EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Sherman Dorn
College of Education
University of South Florida

Volume 14 Number 28 November 3, 2006 ISSN 1068–2341

School Size, Student Achievement, and the "Power Rating" of Poverty: Substantive Finding or Statistical Artifact?

Theodore Coladarci
University of Maine

Citation: Coladarci, T. (2006). School size, student achievement, and the "power rating" of poverty: Substantive finding or statistical artifact?. *Education Policy Analysis Archives, 14*(28). Retrieved [date] from http://epaa.asu.edu/epaa/v14n28/.

Abstract
The proportion of variance in student achievement that is explained by student SES—"poverty's power rating," as some call it—tends to be lower among smaller schools than among larger schools. Smaller schools, many claim, are able to somehow disrupt the seemingly axiomatic association between SES and student achievement. Using eighth-grade data for 216 public schools in Maine, I explored the hypothesis that this in part is a statistical artifact of the greater volatility (lower reliability) of school-aggregated student achievement in smaller schools. This hypothesis received no support when reading achievement served as the dependent variable. In contrast, the hypothesis was supported when the dependent variable was mathematics achievement. For reasons considered in the discussion, however, I ultimately concluded that the latter results are insufficient to affirm the statistical-artifact hypothesis here as well. Implications for subsequent research are discussed.

Keywords: small schools; student achievement; socioeconomic status;, rural education.

Tamaño de la escuela, logro académico de los estudiantes, y la "potencia nominal" de la pobreza: ¿Hallazgos sustantivos o artificio estadístico?

Resumen

La proporción de varianza en el logro académico de los estudiantes que se atribuye al estatus socio económico (SES)—llamado por algunos "la potencia nominal de la pobreza"—tiende a ser menor entre escuelas pequeñas que entre escuelas grandes. Las escuelas más pequeñas, según dicen muchos, logran de alguna manera, romper la aparente asociación axiomática entre el SES y el logro académico de los estudiantes. Usando datos de octavo grado de 216 escuelas públicas del estado de Maine, en este artículo exploro la hipótesis que esto es, en parte, un artificio estadístico que se da debido a la mayor volatilidad (menor confianza) del logro académico de los estudiantes cuando se toma en cuenta el valor agregado de toda la escuela en las escuelas pequeñas. Esta hipótesis no se sustenta cuando el logro académico en lectura es la variable dependiente. Por el contrario, la hipótesis sí se sustenta cuando la variable dependiente es el logro académico en matemáticas. Sin embargo, por las razones argüidas en la discusión, se concluye que los resultados en matemáticas tampoco son suficientes para confirmar la hipótesis de artefacto estadístico. Al final se incluyen sugerencias para continuar la investigación sobre este tema.

Introduction

As every student of education research knows, the relationship between student achievement and socioeconomic status (SES) is well-established in the empirical literature: All things equal, as student SES increases, so does student achievement (e.g., Sirin, 2005; White, 1982). Further, this holds regardless of the unit of analysis employed (e.g., student, school, multilevel). The seemingly axiomatic nature of this relationship notwithstanding, a recurring finding in rural education research is that SES and school size interact in affecting student achievement (e.g., Howley, 1996; Howley & Bickel, 1999; Huang & Howley, 1993; Johnson, Howley, & Howley, 2002; McMillen, 2004; also see Friedkin & Necochea, 1988; Lee & Smith, 1997). In other words, the magnitude of the relationship between SES and achievement depends on the size of the school, or, equivalently, that the magnitude of the relationship between school size and achievement depends on the SES makeup of the school.

How is such an interaction demonstrated? With the school as the unit of statistical analysis, for example, interaction is shown by regressing achievement on SES, school size, and the mathematical product of SES and school size, and then testing the product term for statistical significance. If the slope associated with this term is statistically significant—which researchers have been reporting with remarkable consistency—there is an interaction between SES and school size. A common way to illustrate such an interaction is to show that the school-level correlation between SES and achievement is weaker among smaller schools than among larger schools. That is, SES explains less of the variance in school achievement among smaller schools than it does among larger schools. As Huang and Howley (1993) put it, smaller schools "mitigate" the effect that SES has on student achievement.

The mitigating-effect finding enjoys considerable fanfare by researchers, advocacy groups, and practitioners alike. Johnson, Howley, and Howley (2002), for example, declared this finding to be "among the most consistent ever to be reported in educational research" (pp. 36–37). The Rural School and Community Trust, a strong advocate for rural schools and communities, crafted the phrase "poverty's power rating" to refer to the percentage of variance in achievement that is explained by SES (i.e., the familiar coefficient of determination). In newsletters and press releases, the Rural Trust celebrates the recurring finding that the power rating of poverty is markedly lower—sometimes negligible—among smaller schools than among larger schools. "In study after study," the organization's president recently announced, "small schools have been shown to cut poverty's power over student achievement" (Tompkins, 2006). And in an op-ed published in my local newspaper, a school superintendent and his colleagues summed it up this way: "Small schools are an antidote to the impact of poverty on school achievement" (Butler et al., 2005, p. A9).

Despite my affinity to rural schools and communities, I have always been uneasy with the mitigating-effect finding and, in particular, the markedly lower "power rating" of poverty in smaller schools. As much as I am attracted to the notion that smaller schools, by virtue of their smallness, are somehow able to disrupt the achievement disadvantage of lower-SES students, and as much as I can imagine the many ways in which smaller schools might be able to pull this off (although hard data would be helpful), my immediate suspicion was that the diluted SES-achievement correlation among smaller schools may have little to do with the educational experience characterizing such schools. Rather, I suspected a statistical artifact at play.

Loosely defined, a statistical artifact is where a research result is misleading because of an artificial or extraneous effect due to statistical considerations. For example, if X has modest variance and, further, $r = 0$ between X and Y, the absence of relationship between X and Y very well could be due to restricted range in X (a statistical artifact) rather than to an absence of relationship between the two constructs underlying X and Y. In the present context, the putatively ameliorative role of smaller schools in the SES-achievement relationship would be a statistical artifact if, say, there were much less variability in either student SES or student achievement among smaller schools than among larger schools. This in fact was my immediate suspicion, both because it is so obvious as a plausible rival hypothesis (when subgroup correlations are comparatively small) and because I saw no acknowledgment of this possibility by those who were doing (or celebrating) the research. However, I was unable to find evidence of restricted variance in the statistics reported by the researchers, nor did such evidence surface in my quick reanalysis of Maine data that had been featured in a newsletter of the Rural Trust ("Maine's small schools cut poverty's power," 2005).

My interest in the challenges that small schools face related to the "adequate yearly progress" requirement of No Child Left Behind suggested another possible statistical artifact: the greater volatility of school-level student achievement among smaller schools (Coladarci, 2003). School achievement can differ widely from one year to the next for smaller schools, whereas larger schools enjoy considerably greater stability in this regard (e.g., Hill & DePascale, 2003; Kane & Staiger, 2002; Linn & Haug, 2002).

Consider Figure 1, for example, which shows the relationship between the size of the fourth-grade cohort tested in a Maine school and the one-year change in the proportion of students in that school who are proficient on the Maine Educational Assessment reading test. Although the average change from one year to the next hovers around zero for all schools (dashed line), there is considerably greater variability among smaller schools in the amount of this change. For schools having 15 or fewer fourth graders, for instance, this change ranges from −.47 (declining from 60% proficient to 13% proficient) to +.83 (increasing from 17% proficient to 100% proficient). In contrast, the corresponding figures are only −.07 and +.09, respectively, among schools having 150 or more fourth graders.

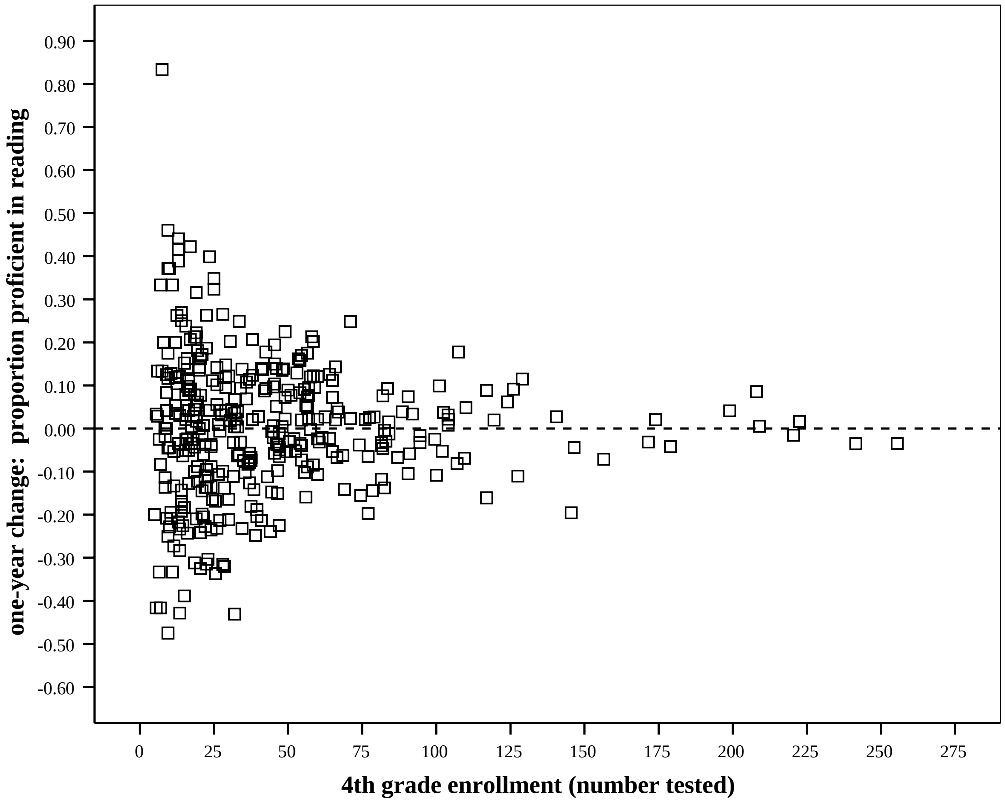


Figure 1. The relationship between the number of fourth-grade students tested in a school and the one-year change in the proportion of students who are proficient. (Source: Coladarci, 2003, Figure 4)

At issue here is the differential reliability of school-aggregated student achievement for smaller versus larger schools. A school's achievement result at any given point in time can be thought of as an estimate of the school's "true score" (Hill, 2002, p. 2). Insofar as school achievement tends to be less stable from one year to next for smaller schools than for larger schools (much as shorter tests tend to have lower test-retest reliability than longer tests), there thus is a greater likelihood that the reported level of achievement for a smaller school will be at variance with the school's true level of achievement—the school's effectiveness as an institution (e.g., see Cronbach, Linn, Brennan, & Haertel, 1997, p. 393). Because reliability places an upper limit on a variable's ability to correlate with other variables (e.g., Thorndike, 1982, p. 222), a plausible conjecture is that the lower SES-achievement correlation among smaller schools is an artifact of the lower reliability of school achievement for such schools. In short, this is the conjecture I investigated in the present study.

In pursuing the statistical-artifact hypothesis, I was not motivated by a desire to debunk popular opinion regarding the virtues of small schools. Rather, I simply wished to determine whether a celebrated proposition in the rural education literature could withstand a sincere attempt to falsify it. If such an attempt were to fail, then we all are entitled to a greater confidence in this proposition—greater *warranted* confidence—than we presently can claim (e.g., see Phillips, 2000, pp. 137–156).

Method

Context, Data Source, and Variables

Maine, the context of this study, is a predominantly rural state in northern New England. With roughly 1.3 million people spread over 33,215 square miles (a geographical area comparable to the remaining New England states combined), the vast majority (96%) of Maine residents are White, the median household income is $39,212 (versus the U.S. average of $43,318), and 10.7% of Mainers live below the poverty level (versus 12.5% in the U.S.). The number of elementary and secondary students is about one fifth that of the U.S. average (198,820 vs. 956,762), with approximately one third (32.34%) of these students qualifying for free or reduced lunch (compared to 37.40% across the nation).[1]

My focus is on eighth-grade achievement in public schools, using reading and mathematics data from the Maine Educational Assessment (MEA) for the 2002–2003 and 2003–2004 school years. (The MEA scale range was 501–580 at that time.) For each public school having an eighth grade, I created a weighted two-year mean for both reading achievement (*reading*) and mathematics achievement (*math*). Similarly, I determined for each school the weighted two-year percentage of students receiving subsidized lunch (*poverty*).

As for operationally defining school size, I immediately faced the distinction between a school's total enrollment across all grades and a school's mean enrollment per grade. Howley (2002, pp. 52–53) argues that the latter is the appropriate measure of school size because per-grade enrollment takes into account a school's grade configuration—that, say, a K–8 school with 270 students (30 per grade) is arguably smaller than a 6–8 school with 270 students (90 per grade). I have yet been able to appreciate the logic of this position, which inevitably must fall on how one conceptualizes "school" and its effects on students. But because most mitigating-effect studies employed the enrollment-per-grade measure of school size, I followed suit in the analyses reported below. Specifically, I determined the mean enrollment per grade for each school, averaged across 2002–2003 and 2003–2004 (*school size*). (I confess that I also ran all analyses using a total-enrollment measure of school size, which yielded results similar to those based on enrollment per grade.)

To estimate a school's volatility in eighth-grade achievement, I determined the difference in mean achievement from 2003–2004 to 2002–2003 for reading and mathematics separately. I then recoded the absolute value of these differences to obtain a volatility rating for each school. There were separate volatility ratings for reading and math (*volatility*), and both were constructed as shown in Table 1.

---

[1] These statistics are available from the Web sites of the State of Maine, http://www.state.me.us; the U.S. Census Bureau, http://www.census.gov; and the National Center for Education Statistics, http://nces.ed.gov.

Table 1
Volatility definitions

Volatility rating	Change in school achievement[a] mean
1	0 to 2.50 points
2	2.51 to 5.00 points
3	5.01 to 7.50 points
4	7.51 to 10.00 points
5	10.01 to 12.50 points
6	12.51 to 15.00 points
7	15.01 to 17.50 points
8	17.51 to 20.00 points

[a]The scale of the Maine Educational Assessment ranges from 501 to 580.

Analyses

I restricted my analyses to public schools in Maine that had (a) an eighth grade in 2002–2003 and 2003–2004, (b) data on all variables for both 2002–2003 and 2003–2004, and (c) neither changed their grade span from one year to the next nor absorbed in 2003–2004 students from a school that had closed at the end of 2002–2003. Finally, I eliminated schools that did not have at least two eighth-grade students in each of the two school years. These restrictions resulted in a final sample of 216 schools from a universe of 233 public schools having an eighth grade in 2003–2004.

The school served as the unit of analysis. After conducting preliminary analyses to establish the trustworthiness of the data, which had been downloaded from state websites, I began by demonstrating the aforementioned interaction between socioeconomic status and school size. I did so using ordinary least-squares regression, where, in the present case, the equation is $\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_1X_2$ (e.g., Aiken & West, 1991). Here, $\hat{Y}$ is the predicted value of the dependent variable (either reading or math); a is the intercept; X_1 and X_2 are poverty and school size, respectively; and X_1X_2 is their mathematical product. Prior to creating the product term and consistent with common practice, I centered poverty and school size at their respective means to reduce the inevitable collinearity engendered by multiplicative terms. All analyses were conducted using *SPSS 14.0 for Windows*.

The statistical significance of b_3, the slope of the product term, indicates the presence of interaction between X_1 and X_2—that the magnitude of b_1 varies with X_2, or, symmetrically, that the magnitude of b_2 varies with X_1. In the present context, this means that the degree of association between poverty and achievement (b_1) depends on school size (X_2), or, equivalently, that the degree of association between school size and achievement (b_2) depends on the socioeconomic status of the school (X_1). By entering the product term on a separate step, I obtained the increment in explained variance (ΔR^2) that is associated with the poverty-size interaction, the statistical significance of which is identical to that of b_3.

To further illustrate the degree of interaction between poverty and school size, and, in particular, to recast this interaction in terms of poverty's power rating, I fit separate achievement-on-poverty regression lines for schools falling above and below the median per-grade enrollment. The magnitude of interaction is shown by the degree to which the two within-group regression lines are nonparallel. From this analysis, I also obtained the within-group correlations between each achievement measure and poverty, which, when squared, represents the power rating of poverty.

To explore my statistical-artifact hypothesis—that poverty's reduced power rating, when examined among smaller schools, reflects the lower reliability of school-level achievement in such schools—I repeated these analyses on successively less-volatile (scorewise) collections of schools. The first set of analyses included all 216 schools (i.e., volatility = 1 through 8); the second set included schools for which volatility = 1 through 7; and so on to the final set of analyses involving the 104 least volatile schools (i.e., volatility = 1). (Again, there were separate volatility ratings for math and reading.) If, in fact, the poverty-size interaction is a statistical artifact due to the lower reliability of school-level achievement among smaller schools, then this interaction should attenuate with successively less-volatile collections of schools—and be negligible for schools having the least volatility.

Results

I begin by portraying the achievement volatility among these schools and, in turn, the relationship between this volatility and school size. To investigate the statistical-artifact hypothesis, I then report the results of the regression analyses on successively less-volatile collections of schools.

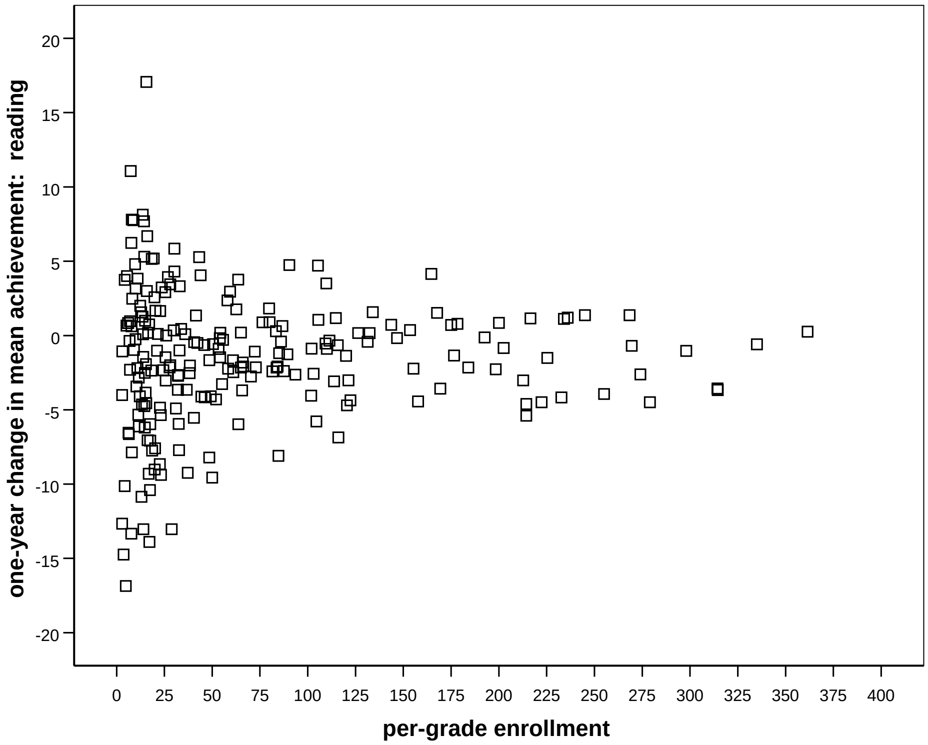
The Volatility of School-Level Achievement

As described above, I estimated a school's volatility in eighth-grade achievement by first calculating the difference in mean achievement from 2003–2004 to 2002–2003 for reading and for mathematics. Among these 216 schools, the change in achievement from one year to the next ranges from roughly –17 to +17 MEA points in reading (M = –1.56, SD = 4.61) and, for math, –19 to +16 MEA points (M = +1.14, SD = 4.79).

The well-established relationship between school size and achievement volatility is clearly evident in the present data (Figure 2). Again, there simply is greater volatility—lower reliability—of school-level achievement among smaller schools than among larger schools. This also can be seen in the correlations between school size and the absolute value of a school's change in achievement from one year to the next: r = –.31 and r = –.29, respectively, for reading and math. In short, Figure 2 and these two correlations underscore the relevance of the statistical-artifact hypothesis that frames the present study.

Table 2
Frequency distribution of volatility ratings (n = 216).

Volatility rating	Reading	Math
1	104	104
2	62	60
3	22	29
4	16	11
5	4	4
6	6	4
7	2	3
8	0	1



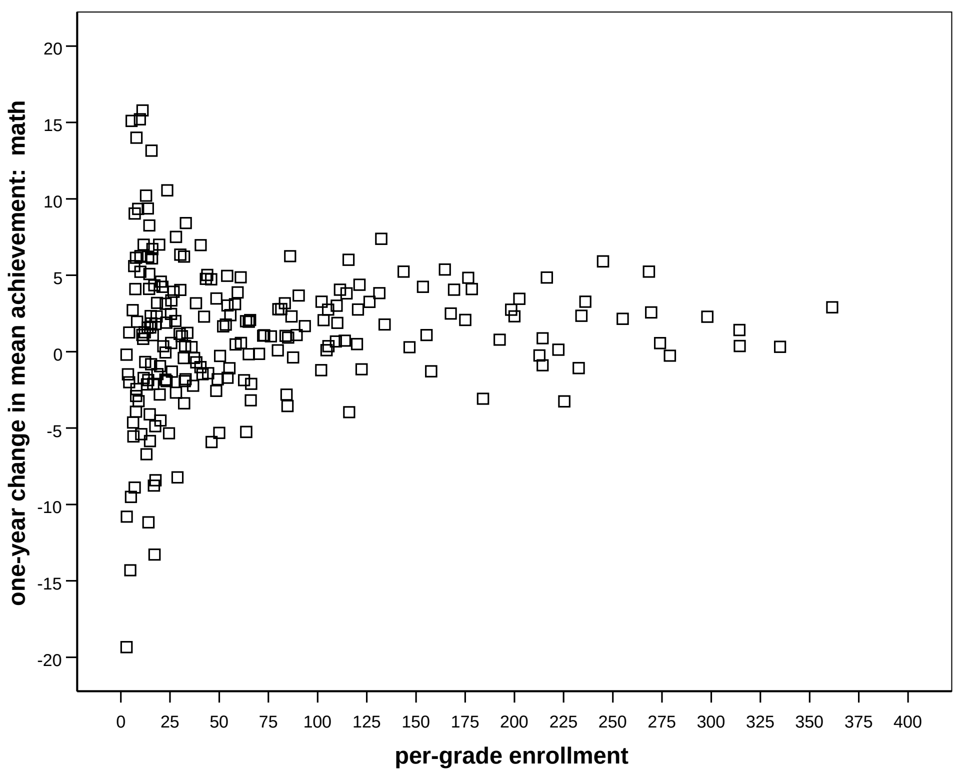


Figure 2. School size and the volatility of achievement in reading (top) and mathematics (bottom).

Regression Analyses: All Schools

The first set of regression analyses is based on all schools, irrespective of volatility. Table 3 presents descriptive statistics for reading, math, poverty, and school size. Not surprisingly, schools vary considerably with respect to both poverty and size: Some schools have as few as 3 students per grade and 3% of their students receiving subsidized meals, whereas other schools have as many as 358 students per grade and 84% of their students receiving subsidized meals. Reading and math correlate highly ($r = .74$), as one would expect, and each correlates with poverty in the customary fashion (Sirin, 2005; White, 1982). There is some tendency for smaller schools to be located in more impoverished communities ($r = –.34$). School size, however, is unrelated to achievement ($r = .07$, $p = .16$).

Table 3
Descriptive statistics: All schools (n = 216).

Variable	*M*	*SD*	Range	Intercorrelations: Reading	Intercorrelations: Math	Intercorrelations: Poverty
Reading	535.96	3.94	522.72, 547.69			
Math	528.16	4.36	514.51, 542.17	.74*		
Poverty	39.52	16.63	2.68, 83.86	-.48*	-.37*	
School size	72.78	77.31	2.94, 358.00	.07	.07	-.34*

Note. For the purpose of this table, poverty and school size are in their original uncentered form (which affects only the mean and range).
* $p < .01$.

Reading. Table 4 shows the regression results for reading. Poverty significantly and independently predicts reading at Step 1, whereas the corresponding effect of school size falls short of statistical significance. An additional 2.21% of the variance in reading is explained by the introduction of the product term at Step 2, which, consistent with prior research, shows a statistically significant interaction between poverty and school size ($p = .01$).

Table 4
Regressing reading on poverty, school size, and their product: All schools (n = 216)

Variable	*b*	s.e.	β	*t*	*p*	ΔR^2
Without interaction						
(constant)	535.96					
Poverty	-0.1216	0.0151	-0.51	-8.07	< .01	
School size	-0.0055	0.0032	-0.11	-1.71	.09	
With interaction						
(constant)	535.74					
Poverty	-0.1273	0.0151	-0.54	-8.45	< .01	
School size	-0.0080	0.0034	-0.16	-2.40	.02	
Poverty x size	-0.0005	0.0002	—	-2.52	.01	.0221

Note. Poverty and school size were centered for this analysis. At Step 2, the unadjusted and adjusted values of R^2 are .260 and .249, respectively.

Because the poverty-size interaction presently enjoys so much attention in the rural education literature, elaboration on the meaning of the various coefficients reported at Step 2 may be helpful. As we saw above, Step 2 estimates the effects for the full equation, $\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_1X_2$, where the last term, $b_3X_1X_2$, reflects the interaction of poverty and school size. As Aiken and West (1991) explain, b_1 is the reading-on-poverty slope for schools having a per-grade enrollment equal to the mean (i.e., centered $X_2 = 0$). For schools of average size, then, reading achievement decreases about .13 MEA points ($b_1 = –0.1273$) with every one-percentage-point increase in the students receiving subsidized meals. In standardized terms, this corresponds to a decline in reading achievement of roughly half a standard deviation ($\beta_1 = –0.54$) for each standard deviation increase in poverty (again, for schools of average size). One interprets b_2 analogously: For schools at the mean for poverty, reading achievement decreases about .01 MEA points (b_2 = -0.0080) for each one-student increase in school size—an achievement decline of 16% of a standard deviation ($\beta_2 = –0.16$) for each standard deviation increase in school size.

The statistical significance of b_3 signals the presence of interaction between poverty and school size. Specifically, the negative coefficient for the product term X_1X_2, coupled with the negative coefficient for poverty, means that the simple slope for poverty—i.e., the reading-on-poverty slope at a specified value of school size—is steeper (more negative) for larger schools than it is for smaller schools.

The concept of simple slope is central to interpreting a statistically significant interaction. The simple slope for poverty derives from the full equation, $\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_1X_2$, which, when recast as the Y-on-X_1 regression at a specified value of X_2, looks like this: $\hat{Y} = (a + b_2X_2) + (b_1 + b_3X_2)X_1$. The critical term here is $(b_1 + b_3X_2)$, which is the Y-on-X_1 slope for the specified value of X_2 (expressed as a deviation from the centered mean of zero). Select a deviation score to represent X_2, plug this value into the expression $(b_1 + b_3X_2)$, and you have the simple slope for poverty at a particular school size.

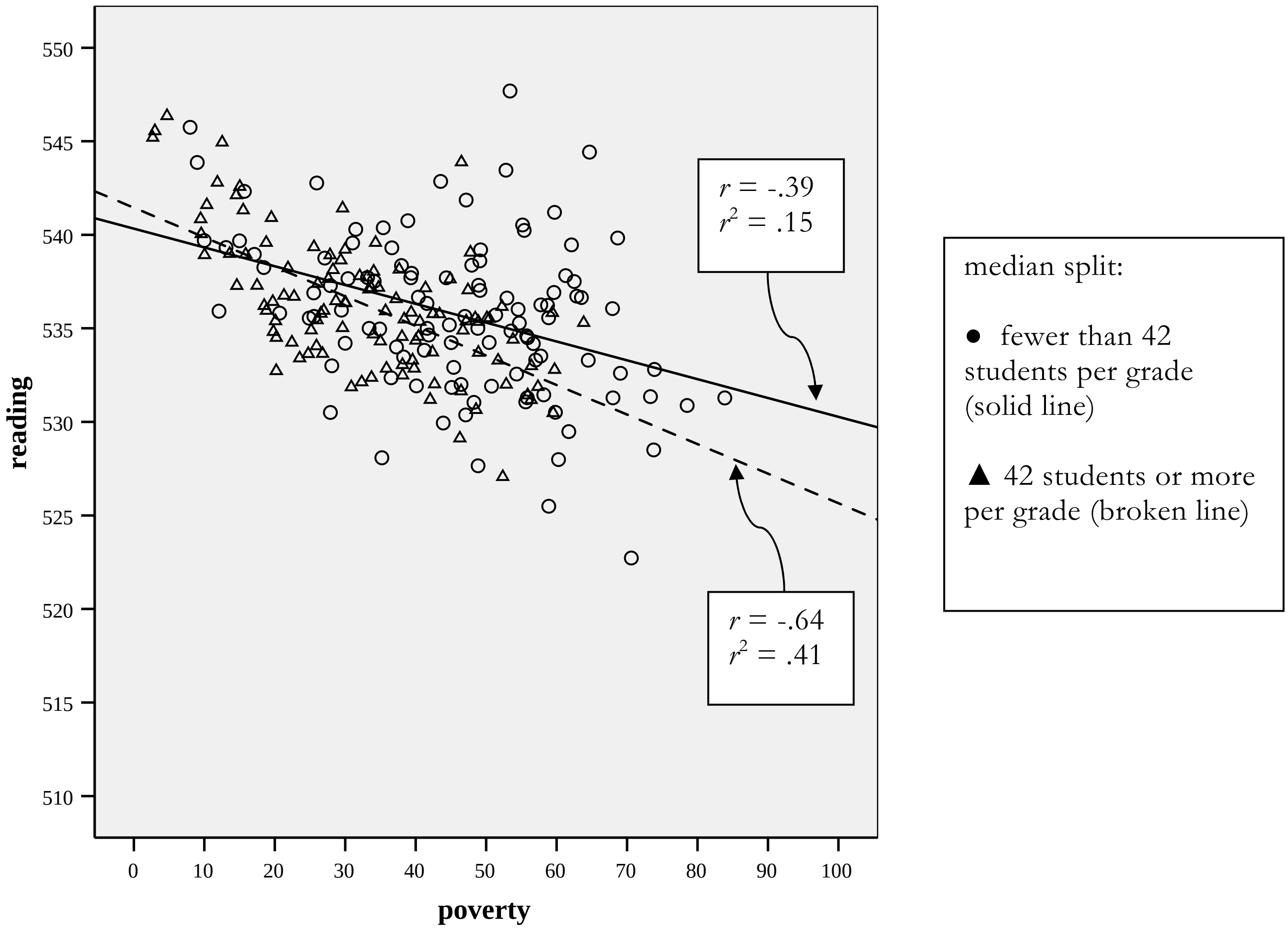


Figure 3. The interaction of poverty and school size (p =.013), reading: All schools (n = 216).

For example, consider a school having 16 students per grade—the 25th percentile in school size and roughly 57 fewer students than the mean ($\bar{X}_2$ = 72.78). The simple slope for schools of this size is $b_{-57} = -0.0982$, which corresponds to a standardized regression coefficient of $\beta_{-57} = -0.41$.[2] Thus, with each standard deviation increase in poverty, reading achievement in these smaller schools decreases approximately 40% of a standard deviation. The simple slope is slightly steeper for schools having 42 students per grade (the median school size, or 50th percentile): $b_{-31} = -0.1115$ or, in standardized terms, $\beta_{-31} = -0.47$. Now consider a school falling at the 75th percentile in school size, or 105 students per grade. Here, the unstandardized and standardized simple slopes are $b_{+32} = -0.1437$ and $\beta_{+32} = -0.61$, respectively. For these larger schools, then, reading decreases approximately 60% of a standard deviation with each standard deviation increase in poverty. Consistent with the statistically significant interaction of poverty and school size, simple slopes estimated at various levels of school size illustrate that reading achievement is increasingly

[2] I introduced the subscript –57 to make explicit the particular value of X_2 at which the Y-on-X_1 slope is estimated. The specified value of X_2 is expressed as a deviation score: $X_2 - \bar{X}_2 = 16 - 72.68 = -56.58$ (rounded to –57).

related to poverty as school size increases, and decreasingly related to poverty as school size decreases.

Figure 3 shows the within-group regression lines for below- and above-median schools in per grade enrollment. As described above, I obtained these by splitting the school-size distribution at the median (42 students per grade) and, for each group of schools, fitting a reading-on-poverty regression line. These within-group regression lines further illustrate the interaction reported in Table 4: There is a flatter slope—a weaker relationship between reading achievement and poverty—for smaller schools than for larger schools. Indeed, the correlation for the former is $r = -.39$ versus $r = -.64$ for the latter, which, when squared, yield power ratings of 15% and 41%, respectively. Although there is considerable within-group variability evident in Figure 3 and, further, the nonparallel displacement of one regression line relative to the other is not great (particularly where most of the data are), there is some tendency for smaller higher-poverty schools to have reading achievement superior to that of larger higher-poverty schools.

Math. Table 5 shows the regression results for math, based on all schools. The pattern of results is similar to those obtained for reading. At Step 1, poverty is significantly related to math whereas school size is not. And at Step 2, the interaction of poverty and school size explains an additional 5% of variance in mathematics achievement ($\Delta R^2 = .0479, p < .01$): As with reading achievement, mathematics achievement is increasingly related to poverty as school size increases, and decreasingly related to poverty as school size decreases. For example, the math-on-poverty slope for median-size schools is $b_{-31} = -0.0860$ ($\beta_{-31} = -0.33$). In contrast, the simple slope for schools at the 25th percentile in school size is $b_{-57} = -0.0643$ ($\beta_{-57} = -0.25$) and, for schools at the 75th percentile, $b_{+32} = -0.1385$ ($\beta_{+32} = -0.53$).

Table 5
Regressing math on poverty, school size, and their product: All schools (n = 216).

Variables	b	s.e.	β	t	p	ΔR^2
Without interaction						
(Constant)	528.16					
Poverty	-0.1025	0.0177	-0.39	-5.78	< .01	
School size	-0.0039	0.0038	-0.07	-1.02	.31	
With interaction						
(Constant)	527.80					
Poverty	-0.1118	0.0175	-0.43	-6.40	< .01	
School size	-0.0080	0.0039	-0.14	-2.05	.04	
Poverty x size	-0.0008	0.0002	—	-3.53	< .01	.0479

Note. Poverty and school size were centered for this analysis. At Step 2, the unadjusted and adjusted values of R^2 are .187 and .176, respectively.

The within-group regression lines are presented in Figure 4, which shows the nonparallel displacement indicative of interaction. The math-on-poverty slope is flatter—signifying a weaker relationship—for smaller schools than for larger schools. The corresponding power ratings are, respectively, 4% for smaller schools ($r = -.19$) and 46% for larger schools ($r = -.68$).

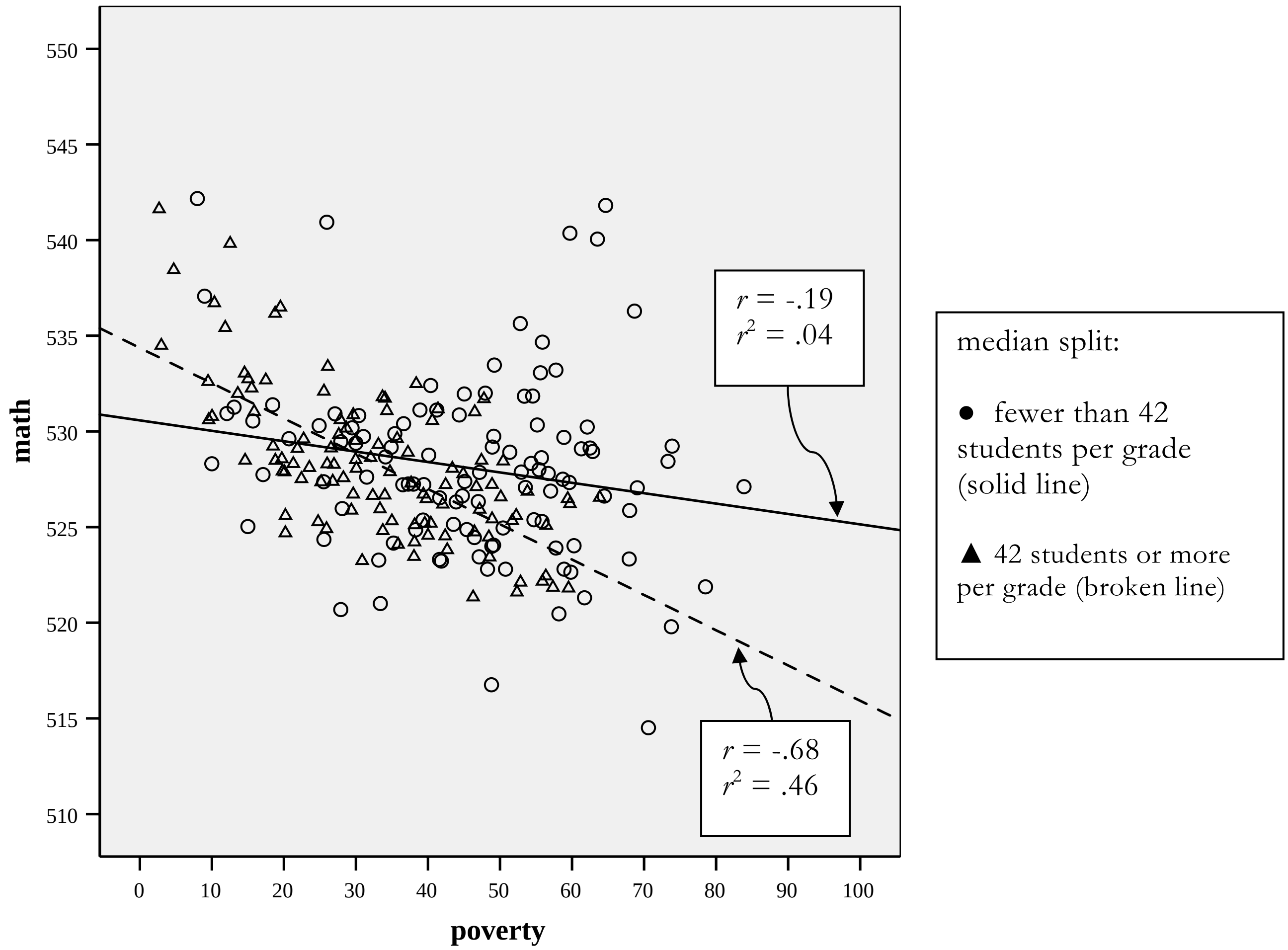


Figure 4. The interaction of poverty and school size (p =.001), math: All schools (n = 216).

The symmetry of b_3. As noted above, the statistical significance of b_3 indicates that the magnitude of the achievement-on-poverty slope (b_1) is a function of school size (X_2) and, symmetrically, the magnitude of the achievement-on-size slope (b_2) is a function of poverty (X_1). My emphasis thus far has been decidedly on the former, given its direct relevance to the concept of poverty's power rating which frames the present study. But many writers blur the distinction between the two interpretations, referring to one and then to the other as their argument develops. I (briefly) will follow suit.

Just as the simple slope for poverty (b_1) at a specified value of school size (X_2) is equal to $b_1 + b_3X_2$, the simple slope for school size (b_2) at specified value of poverty (X_1) is equal to $b_2 + b_3X_1$. At Step 2 of Tables 4 and 5, we see that school size has a negligible, if statistically significant, negative effect on both reading and math for schools of average poverty (i.e., X_1 = 0). But when the simple slope is calculated for a school where 23% of its students receive subsidized meals—approximately one standard deviation, or 17 percentage points, below the mean ($\bar{X}_1$ = 39.52)—school size is unrelated to achievement in either reading or math. Specifically, b_{-17} = 0.0007 and β_{-17} = 0.01 (p = .91) for reading; for math, b_{-17} = 0.0062 and β_{-17} = 0.11 (p = .20). Now consider a comparatively high-poverty school in which 73% of students receive subsidized meals (roughly two standard deviations, or 33 percentage points, above the mean).[3] Here, the effect of school size on reading is statistically significant and large: $b_{+33} = -0.0249$ and $\beta_{+33} = -0.49$ ($p < .01$). For math, the effect is larger still: $b_{+33} = -0.0355$ and $\beta_{+33} = -0.63$ ($p < .01$). Thus, with a standard deviation decrease in school size, reading achievement in these high-poverty schools—unlike their lower-poverty counterpart—increases by half a standard deviation, and math achievement increases almost two-thirds of a standard deviation. This finding, of course, merely restates the poverty-size interaction by focusing on the conditional effect of school size rather than the conditional effect of poverty.

Regression Analyses: Successively Less-Volatile Collections of Schools

To explore the possible operation of a statistical artifact due to the greater volatility in achievement among smaller schools, I repeated the regression analyses reported above for successively less-volatile collections of schools. Rather than exhaustively delineate these results for each value of the volatility measure, I report in Table 6 the primary statistic for each analysis: the increment in R^2 at Step 2 when the product term, X_1X_2, is introduced. I then provide additional details for the results based on the 104 least-volatile schools in reading achievement and the 104 least-volatile schools in math achievement.[4]

Table 6
Volatility in school achievement and the magnitude of ΔR^2

Reading				Math			
Volatility	*n*	ΔR^2	*p*	Volatility	*n*	ΔR^2	*p*
≤ 8	—	—	—	≤ 8	216	.0479	< .01
≤ 7	216	.0221	.01	≤ 7	215	.0470	< .01
≤ 6	214	.0216	.01	≤ 6	212	.0419	< .01
≤ 5	208	.0291	< .01	≤ 5	208	.0392	< .01
≤ 4	204	.0292	< .01	≤ 4	204	.0383	< .01
≤ 3	188	.0300	< .01	≤ 3	193	.0261	.01
≤ 2	166	.0422	< .01	≤ 2	164	.0274	.02
1	104	.0305	.03	1	104	.0139	.19

Note. ΔR^2 is associated with the introduction of the product term (poverty x size) at Step 2 of each regression analysis.

Reading. As Table 6 shows, the interaction between poverty and school size is unrelated to the volatility of school-level achievement in reading: For each successive analysis, the increment in explained variance associated with the introduction of the product term at Step 2 is statistically significant. Further, there is no evidence that ΔR^2, statistical significance notwithstanding, is systematically smaller when based on successively less volatile schools.

---

[3] One should interpret these derived slopes cautiously, of course, given the few data points at the upper end of the poverty scale.

[4] Roughly half (51%) of these 104 schools were "least volatile" in both reading and mathematics achievement.

Table 7
Descriptive statistics: Least volatile schools, reading achievement (n = 104)

				Intercorrelations	
Variable	*M*	*SD*	Range	Reading	Poverty
Reading	535.95	3.76	527.99, 545.95		
Poverty	38.78	15.98	2.68, 78.52	-.59*	
School size	89.19	79.67	2.94, 358.00	.09	-.35*

Note. For the purpose of this table, poverty and school size are in their original uncentered form (which affects only the mean and range).
* $p < .01$.

Tables 7 and 8 show descriptive statistics and regression results, respectively, based on the least-volatile schools in reading achievement ($n = 104$). Again, these are the schools for which mean achievement on the reading measure did not vary more than 2.5 points across the two years examined. The pattern of results here is similar to that reported earlier for all 216 schools, as are the within-group regression lines shown in Figure 5. Indeed, regarding the latter, poverty's power rating differential—16% for smaller schools vs. 42% for larger schools—is almost indistinguishable from the differential based on all schools (15% and 41%, respectively). With respect to reading achievement, then, the statistical-artifact hypothesis is not consistent with the data.

Table 8
Regressing reading on poverty, school size, and their product: Schools having minimal volatility in achievement (n = 104)

Variables	*b*	s.e.	β	*t*	*p*	ΔR^2
Without interaction						
(Constant)	535.94					
Poverty	-0.1492	0.0200	-0.63	-7.45	< .01	
School size	-0.0065	0.0040	-0.14	-1.61	.11	
With interaction						
(Constant)	535.72					
Poverty	-0.1411	0.0200	-0.60	-7.067	< .01	
School size	-0.0074	0.0040	-0.16	-1.875	.06	
Poverty x size	-0.0005	0.0002	—	-2.237	.03	.0305

Note. Poverty and school size were centered for this analysis. At Step 2, the unadjusted and adjusted values of R^2 are .390 and .372, respectively.

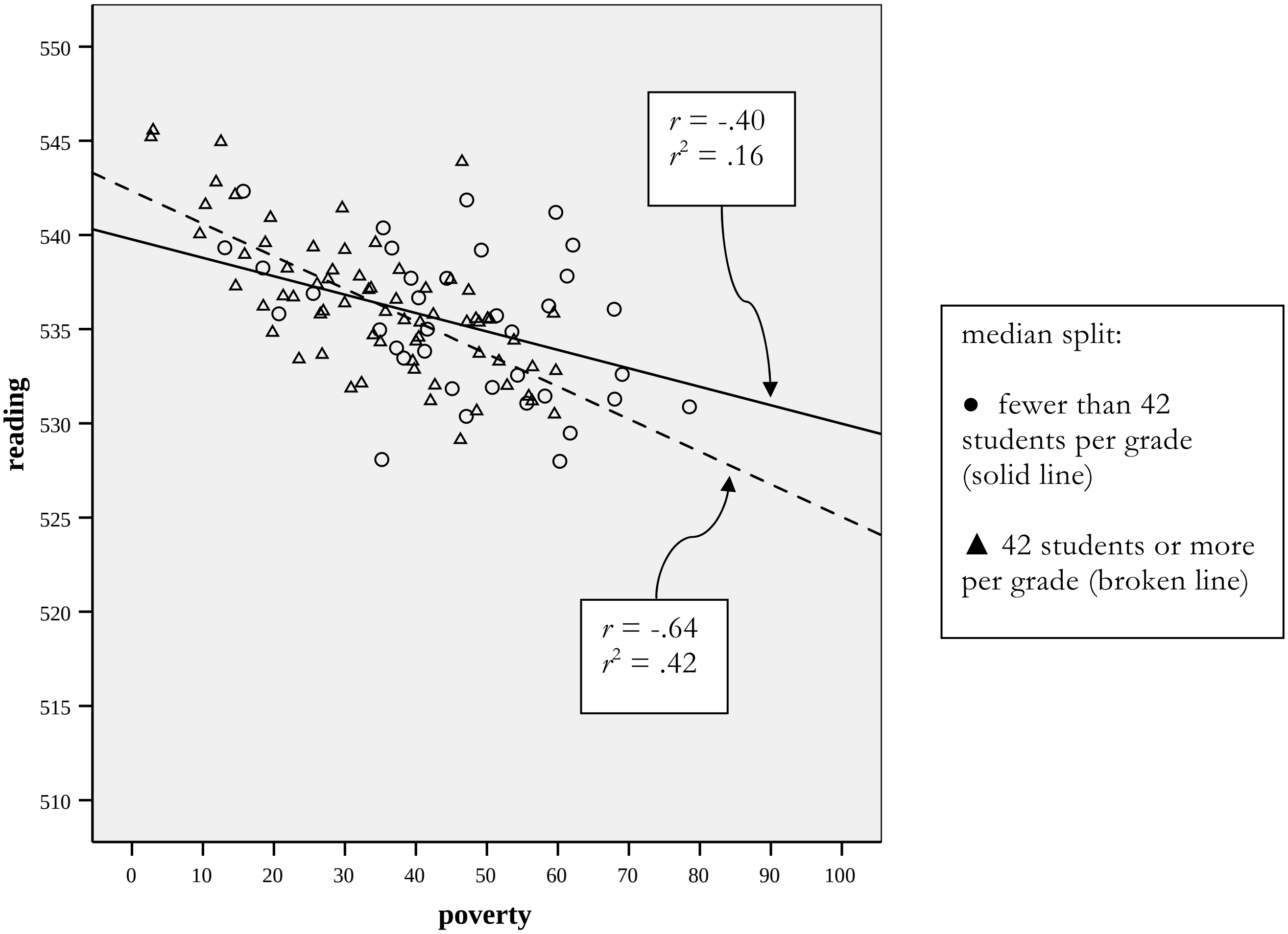


Figure 5. Interaction of poverty and school size (p = .001), reading: Schools having minimal volatility in achievement (n = 104).

Math. A different picture emerges with mathematics achievement, where we see a gradual decline in ΔR^2 with successively less-volatile collections of schools (Table 6)—to the point of statistical nonsignificance when based on the 104 least-volatile schools ($\Delta R^2 = .0139$, $p = .19$). Tables 9 and 10 present the relevant statistics for the latter analysis, where, at Step 2 of Table 10, we see the statistically nonsignificant slope for the product term.

Table 9
Descriptive statistics: Least volatile schools, math achievement (n = 104)

				Intercorrelations	
Variable	*M*	*SD*	Range	Math	Poverty
Math	527.60	4.26	514.51, 542.17		
Poverty	38.25	14.71	7.99, 73.89	-.41*	
School size	82.28	81.72	3.39, 327.50	.06	-.30*

Note. For the purpose of this table, poverty and school size are in their original uncentered form (which affects only the mean and range). The schools in Tables 9 and 10 are the not same 104 schools represented in Tables 7 and 8. (See footnote 4.)
* $p < .01$.

The within-group regression lines are shown in Figure 6. While the power ratings of poverty show some differential between smaller and larger schools, it derives from a poverty-size interaction that failed to reach statistical significance and, therefore, reflects chance variation. Between the general decline in ΔR^2 values (Table 6) and the absence of a statistically significant poverty-size interaction when based on the least volatile schools (Table 10), the statistical-artifact hypothesis is consistent with the data in the case of mathematics achievement.

Table 10
Regressing math on poverty, school size, and their product: Schools having minimal volatility in achievement (n = 104)

Variables	*b*	s.e.	β	*t*	*p*	ΔR^2
Without interaction						
(Constant)	527.47					
Poverty	-0.1249	0.0278	-0.43	-4.54	< .01	
School size	-0.0038	0.0050	-0.07	-0.76	.45	
With interaction						
(Constant)	527.31					
Poverty	-0.1309	0.0278	-0.45	-4.71	< .01	
School size	-0.0070	0.0055	-0.13	-1.26	.21	
Poverty x size	-0.0005	0.0004	—	-1.31	.19	.0139

Note. Poverty and school size were centered for this analysis. At Step 2, the unadjusted and adjusted values of R^2 are .186 and .162, respectively. The schools in Tables 9 and 10 are the not same 104 schools represented in Tables 7 and 8. (See footnote 4.)

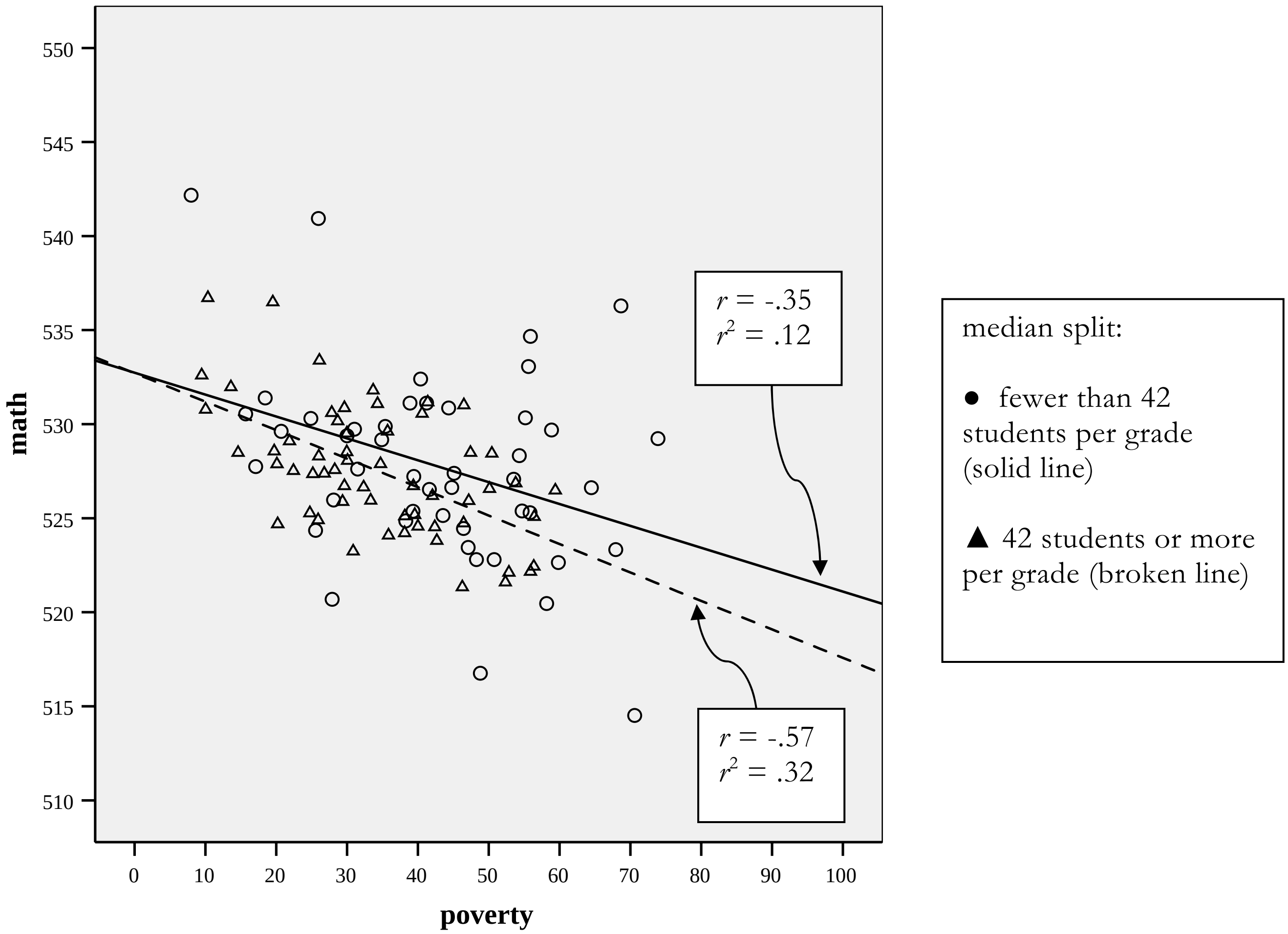


Figure 6. No interaction of poverty and school size (p = .193), math: Schools having minimal volatility in achievement (n = 104)

Discussion

The question posed in the subtitle of this article—substantive finding or statistical artifact?—does not permit an unequivocal answer. When the dependent variable is reading achievement, I find no support for my hypothesis that poverty's power rating is lower in smaller schools because of their greater volatility (lower reliability) in school-level achievement. Thus, the celebrated interaction of socioeconomic status and school size clearly stands with respect to eighth-grade reading achievement in these Maine schools. But for mathematics achievement, the statistical-artifact hypothesis is supported: Among these smaller schools, the lower reliability of school-level achievement appears to be a plausible explanation of the reduced power rating of poverty for these schools.

Unfortunately, the latter conclusion is complicated by plausible rival hypotheses of its own. Two problems immediately come to mind. First, my achievement-volatility measure does not distinguish between random variation and variation due to educational practice. Some of the high-discrepancy schools in Figure 2, as reflected in their alignment on the vertical axis, doubtless are revealing real—not random—improvement or decline in achievement. By treating all variation as *random* variation, I inevitably exclude some schools from the analysis that should have been included (were it possible to make this distinction in practice). That said, the results are not systematically biased as a consequence, insofar as the absence of "real improvement" schools is offset by the absence of "real decline" schools.

The second problem is of greater concern. By conducting the regression analyses on successively less-volatile collections of schools, and because achievement volatility is more pronounced among smaller schools (Figure 2), I successively compromise the full representation of smaller schools as well. For example, 26 schools in the full sample (12.0%) had fewer than 10 students per grade. Among the least volatile schools in mathematics achievement, however, only 6 schools (5.8%) were this small. Would the results of the final analyses, where achievement volatility is minimum, likely differ had the smallest schools been fully represented? It is difficult to say. In short, I arguably exclude some of the very schools required for an adequate test of my statistical-artifact hypothesis (and, in doing so, introduce a certain irony into the present study).

Yet this second problem—the successive underrepresentation of smaller schools—had no effect on the viability of the poverty-size interaction for reading achievement. This inconsistency presents an interesting challenge: how to explain it. If one is inclined to dismiss my findings for mathematics achievement because of this underrepresentation, then the challenge is to explain why a similar outcome was not obtained for reading achievement. After all, smaller-school underrepresentation operates there as well. So, what is it about reading achievement (or related instruction) that makes the poverty-size interaction immune to the successive underrepresentation of smaller schools in these analyses? Or, if one prefers, what is it about mathematics achievement (or related instruction) that makes the poverty-size interaction particularly vulnerable in this regard?

On the other hand, for those readers whose confidence in the statistical-artifact results for mathematics achievement is unshaken by this underrepresentation, the corresponding challenge is to explain why the statistical-artifact hypothesis did not prevail for reading achievement. After all, reading achievement is not appreciably less volatile than mathematics achievement. So, what is it about reading achievement (or related instruction) that explains this apparent invincibility—a greater robustness—of the poverty-size interaction?

Unfortunately, I cannot answer these questions. But insofar as I cannot explain, even with the benefit of hindsight, a statistical-artifact finding that would surface only for mathematics achievement, I am inclined to attach greater import to the successive underrepresentation of smaller schools in these analyses than I had at the outset. Although I cannot explain why this underrepresentation has no concomitant effect on the poverty-size interaction with respect to reading achievement, this anomaly presently perplexes me less than does a mathematics-specific statistical artifact. Furthermore, it is only in the final, most restrictive analysis—where a sizeable number of the smallest schools are lost—that the poverty-size interaction for mathematics achievement fails to reach statistical significance.

In view of these considerations, then, I conclude that my results are insufficient to support the statistical-artifact hypothesis with respect to mathematics achievement. Although this conclusion is not as unequivocal as that for reading achievement, I nevertheless believe it is the reasonable conclusion given the considerations above. In short, the celebrated interaction of poverty and school size has survived a sincere attempt to empirically cast doubt on it. Consequently, we can have greater confidence in this interaction than I believe was warranted before.

Further tests of the statistical-artifact hypothesis would be informative, if only to show that my somewhat equivocal results for mathematics achievement are a mere anomaly. In this spirit, I encourage other researchers who have explored the mitigating-effect phenomenon to conduct, where possible, (re)analyses of their own with the inclusion of an achievement-volatility measure.

If the interaction of socioeconomic status and school size is accepted as an established (if modest) phenomenon, we nonetheless are left wanting for a credible explanation of it. Such an explanation seemingly would draw on the mechanisms through which smaller schools facilitate student achievement and related outcomes, but, unfortunately, we are wanting there as well. As Fowler and Walberg (1991) said in reference to the then-extant research,

> "[a]lthough these studies show a positive relationship between small school size and student outcomes, they do not suggest *why* this may occur. In other studies, which only peripherally included school size, researchers have *suggested* reasons for the beneficial effect that small school size has upon student outcomes" (p. 191; emphasis added).

The situation has changed little in the ensuing 15 years.

As an influence on student achievement, school size clearly is a proxy rather than a causal force in and of itself. To offer credible explanations for the poverty-size interaction, then, we first need stronger evidence regarding the mechanisms—the mediating variables—through which school size putatively influences student achievement (McMillen, 2004, p. 20). Howley (2002, p. 62) offers "care, attention, and respect" as possible mechanisms; Lee and Smith (1997, p. 219) refer to "the academic and social organization and functioning of schools." Doubtless there are other context- and process-related forces at play as well. Whatever the focus, a warranted claim about its relationship to both school size and student achievement must be based on careful empirical investigation, not on casual observation, anecdotal reports, reasonable (but untested) hypotheses, popular opinion, or the will to believe. We need more descriptive research like that conducted by Howley and Howley (2006) and Lee, Smerdon, Alfeld-Liro, and Brown (2000), which should be followed up by analyses that exercise the statistical control necessary to test hypotheses that fundamentally get at cause-and-effect relationships.

Equipped with empirically established mediating variables regarding the relationship between school size and student achievement, we can then craft defensible conjectures regarding the poverty-size interaction. In this regard, of course, one's central obligation will be to argue *why* a mediating variable would be expected to differentially affect student achievement as a function of student SES. For example, if the accumulation of evidence from sound empirical research were to show that smaller schools are characterized by more personalized social relations and, in turn, that these more personalized social relations improve student outcomes, our obligation is to cogently argue why lower-SES students would benefit from such social relations more than higher-SES students would. These conjectures should then be subjected to empirical tests of their own. One could introduce a set of social-relations variables into the full regression equation (in the tradition above) to see whether the poverty-size interaction disappears—as it would if the poverty-size interaction is in fact due to social relations.

In any case, well-crafted arguments followed by equally well-crafted investigations—both premised on warranted claims regarding the mechanisms through which school size influences student achievement—should be the direction of future research on the poverty-size interaction.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, MA: Sage.

Butler, L., Carr, R., Cook, K., Gould, R., Keenan, F., MacArthur, S., Power, L., & Ritchie, K. (2005, June 8). Consolidation can't save money. [Letter to the editor.] *Bangor Daily News*, p. A9.

Coladarci, T. (2003). *Gallup goes to school: The importance of confidence intervals for evaluating "Adequate Yearly Progress" in small schools.* Policy Brief. Washington, D.C.: The Rural School and Community Trust. Retrieved October 26, 2006, from http://files.ruraledu.org/docs/nclb/coladarci.pdf.

Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, *57*, 373–399.

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

Friedkin, N. E., & Necochea, J. (1988). School system size and performance: A contingency perspective. *Educational Evaluation and Policy Analysis*, *10*, 237–249.

Hill, R. K. (2002, April). *Examining the reliability of accountability systems*. Paper presented at the 2002 meeting of the American Educational Research Association, New Orleans. Retrieved September 25, 2003, from http://www.nciea.org/publications/NCME_RHCD03.pdf.

Hill, R. K., & DePascale, C. A. (2003). Reliability of No Child Left Behind Accountability Designs. *Educational Measurement: Issues and Practices*, *22*(3), 12–20.

Howley, C. B. (1996). Compounding disadvantage: The effects of school and district size on student achievement in West Virginia. *Journal of Research in Rural Education*, *12*(1), 25–32.

Howley, C. B. (2002). Small schools. In A. Molnar (Ed.), *School reform proposals: The research evidence* (pp. 49–77). Greenwich, CT: Information Age Publishing.

Howley, C. B., & Bickel, R. (1999). *The Matthew Project: National Report*. Athens, OH: Ohio University, Educational Studies Department. (ERIC Document Reproduction Service No. ED 433 174).

Howley, A., & Howley, C. B. (2006). Small schools and the pressure to consolidate. *Education Policy Analysis Archives, 14*(10). Retrieved March 29, 2006, from http://epaa.asu.edu/epaa/v14n10/.

Huang, G., Howley, C. B. (1993). Mitigating disadvantage: Effects of small-scale schooling on student achievement in Alaska. *Journal of Research in Rural Education*, *9*(3), 137–149.

Johnson, J. D., Howley, C. B., & Howley, A. A. (2002). *Size, excellence, and equity: A report on Arkansas schools and districts*. Athens, OH: Ohio University, Educational Studies Department. (ERIC Document Reproduction Service No. ED 459 987).

Kane, T.J., & Staiger, D.O. (2002). Volatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), *Brookings Papers on Education Policy 2002* (pp. 235–283). Washington, DC: Brookings Institution.

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.

Linn, R. L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis*, *24*(1), 29–36.

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

Phillips, D. C. (2000). *The expanded social scientist's bestiary*. Lanham, MD: Rowman and Littlefield.

Maine's small schools cut poverty's power over student achievement. (2005, February). *Rural Policy Matters*, *7*(2), pp. 3, 5. Retrieved October 30, 2006, from http://www.ruraledu.org/atf/cf/%7BF4BE47E7-FA27-47A8-B662-8DE8A6FC0577%7D/rpm7_2.pdf.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417–453.

Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin.

Tompkins, R. (2006). Small schools, small districts: Good for rural kids, economies, and democracy. *Rural Americans* (Issue 14). Retrieved June 19, 2006, from http://www.demos.org/pubs/KitchenTable014.pdf.

White, K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*, 461–481.

About the Author

Theodore Coladarci
University of Maine

Email: coladarci@umit.maine.edu

Theodore Coladarci is Professor of Educational Psychology at the University of Maine. Since 1992, he has served as editor of *Journal of Research in Rural Education* (http://www.umaine.edu/jrre/). An earlier version of this work was presented at the 2006 meeting of the American Educational Research Association, and the author wishes to thank the discussant, Aimee Howley, for her thoughtful comments and suggestions. The author also is grateful for the feedback provided by Deb Allen, Sandy Ervin, Ed Kame'enui and the four anonymous reviewers.