EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Volume 12 Number 58 October 22, 2004 ISSN 1068-2341

School Size, Achievement, and Achievement Gaps

Bradley J. McMillen
North Carolina Department of Public Instruction

Citation: McMillen, B. J. (2004, October 22). School size, achievement, and achievement gaps. *Education Policy Analysis Archives, 12*(58). Retrieved [date] from http://epaa.asu.edu/epaa/v12n58/.

Abstract

In order to examine the relationship between school size and achievement, a study was conducted using longitudinal achievement data from North Carolina for three separate cohorts of public school students (one elementary, one middle and one high school). Results revealed several interactions between size and student characteristics, all of which indicated that the achievement gaps typically existing between certain subgroups (i.e., more versus less-advantaged, lower versus higher-achieving) were larger in larger schools. Results varied across the grade level cohorts and across subjects, but in general effects were more common in mathematics than in reading, and were more pronounced at the high school level. Study results are discussed in the context of educational equity and cost-effectiveness.

Introduction

Concerns about school size in the educational research literature tend to center on high schools. The most common sentiment expressed is that high schools are too large, and that they are getting larger. The U.S. Department of Education (2000a) reports that the number of public schools serving the secondary grades in the U. S. has largely held steady between 23,000 and 26,000 since 1930. During that same time, however, the number of public high school students in the U. S. nearly tripled, from approximately 4.4 million to over 13 million.

As consolidation trends have created larger schools, the issue of school size has become of great interest to educators and policymakers alike. As the demand for safer schools, the need to help all students reach high achievement standards, and the proliferation of school-level monitoring and accountability systems have increased, so has interest in the contribution of many school-level variables – including school size – to student outcomes. Intuitively, school size would appear to have considerable impact on both student achievement and discipline in the school. Smaller size is often associated with more personal attention, more opportunities for involvement, less anonymity for students, and a more caring environment. These factors are then hypothesized to lead to more positive student outcomes (Finn, 1989; Holland & Andre, 1987). Larger schools, however, are said to offer a broader and deeper curriculum along with economies of scale that often appeal to policymakers.

Studies of student behavior indicate that smaller schools are generally associated with more positive behavioral outcomes for students. Larger schools are reported to have higher dropout and expulsion rates than smaller schools (Fetler, 1989; Fowler & Walberg, 1991; Pittman & Haughwout, 1987; Schoggen & Schoggen, 1988). Larger schools also have been shown to have more problems with most major behavioral issues including truancy, disorderliness, physical conflicts among students, robbery, vandalism, alcohol use, drug use, sale of drugs on school grounds, tobacco use, trespassing, verbal abuse of teachers, teacher absenteeism, and gangs (Haller, 1992; Heaviside, Rowand, Williams, & Farris, 1998; Lindsay, 1982; Page, 1991). There is also a substantial body of research that indicates that students in smaller schools are more likely to be involved in extracurricular activities (Baird, 1969; Barker & Gump, 1964; Grabe, 1981; Lindsay, 1982; Morgan & Alwin, 1980; Schoggen & Schoggen, 1998).

School size has also been studied in relationship to student achievement, at both the elementary and high school levels. The majority of studies at the elementary level point toward an inverse relationship, i.e., smaller elementary schools tend to have higher achievement. For example, a study in New York found that reading and math test scores were higher in elementary schools with smaller enrollments, even after controlling for socioeconomic factors (Kiesling, 1968). Caldas (1993) found a small negative relationship between school size and general achievement among elementary schools in Louisiana. Wendling and Cohen (1981) also found that third graders from smaller schools demonstrated higher achievement in reading and math than their counterparts in larger schools. In that study, the average enrollment in the lower-achieving schools was 776, while the average enrollment of the higher-achieving schools was 447. Fowler (1995) reviewed a number of studies of school size and achievement in elementary schools, all of which again suggested a negative relationship. Several of the studies Fowler reviewed, however, were not widely published or were not published at all. Even so, there is little contrary evidence in the educational research literature to refute the conclusion that smaller elementary schools are associated with higher achievement.

Although the findings for elementary schools would appear fairly consistent, the research on high school size and achievement is less conclusive. Using state achievement test data from 293 public high schools in New Jersey, Fowler and Walberg (1991) found that school size was inversely related to test scores in mathematics and writing. They also found that smaller schools were associated with higher passing rates on the reading portion of the state's Minimum Basic Skills Test as well as on the mathematics and writing portions of the state's High School Proficiency Test. These effects were statistically significant even after controlling for students' family income level, but the actual size of the effects was not clearly reported. The schools in this study had enrollments ranging from 147 to 4,018, with an average enrollment of 1,070.

Other studies have demonstrated similar results. Fetler (1989), in a study of all public high schools in California, found that schools with smaller enrollments tended to have higher achievement scores, although the relationship was not strong and the analysis did not take into account any student background factors. Walberg and Walberg (1994) used data from the 1990 National Assessment of Educational Progress (NAEP) mathematics assessment to examine relationships among size, expenditures and achievement. Their analyses demonstrated that states with larger schools tended to score lower on the NAEP mathematics assessment, even after controlling for per-pupil expenditures and percentage of non-Caucasian students in the state.

Other studies, however, have failed to demonstrate higher levels of achievement for smaller high schools. Lindsay (1984), analyzing data from a nationally representative sample of almost 14,000 high school students, found no meaningful relationship between school size and academic ability. Academic ability in this study was measured by a standardized composite score based on four tests (vocabulary, reading, inductive reasoning, and mathematics) that were used in the National Longitudinal Study conducted by the U. S. Department of Education. A study by Jewell (1989) reached similar conclusions. In examining the relationship between school size and college entrance exam scores across all 50 states and the District of Columbia, he found no significant relationship between high school size and either ACT scores or Scholastic Achievement Test (SAT) scores after controlling for poverty. In another earlier study, Baird (1969) analyzed data from over 21,000 high school students who took the American College Test (ACT) and found that students from smaller schools actually had lower ACT scores. Haller, Monk, and Tien (1993) also found no relationship between high school size and higher-order thinking skills using data from a nationally-representative sample of 10th graders from the Longitudinal Study of American Youth. Compared to the results for elementary schools, the evidence for the size-achievement relationship at the high school level appears to be more mixed.

One of the more sophisticated studies of size and achievement found that students from medium-sized high schools actually demonstrated higher achievement than students in either smaller or larger schools (Lee & Smith, 1997). Using longitudinal data from a nationwide sample of over 9,000 students, the authors studied the relationship between size and achievement gains between 8th grade and 12th grade. The results indicated that after controlling for various student-level and school-level demographic characteristics, students in moderate-sized high schools tended to have higher gains in both reading and mathematics, with the effects for mathematics being somewhat stronger than for reading. Specifically, they found that the highest gains in achievement were found in high schools with enrollments between 600 and 900 students. In addition, the finding of lower mathematics gains in larger schools was especially pronounced for non-Caucasian students and students from lower socioeconomic backgrounds. The Lee and Smith study is also one of the few studies in this area to control for prior achievement.

A recent reanalysis of this same dataset, however, by Howley and Howley (2004) has questioned Lee and Smith's conclusions regarding optimal size, contending in particular that the effects of very small schools were not adequately addressed in the analysis. They concluded that the relationship between size and achievement is in fact more linear and that smaller size (less than the 600-student cutpoint posited by Lee and Smith) does in fact benefit students from lower socioeconomic backgrounds (see Lee (2004) for a critique of this reanalysis and its conclusions).

The interaction between poverty and size was also echoed in a report by the Rural School and Community Trust (Howley & Bickel, 1999) using data from 13,600 public schools in 2,290 districts in Georgia, Montana, Ohio, and Texas. Specifically, schools in less affluent communities in each state demonstrated higher achievement if they were smaller, while the opposite relationship was found in more affluent communities. Howley and colleagues have labeled this phenomenon the "excellence effect" of small school size, and have also demonstrated this result across grade levels using data from other states including West Virginia (Howley, 1995) and Arkansas (Johnson, Howley, & Howley, 2002). This line of research has also forwarded the notion of an "equity effect" of size, showing that the ubiquitous poverty-achievement correlation is much stronger in larger schools and districts than in smaller schools and districts (e.g., Bickel & Howley, 2000; Friedkin & Necochea, 1998).

Overall, school size appears to be related to a host of behavioral and academic outcomes for students, with smaller schools being associated with more positive outcomes in most cases. The research on size and achievement at the high school level appears to be somewhat of an exception, however, with multiple studies reaching different conclusions. In addition, both the Lee and Smith (1997) study and the series of studies by Howley and colleagues point toward the possibility that school size may be associated with different outcomes for students from different backgrounds. Many prior studies, however, have failed to control for prior achievement, have not explored the possibility of differential effects for subgroups of students, and/or have not been able to analyze student-level variables in conjunction with school-level effects. These issues, in conjunction with the federally-driven focus on disaggregated achievement results and progress monitoring, call for further investigation of how the size-achievement relationship may operate among specific types of students.

In an effort to better understand how school size relates to achievement among different subgroups of students across various grade levels, a study was undertaken to examine these relationships using data from the North Carolina public schools. North Carolina provides a particularly interesting venue to study this issue due to the wide ranges in the size of schools across the state, a relatively high average school size (Figure 1), and the availability of longitudinal achievement test data for individual students from the state's testing and accountability program. Two primary research questions were formulated to guide the overall study:

1. What are the relationships between school size and achievement at the elementary, middle and high school levels?
2. Do size-achievement relationships vary among students with differing levels of prior achievement, students of different ethnicities, and students whose parents have different levels of education?

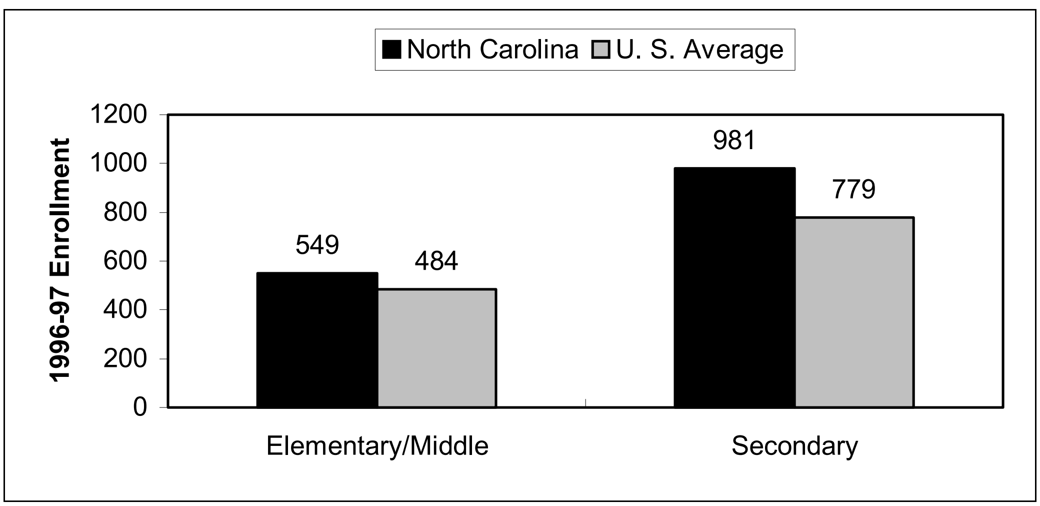

Figure 1. Average Enrollment in North Carolina and U. S. Public Schools, 1997-1998.

Note. Elementary/middle schools are defined as a school in which the lowest grade is no higher than 6 and the highest grade is 8 or lower. Secondary schools are defined as schools in which the lowest grade is no lower than 7. Vocational schools, alternative schools, special education schools, and other schools not reported by grade level are excluded. From U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data.

Method

Student Achievement Data

Data for the study were gathered from several databases maintained by the North Carolina Department of Public Instruction, including reading and mathematics end-of-grade (EOG) testing databases and databases for the state's High School Comprehensive Test (HSCT). EOG tests in reading and mathematics are administered each spring to most North Carolina public school students in grades 3-8[1]. The HSCT is a test of reading and mathematics administered to 10th graders in the North Carolina public schools each spring. These databases also contain a variety of demographic information as well as codes identifying the school attended by each tested student.

Using these databases, three separate cohort samples were constructed. The elementary cohort consisted of all tested 3rd graders from the 1996-97 school year, the middle school cohort consisted of all tested 6th graders in 1997-98, and the high school cohort consisted of all tested students in grade 8 in 1997-98. Each student's achievement data for that school year was then linked to their achievement test scores in the same subject areas two years later (Table 1). Students were included in the final samples only if they a) had available test data in at least one of the two subjects for both the baseline year and two years later; b) had made regular progress from grade to grade between the baseline year and two years later (i.e., were not retained and did not skip any grades); and c) attended the same school for both years following the baseline year.

Table 1
Reading and mathematics test data used for elementary, middle, and high school cohorts

	Grade							
	3	4	5	6	7	8	9	10
Elementary	Pre	→	Post					
Middle				Pre	→	Post		
High						Pre	→	Post

School-Level Data

Data on school size was obtained from state student membership databases. The average daily membership for each school was used as an indicator of school size. Average daily membership is calculated as the number of students officially listed on the daily roster of each school averaged across the entire school year. Since the study covered a two-year span, the size data for each school was averaged across the two-year period to produce a two-year average school size estimate. These data were then appended to the records in the cohort sample datafiles. In cases where schools either closed or where schools gained or lost large numbers of students between the two years due to consolidations, closings, or redistricting, data for those students were filtered out

---

[1] During the years from which the data were drawn for this study, some students who were exempt from being tested based on limited English proficiency status and some special education students who were exempt based on recommendations in their Individualized Education Plans may not have been tested in one or more subject areas.

prior to conducting the analysis. This resulted in the elimination of a very small number of cases (less than 2%) in each of the three analysis datasets. Data on the percentage of students eligible for free or reduced price lunch was also obtained for each school from extant state databases.

Preliminary analyses prior to the calculation of the actual results included screening for univariate and various bivariate outliers and other unusual conditions in the data that may have adversely impacted results (e.g., test scores beyond the range of possible scores, duplicate testing records for the same students in the same year, etc.). These screening analyses resulted in the deletion of a very small number of individual records due to anomalies that could not be reconciled. Characteristics of the students and schools in the final analysis samples are presented in Tables 2 and 3, respectively. Preliminary regression models also indicated some collinearity problems involving the continuous school size predictor. These conditions were ameliorated by creating a four-level categorical school size variable (divided at the quartiles within each cohort grade level configuration) and using it as the indicator of size. Results generated using the four-level school size predictor were not substantively different from those using the continuous predictor; therefore the results reported here are those using the four-level categorical school size variable.

Table 2
Sample characteristics - Students

		Elementary		Middle		High	
		n	%	n	%	n	%
Gender	Male	27,235	50	26,128	49	30,161	49
	Female	27,380	50	27,178	51	28,625	51
Ethnicity	Caucasian	37,808	69	37,129	70	40,861	70
	African-American	14,022	26	13,421	25	14,874	25
	Other	2,785	5	2,756	5	3,051	5
Parent(s) Highest Education Level	Less than HS	4,609	8	3,642	7	3,349	6
	HS Graduate	21,129	39	19,350	36	15,469	26
	Some College	4,719	9	4,886	9	4,787	8
	2-Year Degree	7,955	15	8,813	17	13,294	23
	4-Year Degree	12,913	24	12,938	24	15,141	26
	Graduate Degree	3,290	6	3,677	7	6,746	11
Total		54,615		53,306		58,786	

Note. Student demographic data are from the samples in the Reading analysis. Because a small number of students took only one test or the other due to various exemptions, there are negligible differences between the reading analysis and mathematics analysis samples at each level. Percentages may not add to 100 due to rounding.

Table 3
Sample characteristics - Schools

	Elementary (n = 1,053)		Middle (n = 508)		High (n = 333)	
	Mean	Range	Mean	Range	Mean	Range
Number of students (2-year average)	506	26 - 1,392	570	21 - 1,508	859	27 - 2,352
% of students eligible for free/reduced price lunch	48	0 - 99.7	44	0 - 97.5	30	0 - 94.5

Covariates

In all three sets of analyses, both student-level and school-level variables known to be correlate with student achievement were included as covariates in order to get a more precise estimate of the relationship between school size and achievement. The student-level covariates included gender, ethnicity, the highest level of education for the parent(s) in the home, and the student's prior achievement status (at/above grade level or below) in the same subject[2]. The percentage of students in the school who were eligible for free or reduced price lunch was used as a school-level covariate. In the high school mathematics analyses, indicators of whether each student had taken Algebra and Geometry were also used as covariates to help control for differential coursetaking experiences.

Analysis Procedures

Analyses of each of the three cohort datasets involved the estimation of two-level hierarchical linear models. This approach allows for the proper estimation of effects when units of analysis (e.g., students) are nested within a larger contextual unit (e.g., schools). Adjusting for this nesting allows for proper error estimation as well as the inclusion of both student-level and school-level predictor variables and their interactions in the models. Traditional least-squares regression methods in a multilevel context require either aggregating data to the school level prior to analysis, which results in a loss of statistical power and precision, or disaggregating school-level data down to the individual student level, which often results in spuriously significant results that show relationships between variables which may not truly exist (Hox, 1995). Hierarchical linear modeling methods avoid both of these problems by properly incorporating both school-level and student-level factors in the same analysis (Bryk & Raudenbush, 1992; Singer, 1998). In each case, initial null models were generated as a baseline, with predictors added one by one and level by level, to check for and ameliorate any unusual or problematic conditions in the data that may have hindered interpretation of the results. With respect to the overall models, prior achievement and the other student-level covariates accounted for a notable portion of the explainable between-school and

[2] Dichotomized versions of the ethnicity and parent education level variables were employed in lieu of their original forms because of the uneven distribution of the ethnicity variable in the sample and because of reliability concerns with parent education level data that are reported by the classroom teacher.

within-school variation in achievement in many of the models. These indicators of explained variance are not analogus to a squared multiple correlation in linear regression, however, and should not be interpreted as such. They are merely relative indicators of the proportion of school-level variation and student-level variation that is explained by the variables in the model (Snijders & Bosker, 1994).

Given prior studies in this area which have found a curvilinear relationship between size and achievement (e.g., Lee & Smith, 1997), all models were initially estimated with both linear and nonlinear terms for school size. However, in every case, better model fit was achieved using only the linear term, indicating that these data did not support a curvilinear relationship between size and achievement. The nonlinear terms were therefore omitted from the final models.

Results

Elementary Cohort

For the elementary cohort, the reading achievement analyses yielded no statistically significant relationship between school size and achievement after controlling for school and student demographic characteristics (Table 4). As was expected, higher 3rd grade achievement scores, female gender, White ethnicity, and higher levels of parent education were all associated with higher EOG scores at the end of grade 5. Attending a school with a lower percentage of students eligible for free or reduced price lunch was also associated with higher achievement (Table 4). The size-prior achievement interaction implied that there was a negative size-achievement relationship, but that it was predominantly seen among students who were below grade level in 3rd grade (Figure 2). The size of the effect, however, was rather small (.06 SD[3]). No statistically significant interactions were found between size and ethnicity or size and parent education level; therefore, those terms were dropped from the final model.

In the mathematics analyses, there was no significant main effect for size. There was, however, another significant size-prior achievement interaction (Table 5). This interaction indicated that students who were below grade level in mathematics based on their 3rd grade scores scored better in grade 5 if they attended smaller schools, whereas the pattern for students who scored above grade level in grade 3 was more uneven (Figure 3). The actual magnitude of this interaction, as in the reading model, was rather small (.09 SD). The pattern of relationships for the student and school-level covariates in the mathematics model mirrored that of the reading model, with the exception of the gender variable, which was not significant and subsequently dropped from the final model (Table 5).

---

[3] The expression of effect sizes in standard deviation units in these analyses represents the difference in scale score gaps between students in the smallest school size quartile and the largest. The standard deviation estimate used for these calculations is the statewide standard deviation for each test (North Carolina Department of Public Instruction, 2000; 2001).

Table 4
Elementary cohort 2-level HLM regression model – reading

Level	Variable	b	SE	t	F
Student-level	Prior Reading Achievement - 3rd Grade	10.13	.11	89.41*	
	Gender[a]	-.49	.05	-8.99*	
	Ethnicity[b]	2.33	.07	34.10*	
	Parent Education Level[c]	2.83	.06	48.91*	
School-level	School Size[d]				2.18
	Less than 400 students	.05	.17	.26	
	400-549 students	.12	.16	.74	
	550-699 students	-.05	.16	-.28	
	Free/Reduced Price Lunch (% eligible)	-.02	.003	-5.89*	
Interactions	Size x Prior Achievement				4.70*
Summary	% between-student variation explained				44.8%
	% between-school variation explained				66.2%

[a] 0 = female, 1 = male. [b] 0 = non-White, 1 = White. [c] 0 = high school diploma or less, 1 = at least some post-secondary education. [d] Reference group for school size is 700+ students.
**p* < .05.

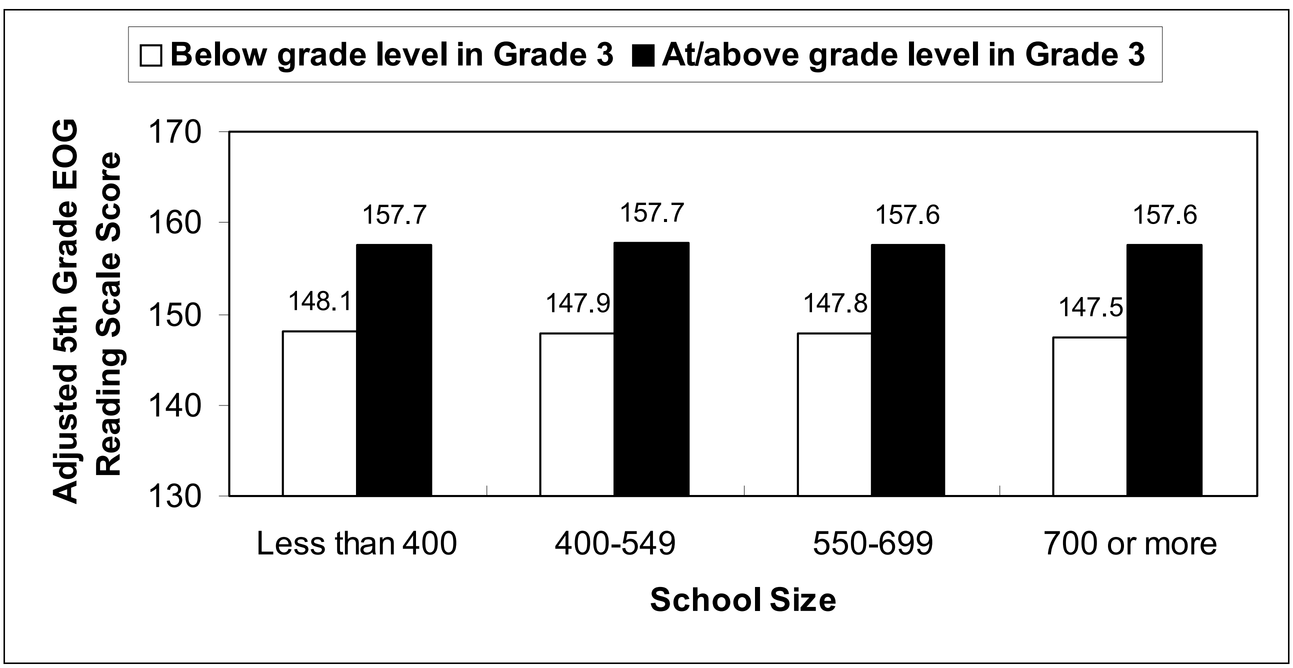

Figure 2. Elementary cohort reading model - interaction between prior achievement and school size.

Table 5
Elementary cohort 2-level HLM regression model - mathematics

Level	Variable	b	SE	t	F
Student-level	Prior Mathematics Achievement - 3rd Grade	11.70	.18	64.65*	
	Ethnicity[a]	2.91	.09	32.61*	
	Parent Education Level[b]	3.63	.07	48.92*	
School-level	School Size[c]				.51
	Less than 400 students	-.26	.27	-.96	
	400-549 students	-.09	.25	-.36	
	550-699 students	.18	.26	.70	
	Free/Reduced Price Lunch (% eligible)	-.01	.004	-3.44*	
Interactions	Size x Prior Achievement				5.60*
Summary	% between-student variation explained				41.5%
	% between-school variation explained				44.2%

[a] 0 = non-White, 1 = White. [b] 0 = high school diploma or less, 1 = at least some post-secondary education. [c] Reference group for school size is 700+ students.
**p* < .05.

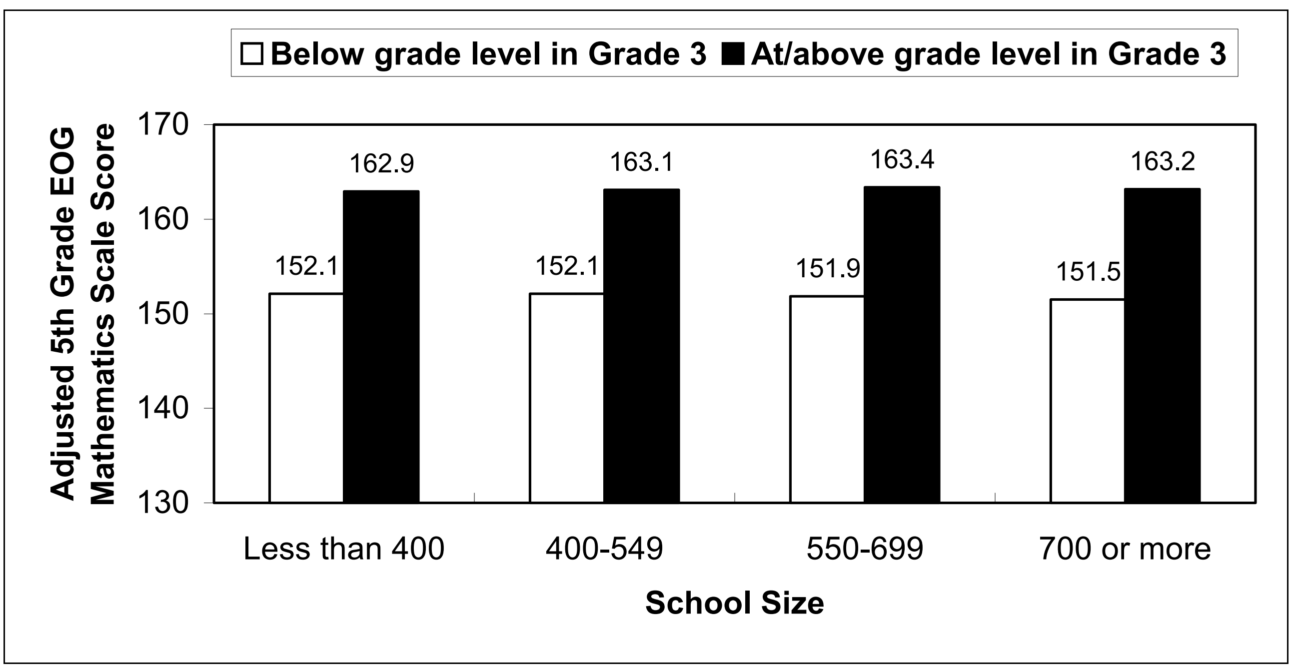

Figure 3. Elementary cohort mathematics model - interaction between prior achievement and school size.

Middle School Cohort

Similar to the elementary results, the reading achievement analyses for middle school students also yielded no overall relationship between school size and achievement after controlling for various school and student demographic characteristics (Tables 6 & 7)[4]. The pattern of relationships found for the covariates was also identical to the elementary cohort analyses with one exception: Male students in the middle school cohort demonstrated slightly higher achievement in mathematics than their female counterparts, which was not the case in the elementary mathematics analysis.

As was true at the elementary level, the middle school models also yielded significant interactions between prior achievement and school size for both reading and mathematics. Students who were scoring on grade level in 6th grade tended to do slightly better in larger middle schools over the next two years, whereas students who were below grade level in 6th grade did slightly better in smaller schools (Figures 4 & 5). Although larger in comparison to the elementary results, the interactions again were not overwhelming (.12 SD for reading and .13 SD for mathematics). Interactions between school size and ethnicity and school size and parent education level were nonsignificant in both the reading and mathematics models and those terms were therefore dropped.

Table 6
Middle school cohort 2-level HLM regression model - reading

Level	Variable	b	SE	t	F
Student-level	Prior Reading Achievement – 6th Grade	10.04	.18	57.01*	
	Gender[a]	-.75	.05	-14.65*	
	Ethnicity[b]	2.22	.07	30.33*	
	Parent Education Level[c]	2.72	.07	41.72*	
School-level	School Size[d]				2.25
	Less than 400 students	-.45	.21	-2.14*	
	400-549 students	-.35	.20	-1.79	
	550-699 students	-.46	.20	-2.30*	
	Free/Reduced Price Lunch (% eligible)	-.01	.003	-4.28*	
Interactions	Size x Prior Achievement				5.68*
Summary	% between-student variation explained				44.8%

[4] Although the coefficients associated with the dummy-coded size variable did indicate that achievement was slightly higher in the largest size category compared to some smaller categories using a standard p level of .05, the overall F test was non-significant. Also, since these t tests largely amount to non-orthogonal a posteori contrasts, a familywise .05 error rate per model is a preferable standard (Kirk, 1995). Using this standard of .0167 (.05 divided by 3), none of the specific size contrasts in the middle school models would have reached significance.

	% between-school variation explained	63.1%

[a] 0 = female, 1 = male. [b] 0 = non-White, 1 = White. [c] 0 = high school diploma or less, 1 = at least some post-secondary education. [d] Reference group for school size is 700+ students. *$p < .05$.

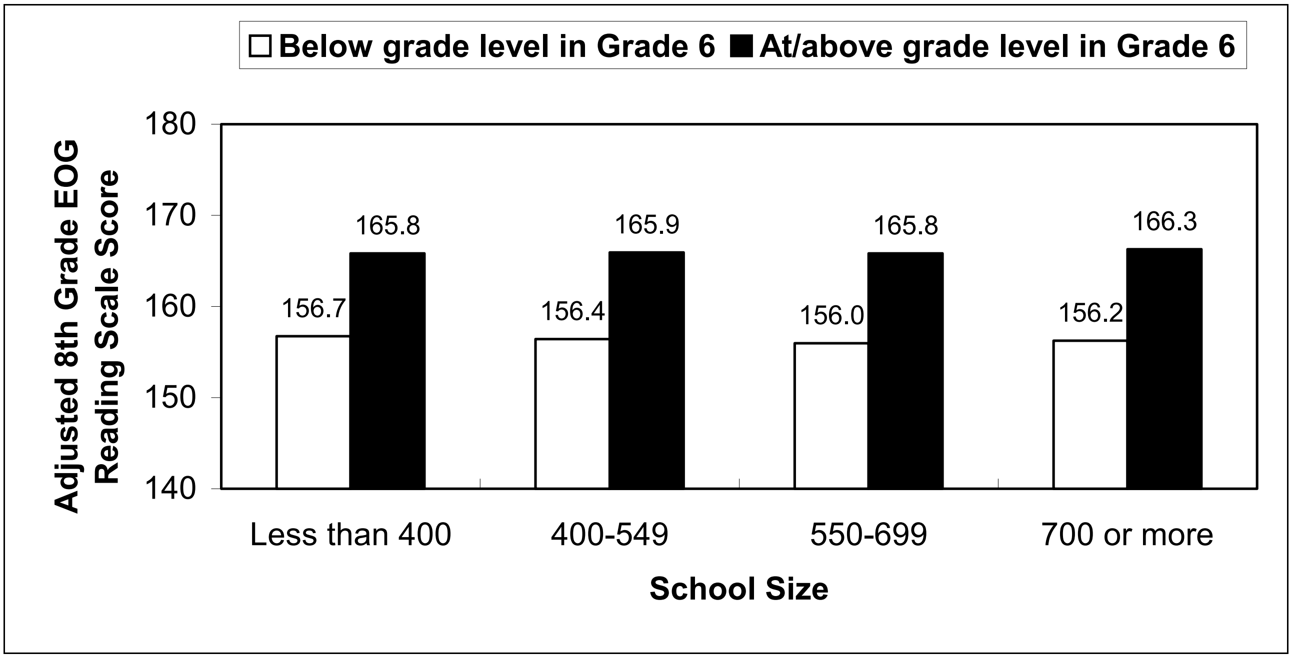

Figure 4. Middle school cohort reading model - interaction between prior achievement and school size.

Table 7
Middle school cohort 2-level HLM regression model - mathematics

Level	Variable	b	SE	t	F
Student-level	Prior Mathematics Achievement – 6th Grade	15.09	.30	-50.75*	
	Gender[a]	.15	.08	1.98*	
	Ethnicity[b]	4.16	.12	35.15*	
	Parent Education Level[c]	4.77	.12	38.87*	
School-level	School Size[d]				1.01
	Less than 400 students	-.91	.39	-2.34*	
	400-549 students	-.69	.37	1.86	
	550-699 students	-.89	.38	-2.33	
	Free/Reduced Price Lunch (% eligible)	-.02	.006	3.27*	
Interactions	Size x Prior Achievement				6.58*
Summary	% between-student variation explained				36.7%

% between-school variation explained	50.1%

[a] 0 = female, 1 = male. [b] 0 = non-White, 1 = White. [c] 0 = high school diploma or less, 1 = at least some post-secondary education. [d] Reference group for school size is 700+ students.
$^{*}p < .05$.

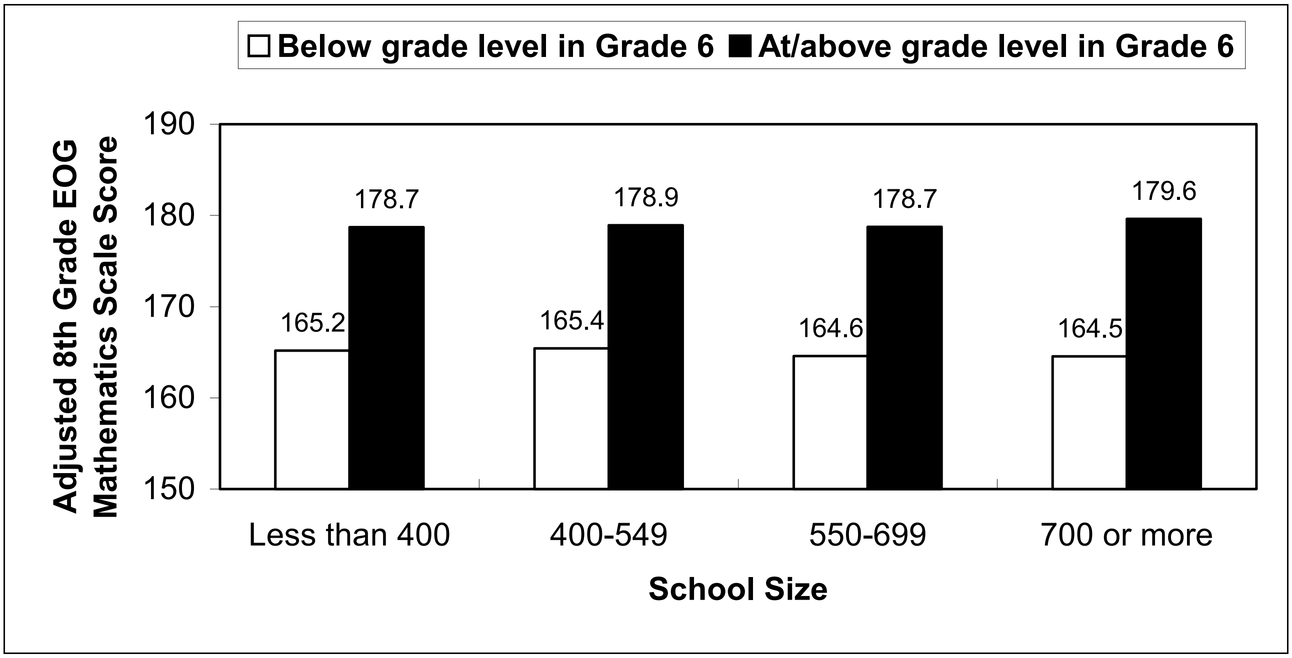

Figure 5. Middle school cohort mathematics model - interaction between prior achievement and school size.

High School Cohort

The high school analyses yielded the largest number of relationships between school size and achievement as well as the largest relationships in terms of effect size. In the reading model, there was a significant and positive main effect for size, along with statistically significant interactions involving size and ethnicity and size and parent education level (Table 8). Taken together, these relationships implied that while students overall performed better in Reading in larger high schools, the benefits accrued more strongly to White students and students whose parents had at least some post-secondary education. Non-White students and students whose parents had a high school education or less showed a more "U-shaped" pattern of performance, with scores being roughly equal in the smallest and largest schools (Figure 6). The size of these interaction effects were as large or larger than any of the interactions found at the middle and elementary levels (.12 SD for the size-parent education level interaction, and .20 SD for the size-ethnicity interaction). The pattern of relationships for the other student and school-level covariates mirrored that of the reading models in the elementary and middle school analyses.

As in the high school reading model, the high school mathematics model also yielded a positive main effect for school size. Significant interactions were also found between size and prior achievement, size and ethnicity, and size and parent education level (Table 9). These interactions, as in previous analyses, again indicated that the benefit of larger school size again accrued disproportionately to students whose prior achievement was higher (.28 SD; Figure 8), White students (.10 SD; Figure 9), and students whose parents had at least some education beyond high school (.11 SD; Figure 10). The size-prior achievement interaction in the high school mathematics model was the largest found in the study. Relationships for other student and school-level covariates were similar to those found in the elementary and middle school models. It should also be noted that indicators representing students' exposure to algebra and geometry courses through grade 10 were also available in the extant database to be used as covariates in the high school mathematics model, thereby controlling for course-taking factors that were not measurable at the elementary and middle school levels. Students who had taken these courses, as expected, demonstrated higher achievement.

Table 8
High school cohort 2-level HLM regression model - reading

Level	Variable	b	SE	t	F
Student-level	Prior Reading Achievement – 8th Grade	11.35	.12	92.49*	
	Gender[a]	-1.89	.08	-23.62*	
	Ethnicity[b]	4.23	.30	14.17*	
	Parent Education Level[c]	3.90	.27	14.33	
School-level	School Size[d]				5.78*
	Less than 700 students	-2.58	.37	-6.96*	
	700-1,199 students	-1.90	.33	-5.72*	
	1,200-1,699 students	-.91	.36	-2.52*	
	Free/Reduced Price Lunch (% eligible)	-.01	.005	-2.09*	
Interactions	Size x Ethnicity				10.18*
	Size x Parent Education Level				7.04*
Summary	% between-student variation explained				32.5%
	% between-school variation explained				66.7%

[a] 0 = female, 1 = male. [b] 0 = non-White, 1 = White. [c] 0 = high school diploma or less, 1 = at least some post-secondary education. [d] Reference group for school size is 1,700+ students. *p < .05.

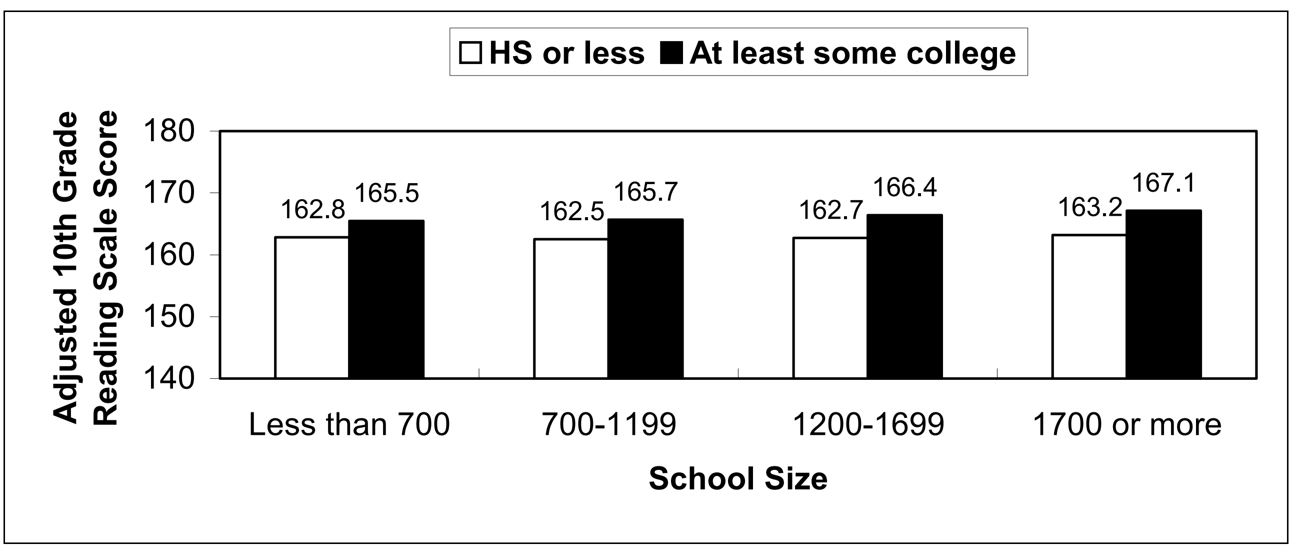

Figure 6. High school cohort reading model - interaction between parent education level and school size.

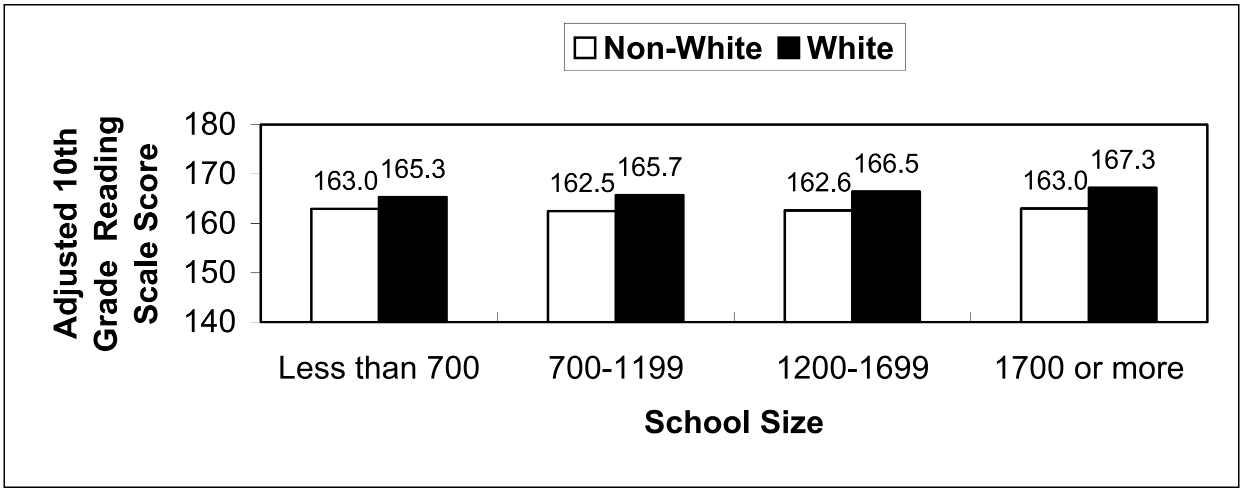

Figure 7. High school cohort reading model - graphic representation of interaction between ethnicity and school size. (*Note.* School size groups are divided based on quartile cutoffs for purpose of illustration. Grade 8 achievement groups are based on a dichotomization of the scale score variable used in the HLM model that corresponds to the official cut point used by the state to determine whether a student is performing at or above grade level.)

Table 9
High school cohort 2-level HLM regression model - mathematics

Level	Variable	b	SE	t	F
Student-level	Prior Mathematics Achievement – 8th Grade	12.55	.56	-22.31*	
	Ethnicity[a]	4.99	.41	12.27*	
	Parent Education Level[b]	4.09	.37	11.06*	
	Completed Algebra I Course	1.27	.11	11.50*	
	Completed Geometry Course	8.59	.19	45.66*	
School-level	School Size[c]				5.90*
	Less than 700 students	-5.14	.63	-8.16*	
	700-1,199 students	-3.93	.57	-6.90*	
	1,200-1,699 students	-2.64	.62	-4.28*	
	Free/Reduced Price Lunch (% eligible)	-.03	.009	-3.89*	
Interactions	Size x Prior Achievement				10.66*
	Size x Ethnicity				2.72*
	Size x Parent Education Level				3.75*
Summary	% between-student variation explained				39.6%
	% between-school variation explained				39.6%

[a] 0 = non-White, 1 = White. [b] 0 = high school diploma or less, 1 = at least some post-secondary education. [c] Reference group for school size is 1,700+ students. $*p < .05$.

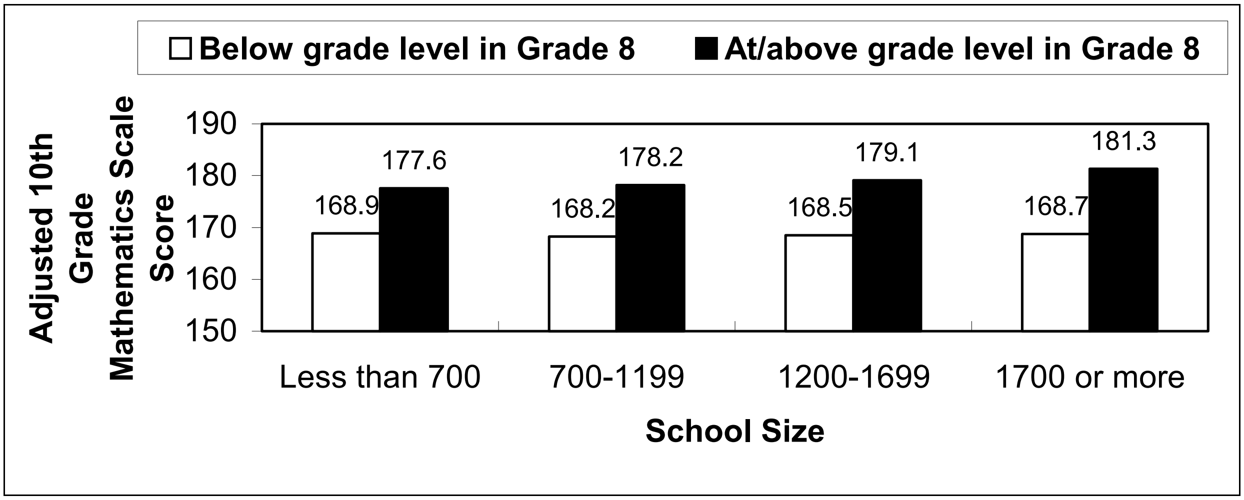

Figure 8. High school cohort mathematics model - interaction between prior achievement and school size.

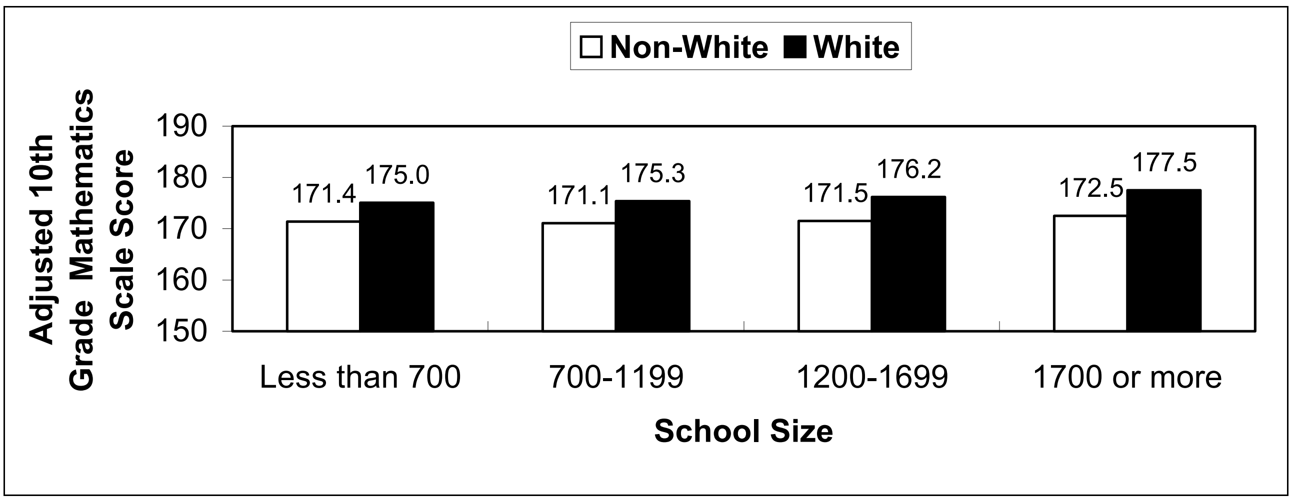

Figure 9. High school cohort mathematics model - interaction between ethnicity and school size.

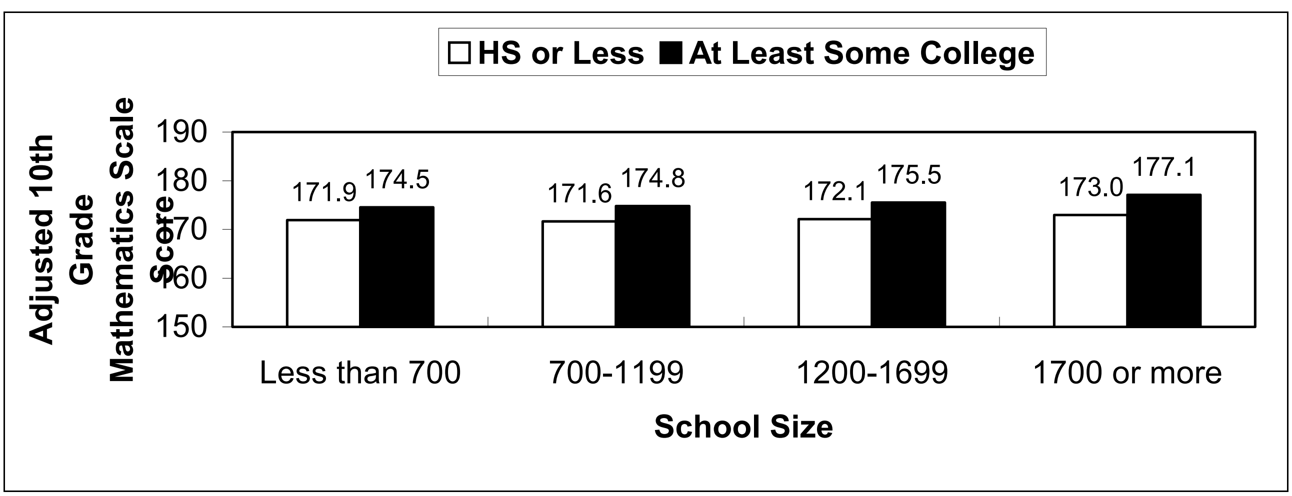

Figure 10. High school cohort mathematics model - interaction between parent education level and school size.

Discussion

According to prior research on school size and its relationship to student achievement and behavior, the majority of studies indicate that smaller is better. There are some inconsistencies with respect to high school size and achievement, but studies of school size in general have demonstrated that smaller schools are associated with better behavioral outcomes, higher rates of participation in extracurricular activities, and higher achievement. In addition, many of these studies have been conducted with large, nationally representative samples of students and schools, which would imply that those results should be fairly robust and applicable to a wide range of educational situations.

Analyses of North Carolina data, however, show a more complex pattern of results. At the elementary and middle school levels, school size was related to achievement but only through interactions with students' prior level of achievement. Students who were scoring on grade level in reading and mathematics in the baseline year tended to score higher two years later if they attended larger schools, whereas students who were scoring below grade level in the baseline year demonstrated slightly lower performance two years later if they attended larger schools. These effects were somewhat stronger in middle school than in elementary school. At the high school level, size was positively related to both reading and mathematics achievement in the overall sample. The benefits of size at the high school level, however, appeared to accrue disproportionately (or in some cases entirely) to higher-achieving students, White students, and students whose parents had more education, especially in mathematics. Effects seen in the high school cohort were the largest in the study. Although the nature of the interactions involving school size in the current study differed by grade level and in some cases were small in magnitude, in each case the interaction implied that learning was less equitable in larger schools. The results of this study provide interesting parallels to previous studies suggesting that student and community characteristics interact with size (e.g., Howley & Bickel, 1999; Lee & Smith, 1997). While the Lee and Smith study attributed greater achievement disparities in larger schools to the relatively low performance of less-advantaged students in those environments, the current study raises the possibility that these disparities may in some cases be due to the relatively high performance of more-advantaged students in larger schools. Correspondingly, while the line of research by Howley and colleagues posits that the poverty-achievement relationship is larger among larger schools, the current study suggests that oft-documented achievement gaps between student subgroups may also larger within larger schools. This suggests that the school-level "equity effects" of size identified by Howley and colleagues may also translate down to student subgroups *within* schools.

Thus, although the same basic achievement gaps are identified across different studies of school size, the possible underlying explanations of these results and their implications could be very different. For example, the observed results may be a function of higher-achieving students in larger schools taking disproportionate advantage of broader and deeper curriculum offerings. The stratification and tracking arrangements that this explanation would suggest may be more easily fostered in larger schools (Haller, Monk, Spotted Bear, Griffith, & Moss, 1990; Monk, 1987). The likelihood of this explanation is further bolstered by the fact that the largest effect in the current study was seen in high school mathematics, where stratification and tracking are particularly prevalent (Gamoran & Hannigan, 2000; Haller et al., 1990; Oakes, Gamoran & Page, 1992). If so, interventions that attempt to raise the level of rigor and breadth of curricula in smaller schools may be warranted (Barker, 1985), or perhaps interventions targeted at promoting greater access to accelerated curricula for historically under-represented groups. Technology applications that allow higher-level offerings such as Advanced Placement courses to be taken via the internet in smaller, more remote schools (or for that matter by larger numbers of students in any school) might be beneficial in this respect, as would programs targeted at better identifying and serving gifted and talented students from more diverse backgrounds (Darity, Castellino, Tyson, Cobb, & McMillen, 2001).

It is also possible that academically-challenged students perform better in smaller schools because of factors related to the school culture and environment. If so, large schools might take advantage of organizational structures such as those discussed by Cawelti (1993) and Goodlad (1984) in order to create a small-school atmosphere within a large school. These may include vertical house plans (i.e. schools-within-schools) which essentially divide a large school into multiple smaller schools on the same campus, each of which operates with its own group of students and with relative autonomy, or special focused curriculum tracks within high schools that could serve as within-school magnet programs to circumvent the enormity of a large school. These approaches assume factors such as the social climate, the personal relationships between students and teachers, and the extent to which students can become engaged and invested in the schooling experience are the true catalysts of positive outcomes in small schools. A recent study by Darling-Hammond, Ancess, & Ort (2002) in New York documents a case where this kind of reorganization strategy was applied to a large urban high school and resulted in improved student outcomes. A five-year evaluation of the Bill and Melinda Gates Foundation program which funds the creation of smaller high schools is also currently underway (American Institutes for Research, 2003) and should help to inform these issues as well. It would seem, then, that the implications of the findings from quantitative studies of the size-achievement relationship may depend on which interpretations of the identified interactions are found to be most plausible. Given the current movement toward closing achievement gaps and getting student subgroups to meet criterion-based academic standards, which has been reinforced by the recent passage of the federal No Child Left Behind Act, proper delineation of the mechanisms underlying these relationships is critical for designing effective interventions for students who are at risk of not meeting those standards.

Further studies of how school size is related to the day-to-day activities of students and teachers may provide greater insight into this issue. In any organization, structural factors such as size tend to have their effects on outcomes indirectly by altering the day-to-day processes and interactions that occur within the organization. Therefore, studies looking for a direct link between school size and student outcomes that fail to include these process factors in the analysis may reach different conclusions about the true role of school size in students' growth and development. Some studies have suggested that factors such as more personal social relations (e.g., teacher-to-student, student-to-student, etc.), stronger internal accountability, and opportunities for more varied approaches to instruction and asssessment may play a mediating role (Darling-Hammond et al., 2002; Lee, Smerdon, Alfeld-Liro, & Brown, 2000; Wasley et al., 2000). Further examination of how these relationships may play out differently for different subject areas or across grade levels would also be important, as the current study as well as some prior investigations both imply that school size effects may be more common and relatively larger in mathematics and at higher grade levels (Howley, 1995; Lee and Smith, 1997; Johnson, Howley, & Howley, 2002). Whether these differential relationships are a function of cumulative developmental effects that are most easily seen in later grades or perhaps qualitative differences in the size-achievement dynamic across levels and subjects is largely unknown at this point.

The findings reported here, along with those of prior school size-achievement studies should also lead local boards of education and other policymakers to at least consider whether efforts to consolidate smaller schools into larger ones might be achieving desired efficiencies at some cost to at-risk student groups. When considering only the financial ramifications, larger schools tend to be less expensive to operate, on a per-pupil basis, "other things being equal" (McGuire, 1993, p. 171). Unfortunately, these "other things" are rarely equal, and financial savings from consolidation will probably not apply equally across all expenditure areas. For example, the consolidation of two schools may save personnel expenses by eliminating a principal's position, but it may simultaneously result in an increase in pupil transportation costs. The consolidation of smaller schools into larger units also may or may not result in cost savings depending on how one defines the outputs of schooling (Lawrence et al., 2002). For example, Stiefel, Berne, Iatarola, & Fruchter (2000) have shown that smaller high schools, although they may not enroll as many students per dollar as larger schools, may be producing more graduates per dollar. Therefore, smaller schools may also be more economically efficient if the output is defined as graduates instead of enrollees while also possibly providing more supportive environments for at-risk students.

In attempting to interpret the results of the current and previous studies of the size-achievement relationship, it has to be acknowledged that size is inextricably intertwined with many other factors that are often associated with academic and behavioral outcomes for students. These complexities are further underscored by the finding in this study that the size-prior achievement interactions were larger and more prevalent than the size-parent education level or size-ethnicity interactions, despite the fact that prior achievement is typically highly correlated with these variables. This overlap, coupled with the fact that school size is typically not manipulated experimentally for research purposes, makes it very difficult to identify which specific factors or combinations thereof might be the most salient. Although including some of these confounding variables in the analysis for control purposes is helpful, it is not a substitute for random assignment. The 2,200-plus public schools in North Carolina show great variability on a number of these potential factors (e.g., urban/rural location, family/community characteristics, student demographics, poverty, etc.), with a good number of schools identified at each point on those spectra. Compared to other states, North Carolina consistently falls at or near the national median on the vast majority of these types of school characteristics (U. S. Department of Education, 2003). The extent to which the results obtained here could be applied to other geographic settings, however, may be influenced by the extent to which those other settings mimic that profile. Analyses of similar data from other states or locales that are more homogeneous on some of these factors may in fact yield different results.

Howley and Howley (2004) would imply that the generalizability of the results of the current study may also be limited by the fact that there are only a handful of truly small high schools (i.e., less than 100 students) in North Carolina compared to states that have been examined in prior research (e.g., Howley & Bickel, 1999), a phenomenon they refer to as size bias. The analyses here do not speak to this issue; however, Lee (2004) argues that it is unclear as to whether or how a small number of small schools would result in actual bias of parameter estimates. With respect to the practical significance of the findings, the elementary and middle school effects, while statistically significant, are small and therefore should be interpreted cautiously. It is unclear whether the effects at these levels are smaller than at the high school level because size is actually less important at those grades. This discrepancy in size of effects may likely be due to statistical factors such as less actual variance in school size at the elementary and middle levels compared to the high school level.

Given the findings here, along with previous studies indicating that the achievement gap exists prior to children entering the K-12 system (e.g., Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998; U. S Department of Education, 2000b), it is unlikely that school size is the primary force behind the well-documented achievement disparities between various student subgroups. Its relationship to achievement and achievement gaps may, however, be mediated by other process-related factors that are either encouraged or stifled by school size. Delineating the specific mechanisms through which school size affects student outcomes and the ways in which those effects might be selectively experienced by different student subgroups or in different subject areas is a potentially rich area of investigation that may help us to better understand how schools need to be structured so that all students can reach high standards.

References

American Institutes for Research. (2003). *High time for high school reform: Early findings from the National School District and Network grants program.* Retrieved on October 21, 2003 from http://www.gatesfoundation.org/nr/downloads/ed/smallschools/Small_schools_eval_2003.pdf.

Baird, L. L. (1969). Big school, small school: A critical examination of the hypothesis. *Journal of Educational Psychology, 60*, 253-260.

Barker, B. (1985). Curricular offerings in small and large high schools: How broad is the disparity? *Research in Rural Education, 3,* 9-12.

Barker, R. G., & Gump, P. V. (1964). *Big school, small school: High school size and student behavior.* Stanford, CA: Stanford University Press.

Bickel, R., & Howley, C. (2000, May 10). The influence of scale on student performance: A multi-level extension of the Matthew Principle. *Education Policy Analysis Archives, 8*(22). Retrieved September 12, 2003 from http://epaa.asu.edu/epaa/v8n22/.

Caldas, S. J. (1993). Reexamination of input and process factor effects on public school achievement. *Journal of Educational Research, 86*, 206-214.

Cawelti, G. (1993). Restructuring large high schools to personalize learning for all. *ERS Spectrum, 11*(3), 17-21.

Darity, W., Castellino, D., Tyson, K., Cobb, C., & McMillen, B. (2001). *Increasing opportunity to learn via rigorous courses and programs: One strategy for closing the achievement gap for at-risk and ethnic minority students.* Raleigh, NC: North Carolina Department of Public Instruction.

Darling-Hammond, L., Ancess, J., & Ort, S. W. (2002). Reinventing high school: Outcomes of the coalition campus school project. *American Educational Research Journal, 39*, 639-673.

Fetler, M. (1989). School dropout rates, academic performance, size, and poverty: Correlates of educational reform. *Educational Evaluation and Policy Analysis, 11*, 109-116.

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

Fowler, W. J. (1995). School size and student outcomes. In B. Levin, W. Fowler, & H. J. Walberg (Eds.), *Advances in educational productivity, Vol. 5* (pp. 3-26). Greenwich, CT: JAI Press.

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

Friedkin, N., & Necochea, J. (1998). School system size and performance: A contingency perspective. *Educational Evaluation and Policy Analysis, 10*, 237-249.

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis, 22*, 241-254.

Goodlad, J. I. (1984). *A place called school.* New York: McGraw-Hill.

Grabe, M. (1981). School size and the importance of school activities. *Adolescence, 16*, 21-31.

Haller, E. J. (1992). High school size and student indiscipline: Another aspect of the school consolidation issue? *Educational Evaluation and Policy Analysis, 14*, 145-156.

Haller, E. J., Monk, D. H., Spotted Bear, A., Griffith, J., & Moss, P. (1990). School size and program comprehensiveness: Evidence from High School and Beyond. *Educational Evaluation and Policy Analysis, 12,* 109-120.

Haller, E. J., Monk, D. H., & Tien, L. T. (1993). Small schools and higher-order thinking skills. *Journal of Research in Rural Education, 9*, 66-73.

Heaviside, S., Rowand, C., Williams, C. & Farris, E. (1998). *Violence and discipline problems in U. S. public schools: 1996-97.* U. S. Department of Education, National Center for Education Statistics, NCES publication #98-030.

Holland, A., & Andre, T. (1987). Participation in extracurricular activities in secondary school: What is known, what needs to be known? *Review of Educational Research, 57*, 437-466.

Howley, C. (1995, November 15). The Matthew Principle: A West Virginia replication? *Education Policy Analysis Archives, 3*(18). Retrieved September 12, 2003 from http://epaa.asu.edu/epaa/v3n18.html.

Howley, C., & Bickel, R.. (1999). *The Matthew Project: National Report.* ERIC Document Reproduction Service (ED 433174).

Howley, C. B. & Howley, A. A. (2004, September 24). School size and the influence of socioeconomic status on student achievement: Confronting the threat of size bias in national data sets. *Education Policy Analysis Archives, 12*(52). Retrieved October 15, 2004 from http://epaa.asu.edu/epaa/v12n52/.

Jewell, R. W. (1989). School and school district size relationships: Costs, results, minorities, and private school enrollments. *Education and Urban Society, 21*, 140-153.

Johnson, J. D., Howley, C. B., & Howley, A. A. (2002). *Size, excellence, and equity: A report on Arkansas schools and districts.* Retreived October 10, 2003 from http://oak.cats.ohiou.edu/~howleyc/ARfin.htm.

Kaufman, P., Kwon, J. Y., Klein, S., & Chapman, C. D. (1999). *Dropout rates in the United States: 1988*. Washington, DC: National Center for Education Statistics, NCES publication #2000022.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Lawrence, B. K., Bingler, S., Diamond, B. M., Hill, B., Hoffman, J. L., Howley, C. B., Mitchell, S., Rudolph, D., & Washor, E. (2002). *Dollars and sense: The cost effectiveness of small schools.* Cincinnati, OH: KnowledgeWorks Foundation.

Lee, V. E. (2004, September 24). Effects of high-school size on student outcomes: Response to Howley and Howley. *Education Policy Analysis Archives, 12*(53). Retrieved October 15, 2004 from http://epaa.asu.edu/epaa/v12n53/.

Lee, V. E., Smerdon, B. A., Alfeld-Liro, C., & Brown, S. L. (2000). Inside large and small high schools: Curriculum and social relations. *Educational Evaluation and Policy Analysis, 22,* 147-171.

Lee, V. E., & Smith, J. B. (1997). High school size: Which works best and for whom? *Educational Evaluation and Policy Analysis, 19,* 205-227.

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

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

McGuire, K. (1989). School size: The continuing controversy. *Education and Urban Society, 21*, 164-174.

Monk, D. H. (1987). Secondary school size and curriculum comprehensiveness. *Economics of Education Review, 6*, 137-150.

North Carolina Department of Public Instruction (2000). *The North Carolina state testing results: 1998-1999*. Raleigh, NC: Author.

North Carolina Department of Public Instruction (2001). *The North Carolina state testing results: 1999-2000*. Raleigh, NC: Author.

Oakes, J., Gamoran, A., & Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570-608). New York: MacMillan.

Page, R. M. (1991). Adolescent use of alcohol, tobacco, and other psychoactive substances: Relation to high school size. *American Secondary Education, 19*(2), 16-20.

Pittman, R. B., & Haughwout, P. (1987). Influence of high school size on dropout rate. *Educational Evaluation and Policy Analysis, 9*, 337-343.

Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black-white test score gap. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 103-145). Washington, DC: Brookings.

Schoggen, P., & Schoggen, M. (1988). Student voluntary participation and high school size. *Journal of Educational Research, 81*, 288-293.

Snijders, T. A., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods and Research, 22*, 342-363.

Stiefel, L., Berne, R., Iatarola, P., & Fruchter, N. (2000). High school size: Effects on budgets and performance in New York City. *Educational Evaluation and Policy Analysis, 22*, 27-39.

U. S. Department of Education. (2000a). *Digest of education statistics: 2000.* Washington, DC: U. S. Government Printing Office.

U. S. Department of Education. (2000b). *America's kindergarteners (NCES #2000-070).* Washington, DC: Author.

U. S. Department of Education. (2003). *Overview of public elementary and secondary schools and districts: School year 2001-02 (NCES # 2003-411).* Washington, DC: Author.

Walberg, H. J., & Walberg III, H. J. (1994). Losing local control. *Educational Researcher, 23*(5), 19-26.

Wasley, P. A., Fine, M., Gladden, M., Holland, N. E., King, S. P., Mosak, E., & Powell, L. C. (2000). *Small schools, great strides.* New York: Bank Street College of Education. Retrieved on February 12, 2001 from http://www.bankstreet.edu/gems/publications/smallschoollow.pdf.

About the Author

Bradley J. McMillen
Senior Evaluation Consultant
North Carolina Department of Public Instruction
Division of Accountabilty Services
6314 Mail Service Center
Raleigh, NC 27699-6314
Phone: (919) 807-3808
Fax: (919) 807-3772
Email: bmcmille@dpi.state.nc.us

Brad McMillen conducts policy and program evaluation studies for the North Carolina Department of Public Instruction. His research interests include school organizational characteristics and students at risk for academic failure.

Education Policy Analysis Archives http://**epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@uconn.edu.