Journal of Educational Psychology
2005, Vol. 97, No. 2, 214–223

0022-0663/05/$12.00 DOI: 10.1037/0022-0663.97.2.214

Small Classes in the Early Grades, Academic Achievement, and Graduating From High School

Jeremy D. Finn and Susan B. Gerber
University at Buffalo—The State University of New York

Jayne Boyd-Zaharias
HEROS, Inc.

This investigation addressed 3 questions about the long-term effects of early school experiences: (a) Is participation in small classes in the early grades (K–3) related to high school graduation? (b) Is academic achievement in K–3 related to high school graduation? (c) If class size is related to graduation, is the relationship explained by the effect of participation in small classes on students' academic achievement? The study included 4,948 participants in Tennessee's class-size experiment, Project STAR. Analyses showed that graduating was related to K–3 achievement and that attending small classes for 3 or more years increased the likelihood of graduating from high school, especially among students eligible for free lunch. Policy and research implications are discussed.

Keywords: academic achievement, small classes, enduring effects, high school, dropout

The purpose of this investigation was to address three questions about the long-term effects of early school experiences: (a) Is participation in small classes in the early grades (K–3) related to the likelihood that a student will graduate from high school? (b) Is academic achievement in the early grades related to high school graduation? (c) If class size in K–3 is related to high school graduation, is the relationship attributable to the effect of small classes on students' academic achievement and the subsequent effect of achievement on graduation?

This study is unique in several ways. Although the relationship of class size with achievement and behavior has been documented elsewhere, no formal examination of early class sizes and graduating or dropping out 6 to 9 years later has been published previously. Also, the study was based on an extraordinary database—a large sample of students followed for 13 years,[1] with norm-referenced and criterion-referenced achievement tests administered annually and graduation/dropout information collected from official school and state records.

Early Academic Achievement and Dropping Out

There is long-standing evidence that students' academic achievement in the early grades sets the stage for much that happens in the ensuing years (see Bloom, 1964). It is also clear that academic achievement throughout the school years is related to students' leaving school without graduating. In an overview of research, the National Research Council (2001) identified a history of poor academic performance as one of three leading school-related characteristics associated with dropping out.[2]

Research on students in the middle grades (5–9) has found that when several antecedents are studied together, academic achievement makes a consistent, independent contribution to graduating from or dropping out of school (Battin-Pearson et al., 2000; Kaplan, Peck, & Kaplan, 1997). Other studies have traced the origins of dropping out to academic performance in the early grades. Barrington and Hendricks (1989) examined retrospectively the permanent records of students entering two high schools in 1981 and then followed the students through high school. Dropouts had been distinct in academic achievement and attendance from as early as third grade but did not differ on home-related characteristics. Significant differences between graduates and dropouts at the .01 level were found in third-grade scores on the Iowa Achievement Tests and, later on, the number of courses failed in Grades 7 to 12. Garnier, Stein, and Jacobs (1997) followed children in 194 families from birth through age 19. Their study used structural equation modeling to describe the relationships of family, individual, and school factors to school noncompletion. A composite of mathematics and reading grades and teacher ratings in Grade 1 had a significant influence on performance in Grade 6 ($r = -0.24$), which, in turn, had a direct impact on dropping out ($r = -0.30$).

Jeremy D. Finn and Susan B. Gerber, Graduate School of Education, University at Buffalo—The State University of New York; Jayne Boyd-Zaharias, HEROS, Inc., Lebanon, Tennessee.

This work was supported by a William T. Grant Foundation "Antecedents and Consequences of High School Gateway Events" grant. An earlier version of this article was presented at the April 2004 annual meeting of the American Educational Research Association, San Diego, California.

Correspondence concerning this article should be addressed to Jeremy D. Finn, Graduate School of Education, University at Buffalo—The State University of New York, 422 Christopher Baldy Hall, Buffalo, NY 14260. E-mail: finn@buffalo.edu

[1] Two other studies of early academic achievement and dropping out are cited in the introductory section of this article, one conducted in Chicago, Illinois (Ensminger & Slusarcick, 1992), and one in Baltimore, Maryland (Alexander, Entwisle, & Horsey, 1997). The sample in the present study was a larger and more diverse sample than in either of those studies.

[2] The other pervasive correlates of high school graduation/dropping out named were educational engagement and academic delay.

Two prospective studies followed urban Grade 1 children through their high school years, examining a range of parent and student behaviors. Ensminger and Slusarcick (1992) studied a sample of 1,242 African American students in Chicago, Illinois, about one half of whom did not graduate from high school. Among the significant predictors of dropping out were poverty, sex (female students were more likely to graduate than were male students), family structure interactively with sex, aggressive behavior in first grade, and school performance from first grade onward. The odds of graduating for male students who received As or Bs in first grade were more than twice as high as the odds for male students who received Cs or Ds; for female students, the odds for those who received As or Bs were more than 1.5 times as great. The authors noted, "Although later educational expectations and assessments of educational performance also mattered, they did not diminish the impact of earlier performance. Children's early school performance and adaptation may help establish patterns that remain relatively stable" (Ensminger & Slusarcick, 1992, p. 110).

As part of the Beginning School Study, Alexander et al. (1997) followed a sample of 790 African American and White students from the time they entered first grade in 1982 through spring of 1996. The study included an extensive set of measures including family stressors, parents' attitudes and practices, children's attitudes and school engagement, and school experiences—data gathered from school records, interviews, and parent and teacher questionnaires. A number of significant antecedents of dropping out were identified, including first-grade marks and first-grade test scores; zero-order correlations with dropping out were in the range from 0.30 to 0.38. This study also found measures of student engagement to be important to graduation, including absences from school and teachers' ratings of engagement in the classroom.

Several theoretical perspectives explain dropping out as the culmination of experiences that may begin in the early grades; the models and the data that support them have been given in Finn (1989); Newmann, Wehlage, and Lamborn (1992); Rumberger (2001); and Wehlage, Rutter, Smith, Lesko, and Fernandez (1989). All give a central role to student engagement (or disengagement) and depict dropping out as the final step in a gradual process of disengagement from school. Student behavior and academic achievement in the early grades are portrayed as important antecedents of engagement (or disengagement) in later years; studies are reviewed that have supported this premise. Student engagement (and disengagement) can also be impacted by school characteristics and practices. For example, the practice of retaining students in one or more grades—as early as first grade—is significantly related to the likelihood of leaving school without graduating (Goldschmidt & Wang, 1999; National Research Council, 2001; Randolph, Fraser, & Orthner, 2004).

The main theme of this research and theory is that dropping out of high school is not a spontaneous event but is often the culmination of a history of school experiences. These experiences may date back to the earliest grades in school or before. The present study examined the relationship of early academic achievement and early class sizes[3] with dropping out in a sample of students followed from kindergarten through high school.

Small Classes in the Early Grades

It is now established that classes of fewer than 20 pupils in Grades K–3 have a positive effect on student achievement. Three phases of research, taken together, have confirmed this relationship. Prior to the 1980s, several hundred studies appeared on the topic; this work was summarized in a meta-analysis by Glass and Smith (1978) and a review by Robinson (1990). The studies showed that classes with fewer than 20 pupils were likely to benefit students' achievement in mathematics and reading. Furthermore, the benefits seemed to be greatest in the early grades and for students from low-income homes. Many of the studies were of poor quality, however, and none was a randomized experiment.

In 1985, the Tennessee State Department of Education undertook a large randomized experiment, Project STAR, to provide more definitive answers to the class-size question. In Project STAR, students entering kindergarten were assigned at random to a small class (13–17 students), a full-size class (22–26 students), or a full-size class with a full-time teacher aide within each participating school. The class size was maintained throughout the day, all year long. Students were kept in the same class arrangement for up to 4 years (Grade 3), with a new teacher assigned at random to the class each year. Norm-referenced and criterion-referenced achievement tests were administered in the spring of each school year. In all, almost 12,000 students participated in the STAR experiment in more than 300 classrooms in schools across the state. All students returned to full-size classes in Grade 4 when the experiment ended.[4]

Project STAR results have been published elsewhere (e.g., Achilles, 1999; Finn & Achilles, 1990; Finn, Gerber, Achilles, & Boyd-Zaharias, 2001; Word et al., 1990). Secondary analysts have confirmed the basic findings using a variety of statistical approaches (Goldstein & Blatchford, 1998; Hedges, Nye, & Konstantopoulos, 2000; Krueger, 1999). Four findings are central: First, small classes were associated with significantly higher academic performance in every school subject in every grade during the experiment (K–3) and in every subsequent grade studied (4–8). Second, many of the academic benefits of small classes were greater for students at risk, that is, minority students, students attending inner-city schools, or students from low-income homes. Krueger and Whitmore (2001) used the STAR data to estimate that the White–minority achievement gap would be reduced by 38.0% if all students attended small classes in K–3. Third, students in small classes were more engaged in learning than were students in larger classes (Evertson & Folger, 1989; Finn, Fulton, Zaharias, & Nye, 1989; Finn, Pannozzo, & Achilles, 2004); this provides a partial explanation of the process by which small classes are

[3] Class size is the number of students who are regularly in a classroom with a teacher and for whom that teacher is responsible. Other writing about pupil–teacher ratios for schools, districts, or states (e.g., Hanushek, 1998) does not pertain to the educational effects of small or large classes. Actual class sizes may vary dramatically within a school or district. At times, even in districts with low pupil–teacher ratios, students may spend most of their school time in large classes (Lewit & Baker, 1997; Miles, 1995).

[4] Achievement scores, behavior ratings, and other data continued to be collected through high school.

academically beneficial. Fourth, no significant differences were found between full-size classes with teacher aides and classes without teacher aides on any test in any grade.

The third phase of research consisted of the district- and state-level class-size-reduction (CSR) initiatives that followed Project STAR. Several initiatives have been accompanied by high-quality evaluations, for example, Tennessee's Project Challenge (Achilles, Nye, & Zaharias, 1995); Wisconsin's Student Achievement Guarantee in Education (SAGE) program (Molnar, Smith, & Zahorik, 1999; Molnar et al., 2000); the CSR initiative in Burke County, North Carolina (Achilles, Harman, & Egelson, 1995; Egelson, Harman, & Achilles, 1996); and the statewide program in California (CSR Research Consortium, 2000). The outcomes of these efforts were highly consistent with STAR findings. For example, SAGE demonstrated greater effects for students at risk, with effect sizes similar to those reported for STAR (Finn et al., 2001). The weak results found in California were also consistent: The California evaluation focused on Grade 3 students, and the (significant) effect sizes were similar to those obtained in STAR for Grade 3 students who spent just 1 or 2 years in small classes (Finn et al., 2001).

Enduring Effects

The primary question of the present study is one of enduring impact: Does attending small classes in the early grades affect the likelihood of graduating from or dropping out of high school? Both empirical findings from Project STAR and theory about interventions that have lasting effects lead to the hypothesis that the answer is yes.

Although the class-size experiment ended in Grade 3,[5] researchers continued to collect achievement test data on the STAR participants through Grade 8. Attending small classes in K–3 was significantly related to academic achievement in all grades (4–8) in all subject areas (Finn et al., 2001; Hedges, Nye, & Konstantopoulos, 1999). The analysis by Finn et al. (2001) took advantage of the fact that some STAR participants attended small classes for 1, 2, or 3 years, as well as the full 4 years, and controlled for student race, socioeconomic status (SES), urbanicity, and movement into or out of STAR. Results showed that the carryover to Grades 4, 6, and 8 was strongest for students who entered small classes in kindergarten or Grade 1 and who remained in small classes for 3 or more years. Krueger and Whitmore (2001) also found that STAR students' likelihood of taking college admissions tests (ACTs/SATs) in high school was increased by participation in small classes in K–3. The increase was especially large for African American students.

To be sure, not all early interventions have long-term effects. With some programs, short-term academic benefits decrease over time even if nonachievement outcomes persist (Barnett, 1992, 1995; Lazar & Darlington, 1982; White, 1986). Both the Perry Preschool Project and most Head Start programs have exhibited this pattern (Haskins, 1989; McKey et al., 1985). Evaluators have found that achievement benefits disappeared 3 years after students left those programs, but students continued to be less likely to be placed in special education, less likely to be retained in grade, and more likely to graduate than were their nonprogram counterparts (Berrueta-Clement, Barnett, Epstein, & Weikart, 1984; McKey et al., 1985).

In contrast, the Chicago Parent–Child Centers (CPC) program has documented continuing academic benefits (Reynolds, 1997; Reynolds, Temple, Robertson, & Mann, 2001). CPC was designed to aid low-income students, especially those not served by Head Start. The program has components for preschool through Grade 3, including high-quality educational, family, and health services. Individual children participate for up to 6 years. In an evaluation of continuing effects, CPC students outperformed nonprogram students in reading and mathematics in Grades 3 and 5 and in mathematics in Grade 8 (Reynolds, 1997). CPC students were also less likely to be retained in grade, were likely to complete more years of education, and were less likely to drop out compared with nonprogram students (Reynolds et al., 2001).

What features of educational programs are likely to produce long-term benefits? Barnett (1995) summarized the evaluations of 36 early childhood programs and concluded that to have any long-term effects at all, school-age services "must actually change the learning environment in some significant ways" (p. 43). Ramey and Ramey (1998) identified six principles of program efficacy for early interventions. The most important principles are (a) developmental timing, that is, start early and continue; (b) program intensity, that is, the importance of many hours per day, days per week, and weeks per year of the intervention; and (c) direct provision of learning experiences rather than relying on intermediary sources—parent training alone, for example, is not likely to have an enduring impact on school performance.

The Perry Preschool Project and most Head Start programs in the evaluation were of limited intensity (see Zigler & Styfco, 1994). Although beginning at an early age (3 or 4 years), the Perry program lasted for 2 years, and most Head Start programs lasted for 1 or, at most, 2 years. Neither program engaged pupils for the full day; the Perry intervention involved about 2.5 hours of school time daily, and typical Head Start programs involve about 3.5 hours of class time, 4 or 5 days per week. When students leave programs such as Perry, CPC, or Head Start, they often enter half-day kindergartens targeted to nonaccelerated children. It comes as little surprise that early advantages are lost after several years in these settings.

Tennessee's Project STAR started early, beginning with full-day kindergartens. By Ramey and Ramey's (1998) definition, STAR was a high-intensity intervention. Children attended small classes for the entire school day every day of the school year, for up to 4 consecutive years. STAR impacted the learning setting directly and influenced all student–teacher interactions taking place in that setting. The present study asked if all-day, multiple-year participation in small classes in K–3 affected the likelihood of dropping out and if the effect on graduation rates was greater for lower than for higher SES students.

[5] All students returned to full-size classes in Grade 4.

Method

The STAR Sample

The sample for this investigation consisted of a subset of students who participated in Tennessee's Project STAR. Although STAR ended when students reached Grade 4, researchers continued to follow as many students as possible through high school. The investigators for this study collected high school transcripts for 5,335 STAR students in 165 schools, 4,948 of whom could be classified clearly as graduating or dropping out and who had achievement data from K–3.[6] When the high school information was unclear, the individual's status was confirmed through Tennessee State Education Department records.

A comparison of the entire STAR sample with the sample for this investigation is shown in Table 1. In general, the two samples had similar compositions. The sample for the present investigation had a somewhat lower percentage of minority students, but the percentage of students receiving free lunch was close to that of the full STAR sample. All demographic characteristics in Table 1 were included in the statistical analysis.

Measures

Student data. In addition to demographic information and the number of years of participation in small or full-sized classes, we computed two achievement composites for each student, one in mathematics and one in reading. Each composite was a principal component obtained from norm-referenced and criterion-referenced achievement tests administered in K–3. The Stanford Achievement Tests (StATs; Psychological Corporation, 1983) were administered to all STAR participants in the spring of each year. In addition, beginning in Grade 1, the Basic Skills First (BSF) tests, a set of curriculum-referenced tests developed by the Tennessee State Education Department, were also administered to each student. These were constructed from well-specified lists of objectives in reading and mathematics at each grade level. The number of objectives covered by a test ranged from 8 to 12 depending on the subject and grade level; a student was considered to have mastered an objective if she or he answered 75.0% of the items correctly. Our analyses used the number of objectives passed on each year's reading and mathematics tests.

For the present study, the reading composite score was the first principal component of four StAT Total Reading scores (Grades K, 1, 2, and 3) and three BSF Reading scores (Grades 1, 2, and 3). Similarly, the mathematics composite was the first principal component of four StAT Total Mathematics scores and three BSF Mathematics scores.[7] These composites accounted for 72.7% of variation in the seven reading tests and 71.0% of variation in the mathematics tests. Each composite had high positive correlations with the seven respective tests, all in the range from 0.73 to 0.92. The second component of each set accounted for less than 10.0% of the variance in the seven tests and had correspondingly low eigenvalues; thus, they were not used in our analyses.[8]

Students who did not participate in STAR for all 4 years were missing 1 or more years of test scores. Scores were imputed for those individuals prior to the principal component analysis using the expectation maximization (EM) method (Schafer, 1997) as implemented in the SPSS Missing Values program (Hill, 1997); this is superior to techniques such as listwise or pairwise deletion or mean substitution (Little & Rubin, 1990). The EM algorithm approach is especially useful in individual studies and for data sets for which the assumptions of data that are missing at random are not strictly met (Little & Schenker, 1995).

In the STAR data, as many as one third of the values were missing on some achievement variables (e.g., in kindergarten). However, we viewed the imputations as adequate for several reasons. For one, the correlations among the 14 reading and mathematics tests were consistently high, thus providing good information for estimating missing test scores, and the squared multiple correlations of each individual test with all others were uniformly strong.[9] Furthermore, a good set of covariates was added to the imputation process on which no values were missing, namely, school urbanicity and student sex, race/ethnicity, free-lunch participation, years in a small class, and years in Project STAR. These helped to adjust for the possibility that missing values were related to SES and student mobility.

After the imputation process, we conducted thorough checks on the reasonableness of the results; 45 students were eliminated who had test scores for just 1 year and imputed values well outside the distribution of observed scores on one or more tests.[10] Also, for each subject area, we correlated the first principal component computed using the original test scores (before the missing values analysis) with the first principal component computed using test scores after the missing values analysis. For both reading and mathematics, the correlation was above 0.99.

School data. Two characteristics of the high schools attended by participants in the study were also examined, total enrollment and school urbanicity. Schools were identified as suburban, rural, or inner city; we created two dummy variables to compare suburban schools with inner-city schools and rural schools with inner-city schools, respectively.

Analyses

The basic model used in the analysis was a logistic regression model for multilevel data, using the HLM5 program (Raudenbush, Bryk, Cheong, & Congdon, 2000). The first level of data comprised students, nested within high schools (the second level). The dependent variable for all analyses was the dichotomous indicator of whether or not the student had graduated from high school.

Table 1
Characteristics of Project STAR Samples

Characteristic	Full Project STAR sample	Transcript sample
Number of students	11,601	4,948
Percentage graduating	—	77.5
Percentage male	52.9	49.8
Percentage minority[a]	36.9	31.6
Percentage free lunch	55.3	55.8
Percentage in small classes[b]	31.7	33.8

[a] Minority students were 98.7% African American. Nonminorities included White students and 0.4% Asian students.
[b] For 1 or more years.

[6] Graduation information was available for 4,993 students, but 45 were eliminated from the analysis due to inadequate K–3 achievement data. Of the 342 students who could not be classified definitively as graduates or dropouts, 7 were listed as deceased.

[7] Components were obtained from the correlation matrices.

[8] We also considered using Grade 3 tests alone, considering them to be a composite that reflected 4 years of learning. The correlations between Grade 3 achievement and the principal component scores were 0.91 in both subjects.

[9] In reading, the correlations ranged from 0.65 to 0.85 for StAT tests and from 0.46 to 0.55 for BSF tests. In mathematics, the correlations ranged from 0.80 to 0.86 for StAT tests and from 0.44 to 0.57 for BSF tests.

[10] Because these students never entered any subsequent analyses, they are not included in the subsample described in Table 1.

The analyses involved a set of computer runs addressing each of the three research questions. The first set addressed the effect of small-class participation on the likelihood of graduation (Question 1), the second set examined the relationship between academic achievement and likelihood of graduation (Question 2), and the third set included both small-class participation and early academic achievement (Question 3). Each set consisted of three computer runs to (a) test main effects alone, (b) test interactions above and beyond main effects, and (c) estimate strength-of-effect measures from a reduced model containing those effects found to be important and significant.

The variables in each analysis are listed in the Appendix. For analyses of class size (Questions 1 and 3), each student was coded as having attended small classes for 0, 1, 2, 3, or 4 years during Grades K–3. Four contrasts were tested to compare students who attended small classes for 1, 2, 3, or 4 years, respectively, with students who attended full-size classes for all 4 years.

All analyses also included student sex, student race/ethnicity, student participation in the free-lunch program, school enrollment, and two school-urbanicity contrasts. Race/ethnicity, free-lunch participation, and school urbanicity provided some control for SES and student mobility into and out of STAR schools or between one STAR school and another. Interactions between small-class participation and eligibility for free lunch and between small-class participation and race/ethnicity were tested in the interaction model (Step b).

For the analysis of academic achievement (Question 2), the reading and mathematics composite scores were used in place of the class-size variable, and the interactions of mathematics and reading with free-lunch and race/ethnicity replaced the interactions of class size with free-lunch and race/ethnicity. All other effects were the same. A combined model with class-type contrasts and achievement test scores was tested to address Question 3. This model included the interactions of race and free-lunch participation with class type and with the achievement tests.

All terms in the hierarchical linear modeling (HLM) models were treated as fixed effects except for the intercepts at the student and school level, which were treated as random. All student-level characteristics were centered around the school means. Although the sample sizes were large, an alpha level of .05 was used for tests of significance. The outcomes, if they occurred, would take place 7 to 9 years after Grade 3 and may have been difficult to detect.

Strength-of-effect measures were obtained from final regression models after eliminating nonsignificant main effects and interactions; they were computed holding constant all predictor variables that remained in the model. Because the outcome measure was dichotomous (graduate or drop out), the strength-of-effect measures were odds ratios. This is the common strength-of-effect measure for logistic regression and has a direct relationship to the logistic regression coefficients (odds ratio $= e^{\beta}$ where β is a specific regression weight; see Hosmer & Lemeshow, 2000).[11] The odds that a member of one group (e.g., White students) would graduate are the estimated percentage of Whites who graduate divided by the percentage who drop out. The odds for the second group (e.g., minority students) are the estimated percentage of minorities who graduate divided by the percentage who drop out. The odds ratio is the ratio of the two (White odds/minority odds). When the independent variable is numerical rather than categorical (e.g., mathematics achievement or reading achievement), the odds ratio is the change in odds associated with a one-standard-deviation change in the respective achievement scale.

Results

The percentage of all students who graduated from high school was 77.5% in the transcript sample. Graduation rates were higher for female students (81.8%) than for male students (73.1%), higher for White students (81.8%) than for minority students (67.9%), and higher for students who did not receive free lunches (83.4%) than for students who received free lunches (72.8%).

Table 2 shows the graduation rates for students who attended full-size classes or small classes for 1 or more years. Graduation rates (and academic achievement) increased monotonically with additional years in a small class. Furthermore, the benefit of 3 or 4 years in a small class was greater for free-lunch students than for non-free-lunch students. Indeed, after 4 years in a small class, the graduation rate for free-lunch students was as great as or greater than that for non-free-lunch students. These effects were tested for significance in the regression analyses.

Table 2 also shows that the graduation rates of students in full-size classes were higher than those of students who spent 1 year in a small class. This may be due to the fact that students who attended small classes for 1 year were more transient than others; their families most likely moved into or out of the school's catchment area during the STAR years. In contrast, the full-size class group included the whole range of transience, including none. On average, the full-size group was less transient than groups who had 1 (or 2) years in small classes, and the lower graduation rate for 1 year in a small class may reflect transience as well as class size.[12]

The results for background demographic characteristics of schools and students (sex, race/ethnicity, free lunch) were consistent across the three sets of analyses, whether class size was the main independent variable (Question 1), academic achievement was the main independent variable (Question 2), or both (Question 3). (See Table 3.)

With respect to school characteristics, graduation rates were significantly higher in suburban and rural schools than they were in inner-city schools and were positively related to school size. With respect to student characteristics, female students had a significantly higher graduation rate than did male students (odds ratio = 1.67); in the sample, the difference was 8.7%. The difference between White and minority students was not statistically

[11] This is analogous to the way that effect-size measures are directly related to regression coefficients in ordinary least squares by dividing β by the standard deviation of the outcome variable.

[12] A rough approximation to transience rates was used to confirm this. School identifiers were not available for students when they were not attending a STAR school. However, we computed the number of years each student participated in Project STAR out of 4 possible (or 3 possible if the student began in Grade 1). A transience indicator was defined as 0 if the student participated in STAR for all 4 years (or 3 years if the student began in Grade 1) and 1 otherwise. Of students in full-size classes, 52.0% had made one or more school moves according to this indicator. Of 537 students in small classes for 1 year, 74.9% had made one or more school moves—clearly more than students in full-size classes. For students in small classes for 2, 3, and 4 years, the transience percentages were 70.5%, 20.7%, and 0.0%, respectively. Also, the actual differences in dropout rates were tested for significance in the regressions, both in the total sample and in each free-lunch group (see the *Analyses* section). None of the differences between 0 years in a small class and 1 or 2 years in a small class were statistically significant.

Table 2
Graduation Rates and Academic Achievement by Small-Class Participation

	Percentage graduating			Mean achievement score[a]	
Years in a small class	Free lunch	No free lunch	All	Reading	Mathematics
0 (full-size classes)	70.2	83.7	76.3	49.58	49.59
1	68.1	78.3	72.8	49.33	49.32
2	70.1	85.2	76.8	50.00	50.01
3	79.6	82.8	81.1	50.75	50.72
4	88.2	87.0	87.8	52.83	52.81

[a] Principal component scores plus constant (50). Standard deviations are 5.19 (Reading) and 5.18 (Mathematics).

significant.[13] Free-lunch status was significantly related to likelihood of graduation. In the sample, the graduation rate for non-free-lunch students was 10.6% higher than for free-lunch students; the odds ratio was 1.89.

Class size. In the class-size analysis (Question 1), neither 1, 2, nor 3 years in a small class was significantly different from full-size classes. However, 4 years in a small class was associated with a significantly higher graduation rate than attending full-size classes. Table 4 displays odds ratios for this effect, obtained from the final reduced model. The overall odds ratios were greater than 1.00 for 2, 3, and 4 years in a small class, ranging from 1.08 to 1.21 to 1.80. These figures show that more years in small classes had an increasing effect on the odds of completing high school.

None of the interactions of class size with race/ethnicity were statistically significant; the effect of small classes on graduation rates did not impact White and minority students differently. In contrast, several of the interactions of class size with free lunch were statistically significant, specifically those for 3 and 4 years in a small class. To examine this interaction further, we conducted tests of significance separately for free-lunch and non-free-lunch groups (simple main effects). Overall, there was a significant effect for participation in small classes for students eligible for free lunches, $\chi^2(4) = 30.34$, $p < .001$, but not for students who were not eligible, $\chi^2(4) = 4.08$, $p > .05$.[14]

Odds ratios for small-class participation were also computed separately by free-lunch status (Table 4 and Figure 1). The odds ratios for students not receiving free lunch do not differ significantly from 1.0. That is, for non-free-lunch students there were no significant differences in graduation rates based on participation in small classes. Among free-lunch students, however, the odds ratio for 3 years in a small class was large and significant, $t(2737) = 2.50$, $p < .02$, odds ratio = 1.67, and was larger still for 4 years, $t(2737) = 5.10$, $p < .001$, odds ratio = 2.49. The odds of graduating were 67.0% greater for students attending small classes for 3 years and almost 2.5 times greater for students attending small classes for 4 years. Table 2 shows that the difference between 3 years in a small class and attending full-size classes was associated with a 9.4% difference in graduation rates; for 4 years, it was 18.0%.

Achievement tests. The analysis of achievement scores showed that both reading achievement and mathematics achievement in K–3 were significantly, positively related to the likelihood of graduating from high school. Odds ratios were computed for a 5-point interval on the achievement measures; this is approximately equal to one standard deviation on the achievement composites (5.09 and 5.03, respectively).[15] The odds ratios were 1.32 for reading and 1.35 for mathematics. A one-standard-deviation increase in achievement in either area increased the odds of graduating from high school by about one third. The origins of high school graduation or dropping out can be seen clearly in academic achievement in the early grades.

When both class size and academic achievement were entered in a single HLM model (Question 3), the pattern of significant results was largely unchanged: Four years of small-class participation remained significantly superior to attending full-size classes, although at a lower level of significance, and the interaction of class size with free-lunch participation remained significant for the 3-year and 4-year contrasts. That is, after controlling for academic achievement in K–3, consistent small-class participation still increased the likelihood of graduating from high school. Stated another way, attending small classes for 3 or 4 years in the early grades had a positive effect on high school graduation above and beyond the effect on early academic performance.

Conclusions

This investigation addressed three questions about the relationships of early school experiences with graduating from or dropping out of school. One was a basic question about students' academic achievement in the primary grades. On the basis of a large sample of White and minority students followed from kindergarten through Grade 12, the analysis revealed a strong relationship between mathematics and reading achievement in K–3 and graduation from high school. The nonsignificant interactions indicated that these relationships held for White and minority students and for higher and lower SES students alike. The results are consistent with the findings of other research (Alexander et al., 1997; Barrington & Hendricks, 1989; Ensminger & Slusarcick, 1992; Garnier et al., 1997) and support theories that explain dropping out or

[13] As follow-up, we tested the significance of racial/ethnic differences in a model that did not include free-lunch status. The relationship was still nonsignificant.

[14] Tests were conducted using a Wald test of overall differences among the percentages.

[15] Also, if thirds of the scores distributions are viewed as achievement levels, the 5-point interval is the number of points required to move from the middle of one achievement level to the bottom of the next higher level.

Table 3
Logistic HLM Results

Independent variable	Class-size analysis	Achievement analysis	Class size and achievement
School level			
Urbanicity			
Suburban–inner city	0.54*	0.56*	0.56*
Rural–inner city	0.60**	0.73**	0.73**
Enrollment	0.001*	0.001*	0.001*
Student level			
Sex (M–F)	−0.58***	−0.52***	−0.51***
Race (White–minority)	0.05	−0.24	−0.22
Free-lunch status (nonfree–free)	0.66***	0.64***	0.65***
Years in small class			
1–none	−0.18		−0.21
2–none	0.05		0.00
3–none	0.22		0.07
4–none	0.68***		0.36*
Mathematics		0.06***	0.06***
Reading		0.06***	0.05***
Free-Lunch Status × Years in Small Class			
1–none	−0.18		−0.20
2–none	0.33		0.22
3–none	−0.73*		−0.83*
4–none	−0.74*		−0.87**
Free-Lunch Status × Mathematics		0.05	0.06*
Free-Lunch Status × Reading		−0.03	−0.02
Race × Years in Small Class			
1–none	−0.42		−0.25
2–none	0.12		0.22
3–none	0.05		0.06
4–none	0.10		0.13
Race × Mathematics		−0.04	
Race × Reading		0.05	

Note. All results are regression coefficients. Main-effects results from main-effect analyses (Step a). Interaction results from analysis (Step b). HLM = hierarchical linear modeling; M = male; F = female.
* $p < .05$. ** $p < .01$. *** $p < .001$.

withdrawing from school as a process that may begin in the early grades. The findings point once again to the need to identify and address learning and behavior problems at the earliest time feasible.

The other questions concerned the impact of small classes in the early grades on the likelihood of graduating from high school. The hypothesis of a positive long-term impact was predicated on research showing long-term effects of small classes on other outcomes and theory about the types of programs likely to have lasting benefits (e.g., Barnett, 1995; Ramey & Ramey, 1998).

Table 4
Odds Ratios for Small Class × Free-Lunch Participation

Years in small class	Free lunch	No free lunch	All
1	0.89	0.75	0.83
2	0.94	1.27	1.08
3	1.67	0.80	1.21
4	2.49	1.19	1.80

Note. Odds ratios computed from final reduced models; each odds ratio is the comparison of small-class participation to full-size classes.

The results support this hypothesis. For all students combined, 4 years in a small class in K–3 were associated with a significant increase in the likelihood of graduating from high school; the odds of graduating after having attended small classes for 4 years were increased by about 80.0%. Furthermore, the impact of attending a small class was especially noteworthy for students from low-income homes. Three years or more of small classes affected the graduation rates of low-SES students, increasing the odds of graduating by about 67.0% for 3 years and more than doubling the odds for 4 years. These findings are consistent with research showing that the immediate academic impact of small classes is greater for minority students and low-SES students (Finn & Achilles, 1990; Krueger & Whitmore, 2001) and that the percentage of minority students taking college entrance exams is increased by small-class participation (Krueger & Whitmore, 2001).

It is possible that the small-class effect on graduation rates is more far reaching than found in this investigation. Student mobility in the study may have played a role: Students who spent 1 or 2 years in small classes probably had higher mobility, on average, than students with 4 years of full-size classes. Thus, mobility may have decreased the graduation rates of 1- and 2-year groups

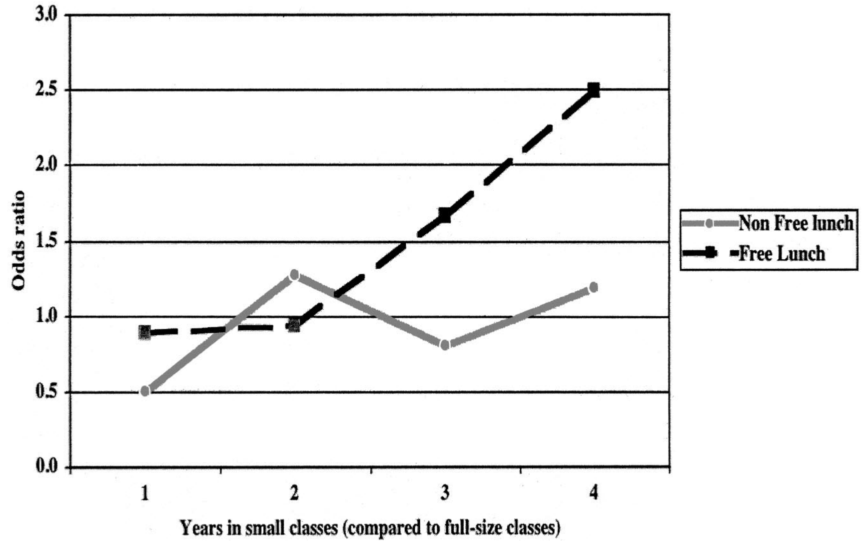

Figure 1. Odds ratios for years in small classes compared with years in full-size classes by free-lunch status.

relative to the comparison group, contributing to nonsignificant findings. Viewed in isolation, graduation rates did increase monotonically with additional years in a small class for all students combined and for each subgroup.

Finally, the long-term effects of small classes on dropout rates were not explained entirely by improvements in academic performance, even if the improvements carried through later grades. Other dynamics must have been occurring as well, for example, effects on students' attitudes and motivation, students' pro- or antisocial behavior, or students' learning behavior. The latter two have been termed social and academic engagement in school and have been posed as possible explanations in a recent review article (Finn et al., 2004). Further research is needed to understand the processes that connect early school experiences with long-term benefits. Even studies that have demonstrated connections between them have not given adequate consideration to the processes that lead from one to the other.

This study contradicts the argument that 1 year in a small class is sufficient to realize all the noteworthy benefits (Hanushek, 1999). Three or 4 years of small classes are needed to affect graduation rates, and 3 or 4 years have been found necessary to sustain long-term achievement gains (Finn et al., 2001). Our findings also raise a question about attempts to analyze the costs of small classes (e.g., Brewer, Krop, Gill, & Reichardt, 1999). To our knowledge, no cost analysis has weighed the benefits of small-class participation, which include increased high school graduation rates and increased aspirations to attend postsecondary school (Krueger & Whitmore, 2001).

This study did not ask whether the findings would be the same in locales with different populations or with particular programs to increase graduation rates. The results are in agreement with other research on academic achievement and dropping out, and the short-term impact of small classes on academic achievement has been found in other large-scale programs. However, the connections between early educational interventions—small classes among them—and long-range outcomes remain to be examined in other settings.

Furthermore, the theory about long-term impact and these findings raise the question, How can the magnitude of the effect be increased still further, perhaps by continuing small classes into later grades or by combining small classes with other educational interventions? Little if any research has examined the joint impact of reduced-size classes with programs such as an intensive reading curriculum in kindergarten (see Hanson & Farrell, 1995), full-day kindergartens, intensive preschool programs, or others. It seems that the potential for improved educational outcomes, especially among low-SES students, is considerable.

References

Achilles, C. M. (1999). *Let's put kids first, finally: Getting class size right.* Thousand Oaks, CA: Corwin Press.

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), 2–30.

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 meeting of the American Educational Research Association, San Francisco, CA.

Alexander, K. A., Entwisle, D. R., & Horsey, C. S. (1997, April). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*(2), 87–107.

Barnett, W. S. (1992). Benefits of compensatory preschool education. *Journal of Human Resources, 27,* 279–312.

Barnett, W. S. (1995, Winter). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children, 5*(3), 25–50.

Barrington, B. L., & Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts, and nongraduates. *Journal of Educational Research, 82,* 309–319.

Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalano, R. F., & Hawkins, J. D. (2000). Predictors of early high school dropout: A test of five theories. *Journal of Educational Psychology, 92,* 568–582.

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

Bloom, B. S. (1964). *Stability and change in human characteristics.* New York: Wiley.

Brewer, D., Krop, C., Gill, B. P., & Reichardt, R. (1999). Estimating the cost of national class size reductions under different policy alternatives. *Educational Evaluation and Policy Analysis, 21,* 179–192.

CSR Research Consortium. (2000). *Class size reduction in California: The 1998–99 evaluation findings.* Sacramento, CA: California Department of Education.

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

Ensminger, M. E., & Slusarcick, A. L. (1992, April). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*(2), 95–113.

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, CA.

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

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

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

Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record, 103,* 45–83.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2004). The "why's" of class size: Student behavior in small classes. *Review of Educational Research, 73,* 321–368.

Garnier, H. E., Stein, J. A., & Jacobs, J. K. (1997). The process of dropping out of high school: A 19-year perspective. *American Educational Research Journal, 34,* 395–419.

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

Goldschmidt, P., & Wang, J. (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal, 36,* 715–738.

Goldstein, H., & Blatchford, P. (1998). Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal, 24,* 255–268.

Hanson, R. A., & Farrell, D. (1995). The long-term effects on high school seniors of learning to read in kindergarten. *Reading Research Quarterly, 30,* 908–933.

Hanushek, E. A. (1998). *The evidence on class size.* Rochester, NY: University of Rochester, W. Allen Wallis Institute of Political Economy.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21,* 143–164.

Haskins, R. (1989). Beyond metaphor: The efficacy of early childhood education. *American Psychologist, 44,* 274–282.

Hedges, L. V., Nye, B., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 21,* 127–142.

Hedges, L. V., Nye, B., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37,* 123–151.

Hill, M. (1997). *SPSS missing value analysis 7.5.* Chicago: SPSS.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression.* New York: Wiley.

Kaplan, D. S., Peck, B. M., & Kaplan, H. B. (1997). Decomposing the academic failure–dropout relationship: A longitudinal analysis. *Journal of Educational Research, 90,* 331–343.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114,* 497–532.

Krueger, A. B., & Whitmore, D. (2001, March). *Would smaller classes help close the Black–White achievement gap?* (Princeton University Industrial Relations Section Working Paper No. 451). Retrieved January 18, 2005, from www.irs.princeton.edu/pubs/frame.html

Lazar, I., & Darlington, R. (1982). *Lasting effects of early education: A report from the Consortium for Longitudinal Studies.* Chicago: University of Chicago Press.

Lewit, E. M., & Baker, L. S. (1997, Summer/Fall). Class size. *Future of Children, 7*(2), 112–121.

Little, R. J. A., & Rubin, D. B. (1990). The analysis of social science data with missing values. In J. Fox & J. S. Long (Eds.), *Modern methods of data analysis* (pp. 374–409). Newbury Park, CA: Sage.

Little, R. J. A., & Schenker, N. (1995). Missing data. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 39–75). New York: Plenum Press.

McKey, R. H., Condelli, L., Granson, H., Barnett, B., McConkey, C., & Plantz, M. (1985) *The impact of Head Start on children, families, and communities.* Washington, DC: Head Start Bureau, U. S. Department of Health and Human Services.

Miles, K. H. (1995). Freeing resources for improving schools: A case of teacher allocation in Boston public schools. *Educational Evaluation and Policy Analysis, 17,* 476–493.

Molnar, A., Smith, O., & Zahorik, J. (1999). *1998–99 evaluation results of the Student Achievement Guarantee in Education (SAGE) program.* Milwaukee: University of Wisconsin School of Education.

Molnar, A., Smith, O., Zahorik, J., Ehrle, K., Halbach, A., & Kuehl, B. (2000). *1999–2000 evaluation of the Student Achievement Guarantee in Education (SAGE) program.* Milwaukee: University of Wisconsin School of Education.

National Research Council. (2001). *Understanding dropouts: Statistics, strategies, and high-stakes testing.* Washington, DC: National Academy Press.

Newmann, F. M., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. M. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11–39). New York: Teachers College Press.

Psychological Corporation. (1983). *Stanford Achievement Tests* (7th ed.). San Diego, CA: Author.

Ramey, C. T., & Ramey, S. (1998). Early intervention and early experience. *American Psychologist, 53,* 109–120.

Randolph, K. A., Fraser, M. W., & Orthner, D. K. (2004). Educational resilience among youth at risk. *Substance Use and Misuse, 39,* 747–767.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2000). *HLM5: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Reynolds, A. J. (1997). *The Chicago child–parent centers: A longitudinal study of extended early childhood intervention* (Discussion Paper No. 1126–97). Madison, WI: Institute for Research on Poverty.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001).

Long-term effects of an early childhood intervention on educational achievement and juvenile arrest. *JAMA, 285,* 2339–2346.

Robinson, G. E. (1990). Synthesis of research on effects of class size. *Educational Leadership, 47*(7), 80–90.

Rumberger, R. W. (2001, January). *Why students drop out of school and what can be done.* Paper presented at the "Dropouts in America: How Severe Is the Problem? What Do We Know About Intervention and Prevention?" conference, Cambridge, MA.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

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

White, K. R. (1986). Efficacy of early intervention. *Journal of Special Education, 19,* 400–416.

Word, E., Johnson, J., Bain, H. P., Fulton, D. B., Boyd-Zaharias, J., Lintz, M. N., et al. (1990). *Student/teacher achievement ratio (STAR): Tennessee's K–3 class-size study.* Nashville, TN: Tennessee State Department of Education.

Zigler, E., & Styfco, S. J. (1994). Is the Perry Preschool better than Head Start? Yes and no. *Early Childhood Research Quarterly, 9,* 269–287.

Appendix

Variables Included in HLM Analyses

Variable	Class size	Achievement	Class size and achievement
Student level			
Sex (0 = female, 1 = male)	X	X	X
Race/ethnicity (0 = minority, 1 = nonminority)	X	X	X
Free-lunch status (0 = eligible, 1 = not eligible)	X	X	X
Years in small classes[a]			
One–none	X		X
Two–none	X		X
Three–none	X		X
Four–none	X		X
Mathematics achievement[b]		X	X
Reading achievement[b]		X	X
Free Lunch × Years in Small Classes[a]	X		X
Race/Ethnicity × Years in Small Classes[a]	X		X
Free Lunch × Mathematics Achievement[b]		X	X
Free Lunch × Reading Achievement[b]		X	X
Race/Ethnicity × Mathematics Achievement[b]		X	X
Race/Ethnicity × Reading Achievement[b]		X	X
School level			
School enrollment	X	X	X
Urbanicity			
Suburban–inner city	X	X	X
Rural–inner city	X	X	X

Note. HLM = hierarchical linear modeling.
[a] Four contrasts. [b] Achievement measures are numerical composite scores.

Received July 29, 2004
Revision received December 1, 2004
Accepted December 4, 2004 ■