multiple risk factors are likely to occur in the same individual especially over time. And risk factors "track;" that is, they have early forms that evolve into fully developed forms over time that are increasingly difficult to alter. Thus it is essential that educators identify and understand forms of disengagement from school in the early grades and do all that is feasible to intervene at that point. The central question then is:

> Can small classes in the early grades begin students on a positive trajectory that persists through the school years?

Three key issues should be explored further. First, we need to assess the short-run and long-run likelihood of adverse consequences of early risk behavior. The relationship of status and behavioral risk factors in the early grades with absenteeism, suspensions, retention in grade, loss of identification with school and dropping out, and even drug use and contacts with police in later years should be studied carefully. Patterns of tracking and clustering of risk factors should also be documented.

Second, we need to understand why some students at risk succeed

academically in spite of the obstacles they may face because of group status characteristics. Such students have been termed "educationally resilient" (see Nettles & Pleck, 1994; Rutter, 1990). With respect to resilient students we should ask whether they exhibit positive engagement behaviors beginning in the early grades. What sorts of preschool and early school experiences did they participate in? What sorts of support for learning did they receive from their teachers, parents, peers, and others?

Finally, we need to ask whether small classes in the early grades interrupt patterns of disengagement, decrease the likelihood of adverse consequences, and increase the likelihood of positive behaviors (and achievement) over subsequent years.

Issue 5: Assessing the Costs of a Small-Class Initiative

Every school, district or state planning to undertake a class-size initiative confronts the budget question. However, the question is not as simple as asking "How much more will additional teachers and classrooms cost?" because associated benefits may produce savings and careful planning may be able to contain the expense. There are a number of related issues, elaborated below, about which there is a small but growing base of knowledge. Additional research is needed to address these questions more fully.

Do classes of 15 18 pupils really cost more if weighed against the benefits that accrue?

Researchers have not yet assessed the total impact of small classes, but research has demonstrated academic benefits in all subjects that persist into later grades, and improved learning behavior at least through grade 4. Related studies previously discussed have indicated fewer grade retentions and fewer disciplinary referrals. If, in the long run, the need for remedial and special education teachers is reduced, discipline problems and violence are reduced, and/or fewer students leave school without graduating, then there is a real gain on the output side of the equation. Most of these effects are well-documented while some require further research. It is clear, however, that small classes produce an array of academic and behavioral benefits that have cost-savings value.

How can the costs of implementing small classes be contained?

If hiring more teachers is the only strategy used to reduce class size, a small-class initiative undoubtedly will be expensive. Again, however, it may not be expensive in relation to the benefits that accrue or in comparison to other interventions with an equally broad array of outcomes. Although, at present, there are no prescribed solutions to the issue of cost, a number of districts have found ways to achieve small classes, even within the usual per-pupil expenditures. Some schools have experimented with creative scheduling plans. Others have redeployed staff in order to achieve smaller class sizes; for example, by assigning Title I teachers or specialty teachers to small classes, using supplemental state funds for additional teachers, or allocating part-time teacher aide funds to full-time teaching positions (see also Miles, 1995).

While reassignments such as these do challenge people's thinking about "business as usual," initial reactions from these sites indicate that both teachers and administrators are satisfied with the decisions. However, the

experiences of these schools and districts must be systematically documented in order for us to obtain further answers to the question of how costs can be contained. Additional field-based research is needed urgently to build a broader knowledge base that educators can use for decision making. Further, a mechanism is needed for compiling the experiences of local sites into a central database that can be tapped by researchers and policymakers alike.

References

Achilles, C. M. (no date). Summary of recent class-size research with an emphasis on Tennessee's Project STAR and its derivative research studies. Nashville, TN: Tennessee State University, Center of Excellence for Research and Policy on Basic Skills.

Achilles, C. M., Kiser-Kling, K., Owen, J., & Aust, A. (1994). Success Starts Small: Life in a small class. Final Report. Greensboro, NC: University of North Carolina.

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

Achilles, C. M., Nye, B. A., & Bain H. P. (1994 1995). "Test-score 'value' of kindergarten for pupils in three class conditions at grades 1, 2, 3." National Forum of Educational Administration and Supervision Journal (NFEAS), 12(l), 3 15.

Achilles, C. M., Nye, B. A., & Zaharias, J. B. (1995, April). Policy use of research results: Tennessee's Project Challenge. Paper presented at the annual convention of the American Educational Research Association, San Francisco.

Bain, H. P., Achilles, C. M., Zaharias, J. B., & McKenna, B. (1992). "Class size does make a difference." Phi Delta Kappan, 74, 253 256.

Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. A., & Weikart, D. P. (1984). Changed lives: The effects of the Perry preschool program on youths through age 19. Ypsilanti, MI: High/Scope.

Boozer, M., & Rouse, C. (1995, June). Intraschool variation in class size: Patterns and implication. (Working Paper No. 344). Washington, DC: National Bureau of Economic Research, Industrial Relations Section. (ERIC document number ED 385935).

Bourke, S. (1986). "How smaller is better: Some relationships between class size, teaching practices and student achievement." American Educational Research Journal, 23, 558 571.

Brophy, J., & Evertson, C. (1981). Student characteristics and teaching. New York: Longman.

Brophy, J., & Rohrkemper, M. (1989, October). Teachers' strategies for coping with shy/withdrawn students. East Lansing, MI: The Institute for Research on Teaching. (Research Series No. 199).

Center for School Assessment. (1986, July). PRIME TIME: The relationship

between class size and achievement for first grade students in Indiana 1984 85. Indianapolis, IN: Indiana Department of Education.

Chase, C. I., Mueller, D. J., & Walden, J. D. (1986, December). PRIME TIME: Its impact on instruction and achievement. Final report. Indianapolis, IN: Indiana Department of Education.

Cockman, D., Bryson, J., & Achilles, C. M. (1989). High school size and student participation in school activities: North Carolina, 1988 89. Paper presented to the North Carolina Association of Research in Education Annual Conference, Research Triangle Park.

Cohen, E. G., & Lotan, R. A. (1995). "Producing equal-status interaction in the heterogeneous classroom." American Educational Research Journal, 32, 99 120.

Cooley, W., & Bickel, W. (1986). Decision-oriented education research. Boston: Kluwer-Nijhoff.

Egelson, P., Harman, P., & Achilles, C. M. (1996). Does class size make a difference? Recent findings from state and district initiatives. Greensboro, NC: Southeastern Regional Vision for Education (SERVE).

Evertson, C. M., & Folger, J. K. (1989, March). Small class, large class: What do teachers do differently? Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Farkas, G., Grobe, R. P., Sheehan, D., & Shuan, Y. (1990). "Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban school district." American Sociological Review, 55, 127 142.

Ferguson, R. F., & Ladd, H. F. (1996). "How and why money matters: An analysis of Alabama schools" in H.F. Ladd (Ed.), Holding schools accountable. Performance-based reform in education (pp 265 298). Washington, DC: The Brookings Institution.

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

Finn, J. D. (1993). School engagement and students at risk. Washington, DC: National Center for Educational Statistics, U.S. Department of Education. (NCES 93 470).

Finn, J. D., & Achilles, C. M. (1990). "Answers about questions about class size: A statewide experiment." American Educational Research Journal, 27, 557 577.

Finn, J. D., Folger, J., & Cox, D. (1991). "Measuring participation among elementary grade students." Educational and Psychological Measurement, 51, 393 402.

Finn, J. D., Fulton, D., Zaharias, J., & Nye, B. A. (1989). "Carry-over effects of small classes." Peabody Journal of Education, 67(l), 75 84.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). "Disruptive and inattentive-withdrawn behavior and achievement among fourth graders." The Elementary School Journal, 95, 421 434.

Finn, J. D., & Rock, D. A. (1997). "Academic success among students at risk." Journal of Applied Psychology, 82, 221 234.

Finn, J. D., & Voelkl, K. E. (1993). "School characteristics related to student engagement." Journal of Negro Education, 62, 249 268.

Finn, J. D., & Voelkl, K. E. (1994). "Class size" in T. Husen & T. N. Postlethwaite (Eds.), International encyclopedia of education (2nd edition). Oxford: Pergamon Press.

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

Fowler, W. J., Jr. (1992, April). What do we know about school size? What should we know? Paper presented at the American Educational Research Association, San Francisco.

Glass, G. V., & Smith, M. L. (1978). Meta-analysis of research on the relationship of class size and achievement. San Francisco: Far West Laboratory for Educational Research & Development.

Hanushek, E. A. (1986). "The economics of schooling: Production and efficiency in public schools." Journal of Economic Literature, 24, 1141 1177.

Harvey, B. (1993, December). An analysis of grade retention for pupil in K-3. Unpublished doctoral dissertation. University of North Carolina, Greensboro.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994, April). "Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes." Educational Researcher, 23, 5 14.

King, J. A. (1994). "Meeting the educational needs of at-risk students: A cost analysis of three models." Educational Evaluation and Policy Analysis, 16(l), 1 19.

Kiser-Kling, K. (1995). Life in a small teacher-pupil ratio class. Unpublished Ed.D. dissertation. University of North Carolina, Greensboro.

Laine, R. D., Greenwald, R., & Hedges, L. V. (1995). "Money does matter: A research synthesis of a new universe of education production function studies" in L. O. Picus & J. L. Wattenbarger (Eds.), Where does the money go? Resource allocation in elementary and secondary schools (pp. 44 70). Thousand Oaks, CA: Corwin Press.

Lamborn, S. D., Brown, B. B., Mounts, N. S., & Steinberg, L. (1992). "Putting school in perspective: The influence of family, peers, extracurricular participation, and part-time work on academic engagement" in F. M. Newman (Ed.), Student engagement and achievement in American secondary schools (pp. 153 181). New York: Teachers College Press.

Levin, H. (1988). "Cost-effectiveness and educational policy." Educational Evaluation and Policy Analysis, 10(l), 51 69.

Lindsay, P. (1982). "The effect of high school size on student participation, satisfaction, and attendance." Educational Evaluation and Policy Analysis, 4(l), 57 65.

Lindsay, P. (1984). "High school size, participation in activities, and young adult social participation: Some enduring effect of schooling." Educational Evaluation and Policy Analysis, 6(l), 73 83.

Malloy, L., & Gilman, D. (1989). "The cumulative effects on basic skills achievement of Indiana's PRIME TIME A state sponsored program of reduced class size." Contemporary Education, 60, 169 172.

McFadden, A. C., Marsh, G. E., Price, B. J., & Hwang Y. (1992). "A study of race and gender bias in the punishment of school children." Education and Treatment of Children, 15, 140 146.

McGiverin, J., Gilman, D., & Tillitski, C. (1989). "A meta-analysis of the relation between class size and achievement." Elementary School Journal, 90, 47 56.

Milchus, N., Farrah, G., & Reitz, W. (1968). The self-concept and motivation inventory: What face would you wear? Preschool/kindergarten form. Dearborne Heights, MI: Person-0-Metrics.

Miles, K. H. (1995). "Freeing resources for improving schools: a case study of teacher allocation in Boston Public Schools." Educational Evaluation and Policy Analysis, 17, 476 493.

Mosteller, F. (1995). "The Tennessee study of class size in the early school grades." The Future of Children, 5(2), 113 127.

Mueller, D. J., Chase, C. I., & Walden, J. D. (1988). "Effects of reduced class size in primary classes." Educational Leadership, 45(7), 48 53.

Nettles, S. M., & Pleck, J. H. (1994). "Risk, resilience, and development: The multiple ecologies of black adolescents in the United States" in R. J. Hagerty, N. Garmezy, M. Rutter & L. R. Sherrod (Eds.), Stress, risk, and resilience in children and adolescents: Processes, mechanisms. and intervention (pp. 147 181). New York: Cambridge University Press.

Nye, K. E. (1995). The effect of school size and the interaction of school size and class type on selective student achievement measures in Tennessee elementary schools. Unpublished doctoral dissertation. University of Tennessee: Knoxville.

Odden, A. (1990). "Class size and student achievement: Research-based policy alternatives." Educational Evaluation and Policy Analysis, 12, 213 227.

Pinnell, G. S., deFord, D. E., & Lyons, C. A. (1988). Reading recovery: early intervention for at-risk first graders. Arlington, VA: Educational Research Service.

Ralph, J. (1989). "Improving education for the disadvantaged: Do we know whom to help?" Phi Delta Kappan, 70, 395 401.

Robinson, G. E. (1990). "Synthesis of research on effects of class size."

Educational Leadership, 47(7), 80 90.

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

Rutter, M. (1990). "Psychosocial resilience and protective mechanisms" in J. Rolf, A. S. Masten, D. Cicchetti, K. H. Neuchterlein, & S. Weintraub (Eds.), Risk and protective factors in the development of psychopathology (pp. 181 214). New York: Cambridge University Press.

Shapson, S. M., Wright, E. N., Eason, G., & Fitzgerald, J. (1980). "An experimental study of effects of class size." American Educational Research Journal, 17, 141 152.

Slavin, R. E. (1989). (Ed.). School and classroom organization. Hillside, NJ: Lawrence Erlbaum Associates.

Slavin, R. E., & Madden, N. A. (1989). "What works for students at risk: A research synthesis." Educational Leadership, 46(5), 4 13.

Slavin, R. E., & Madden, N. A. (1995). "Success for all: Creating schools and classrooms where all children can read" in Oakes & Quarts (Eds.), Creating New Educational Communities. NSSE Yearbook, Part 1. Chicago, IL: University of Chicago Press, 70 86.

Slavin, R. E., Madden, N. A., Karweit, N. J., Livermon, B. J., & Dolan, L. J. (1990). "Success for All: First-year outcomes of a comprehensive plan for reforming urban education." American Educational Research Journal, 27, 255 278.

Steele, C. M. (1992, April). "Race and the schooling of black Americans." The Atlantic Monthly, 68 78.

Swift, M. S., & Spivack, G. (1968). "The assessment of achievement-related classroom behavior." Journal of Special Education, 2(2) 137 149.

Tomlinson, T. M. (1988). Class size and public policy: Politics and panaceas. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Tomlinson, T. M. (1989). "Class size and public policy: Politics and panaceas." Educational Policy, 3, 261 273.

Tomlinson, T. M. (1990). "Class size and public policy: The plot thickens." Contemporary Education, LXII(1), 17 23.

Trueba, H. T. (1983). "Adjustment problems of Mexican and Mexican-American students: An anthropological study." Learning Disability Quarterly, 6, 395 404.

Velez, W. (1989). "High school attrition among Hispanic and non-Hispanic White youths." Sociology of Education, 62, 119 133.

Voelkl, K. E. (1996). "Measuring students identification with school." Educational and Psychological Measurement, 56, 760 770.

Wasik, B. & Slavin, R. E. (1990, April). Preventing early reading failure with one-on-one tutoring: A best-evidence synthesis. Paper presented at the annual convention of the American Educational Research Association, Boston.

Wehlage, G. G., Rutter, R. A., Smith, G. A., Lesko, N., & Fernandez, R. R. (1989). Reducing the risk. Schools as communities of support. London: Farmer Press.

Wenglinsky, H. (1997). When money matters. How educational expenditures improve student performance and when they don't. Princeton, NJ: Educational Testing Service, Policy Information Center.

Word, E., Johnston, J., Bain, H., Fulton, D. B., Boyd-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. Nashville, TN: Tennessee State Department of Education.

Zaharias, J. B., Achilles, C. M., Bain, H. P., & Fulton, B. D. (1995). "Quality schools build on a quality start" in E. Chance (Ed.), Creating the quality school (pp. 117 124). Madison, WI: Magna Publishing.

1At the time of the Slavin analysis, Project STAR had not been completed.

2Slavin also commented that while teachers may change their behavior in small classes, the changes are so slight that they are unlikely to make important differences in student achievement. This issues is discussed more fully in a later section of this paper.

3Of the studies described in the next section, Project PRIME TIME maniupulated pupil/teacher ratios but failed to find a significant impact on academic achievement. In contrast, Project STAR controlled the number of pupils in each classoom; this was accompanied by differences in student performance.

4There was a training component for some teachers in grade 2. The effects on student achievement were found to be negligible. The results reported here do not include classes taught by that subsample of teachers.

5Several longitudinal analyses have been completed as well, including a K-1 analysis (Finn & Achilles, 1990)and a K-2 analysis (Finn, et al., 1990). Many important longitudingal analyses remain to be conducted.

6The exceptions did not contradict the finding of a small-class advantage. They indicated that, to some extent, the advantage was greater for students attending inner-city schools.

7One possible reason for the negative findings may lie in the difficulties in assessing noncognitive characteristics of young children. Of course it is also possible that small classes improved learning but did not affect pupils' motivation or self-concepts.

8Unpublished table obtained directly from the analyses.

9Although precise grade equivalents are not available, these differences correspond to an advantage of about .1 grade equivalents (or about 1

month)by the end of kindergarten, about 0.2 grade equivalents (or about 2 months) at the end of first grade, and somewhat more by the end of grade 2.

10Including several subtests not listed in Table 1.

11In the range 0.3* and upward (Zaharias, et al., 1995).

12Each year (1990-1994) the number of students tested was between approximately 4200 and 6000.

13Later follow-ups through grade 11 are being conducted by H.P. Bain and J.B. Zaharias of HEROS, Inc. Preliminary results indicate that the positive effects of small classes persisted at least through grade 10.

14This finding is discussed further in the later section on instructional practice and student behavior.

15Note that "indicators" are not the same as actual expenditures, and the relationship between the two may be complex.

16The step cannot even be taken "in theory" since reductions in class size would change the values of other important inputs as well.

17Ferguson and Ladd (1996) is an exception.

18Brophy and Evertson (1981) termed such students "invisible" students.

19This is not a necessary assumption since no conculsions in this paper rest on it being correct. The cycle depicted here is part of the "participation-identification model" presented in Finn (1989; 1993).

20The total sample size (approximately 12,000) exceed the number of students for any given year, since it includes pupils who were in STAR classes for 1 or 2 years but not other years.

21Brophy and Rohrkemper (1989) have produced one of the few lists of strategies to encourage participation by students who are shy and/or withdrawn.

22Both observational and interview data are needed to address these questions. Since small classes are being implemented in many states and districts across the country, further observational data should be readily accessible.

23Teachers and admnistrators may actually behave in warmer and more supportive ways in smaller schools; that is, there may be a factual basis to this perception. This has not been studied.

24At the time Project STAR began, Tennessee did not require that children attend kindergarten.