

## Do smaller schools really reduce the “power rating” of poverty?

Theodore Coldarci

*University of Maine*

*The percentage of variance in student achievement that is explained by student SES—“poverty’s power rating,” as some call it—tends to be less among smaller schools than among larger schools. Smaller schools, we are told, are able to somehow disrupt the association between SES and student achievement. Using eighth-grade data for 215 public schools in Maine, I explored the hypothesis that this finding is in part a statistical artifact of the lower reliability of school-aggregated student achievement in smaller schools. This hypothesis was supported for mathematics achievement but seemingly not for reading achievement. Implications are discussed.*

As every student of education research knows, the positive relationship between student achievement and socioeconomic status (SES) is well-established: Higher-SES students tend to achieve more highly than lower-SES students (e.g., Sirin, 2005). Nevertheless, 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, 2002; McMillen, 2004; also see Friedkin & Necochea, 1988; Lee & Smith, 1997). That is, 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.

A common way to illustrate this interaction is to show that the correlation between SES and achievement (calculated with the school as the unit of analysis) 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 the Rural School and Community Trust calls it, poverty’s “power rating”—the percentage of SES-explained variance in student achievement—is lower for smaller schools than it is for 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). Indeed, Johnson, Howley, and Howley (2002) declared this finding to be “among the most consistent ever to be reported in educational research” (pp. 36–37). In the words of a Maine school superintendent and his colleagues, “[s]mall schools are an antidote to the impact of poverty on school achievement” (Butler et al., 2005, p. A9).

I must confess that, despite my affinity to rural schools and communities, I have always been uneasy with this finding. 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 higher-poverty students, and as much as I can imagine the many ways in which smaller schools might be able to pull this off, my immediate suspicion was that the weaker SES-achievement correlation among smaller schools may have

little to do with student experience in such schools. Rather, I suspected a statistical artifact at play.

Just what is a statistical artifact? It is where a research result is misleading because of an artificial or extraneous effect due to statistical considerations. For example, imagine that the values on variable *X* do not vary much and, in turn, we find that there is absolutely no correlation between this variable and variable *Y*. The absence of relationship between *X* and *Y* very well could be due to insufficient variance 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 assumed role of smaller schools in weakening the SES-achievement relationship would be a statistical artifact if, say, there were much less variance in either student SES or student achievement among smaller schools than among larger schools. This in fact was my immediate suspicion, but I subsequently ruled it out when I was unable to find evidence of restricted variance in the statistics reported by the researchers. Further, I found no evidence of restricted variance in a quick analysis of Maine data that had been featured in a 2005 Rural Trust news release regarding the poverty power-rating phenomenon (Rural School and Community Trust, 2005).

My interest in the challenges that small schools face in complying with 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). In short, school achievement often jumps around quite a bit from one year to the next in smaller schools, whereas larger schools enjoy much greater stability in this regard (e.g., Hill & DePascale, 2003; Kane, Staiger, & Geppert, 2002; Linn & Haug, 2002). At issue here is the *reliability* of school-aggregated student achievement. Insofar as any measure of school achievement is less reliable—i.e., less stable—for a smaller school than for a larger school and, further, because a measure’s reliability places an upper limit on its 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. This is the conjecture I investigated in the present study.

**Method**

*Data Source and Variables*

My focus is on eighth-grade achievement in Maine 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 is 501–580.) 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 meals (*poverty*). As for school size, I determined the mean enrollment per grade for each school, averaged across 2002–2003 and 2003–2004 (*size*).

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 (*volatility*). There were separate volatility ratings for reading and math, and both were constructed as shown in Table 1.

Table 1  
*Volatility definitions.*

Volatility rating	Change in school mean-achievement <sup>a</sup>
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

<sup>a</sup>The scale of the Maine Educational Assessment ranges from 501 to 580.

*Analyses*

I restricted my analyses to public schools in Maine that (a) had an eighth grade in 2002–2003 and 2003–2004, (b) had data on all variables for both 2002–2003 and 2003–2004, and (c) had 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. I began by testing for the interaction between socioeconomic status and school size. I did so using ordinary least squares regression (e.g., Aiken & West, 1991), where I regressed math and reading (in separate analyses) on three independent variables: poverty, size, and their mathematical product (i.e., poverty x size). The statistical significance of the product term indicates the presence of a poverty-size interaction—that the degree of association between poverty

and student achievement depends on school size, or, equivalently, that the degree of association between school size and student achievement depends on the socioeconomic status of the school.

To illustrate the magnitude of this interaction, I did a median split on school size and then regressed reading and math (separately) on poverty for below-median schools and for above-median schools. 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, is the aforementioned power rating of poverty.

To explore my statistical-artifact hypothesis—that poverty’s reduced power rating, when examined among smaller schools, reflects in part the lower reliability of school-level achievement in smaller schools—I repeated these analyses on successively less-volatile collections of schools. The first set of analyses included all 216 schools (i.e., volatility = 1, 2, 3, 4, 5, 6, 7, or 8); the second set included schools for which volatility = 1, 2, 3, 4, 5, 6, or 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 disappear among schools having the least volatility.

**Results**

I begin with a brief note on the well-established relationship between school size and achievement volatility, which is clearly evident in the present data. As Figure 1 shows, there are wide variations in achievement from one year to the next for smaller schools. For the smallest

schools, mean achievement can vary by almost 20 points in one direction or the other (on a test whose scale is 501–580). Larger schools, in contrast, demonstrate considerably more stability. (See Coladarci, 2003, for a discussion of the corresponding implications for the adequate-yearly-progress requirement of No Child Left Behind.)

The distribution of the 8-point volatility ratings are shown in Table 2 for both reading and math. Each distribution reflects extreme positive skew: While the vast majority of these 216 schools have rather stable levels of achievement ( $\pm 5$  points from one year to the next), some schools' achievement vary widely in this regard. Only one school falls in the highest volatility category for mathematics achievement; none does for reading achievement.

Table 2  
*The frequency distribution of volatility ratings for reading and math*

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

(n = 216)

*All Schools*

The first set of analyses is based on all schools, irrespective of volatility. Table 3 presents descriptive statistics for reading, math, poverty, and size. As would be expected, schools vary with respect to both poverty and size,

reading and math correlate highly, and reading and math each correlates moderately with poverty. Smaller schools are somewhat more likely to be located in higher-poverty communities ( $r = -.34$ ), and school size is unrelated to achievement ( $r = .07, p = .16$ ).

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

	<i>M</i>	<i>SD</i>	Range	<u>Intercorrelations</u>		
				Reading	Math	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*	
Size	72.78	77.31	2.94, 358.00	.07	.07	-.34*

\*  $p < .01$ .

*Reading.* With reading as the dependent variable, the interaction between poverty and size is statistically significant ( $t = -2.52, p < .05$ ). Figure 2 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. Consistent with the statistically significant interaction from the regression analysis, Figure 2 reveals 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. That is, poverty explains only 15% of the variance in reading achievement among smaller schools versus 41% among larger schools.

*Math.* A similar pattern of results is found for math. The statistically significant ( $t = -3.53$ ,  $p < .01$ ) poverty-size interaction from the regression analysis is illustrated in Figure 3. As with reading, 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 ( $r = -.68$ ).

Thus, for both achievement measures, the familiar interaction between poverty and school size clearly surfaces when all schools are included in the analysis. Consistent with popular rhetoric, the power rating of poverty is considerably weaker among smaller schools than among larger schools.

#### *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 analyses reported above for successively less-volatile collections of schools. Rather than exhaustively delineate the results for each value of volatility, I instead characterize the upshot of these analyses. *Reading.* The poverty-size interaction is statistically significant for each successive analysis, even when assessed on the 104 least volatile schools ( $t = -2.24$ ,  $p < .05$ ). Consider Figure 4, for example, which shows the within-group regression lines for these least volatile schools. Poverty's power rating differential here—16% for smaller schools vs. 42% for larger schools—is virtually indistinguishable from the differential based on all schools (15% and 41%, respectively). With respect to reading achievement, then, my statistical-artifact hypothesis is not supported: When the lower reliability of school-level achievement among smaller schools is taken into account, these schools still enjoy a reduced power rating of poverty.

*Math.* The picture is different for mathematics achievement. Although statistically significant, the magnitude of the poverty-size interaction systematically declines with each successive analysis. In the final analysis, based on the 104 least volatile schools, this interaction fails to reach statistical significance ( $t = -1.31$ ,  $p = .19$ ). Thus, my statistical-artifact hypothesis is supported when the dependent variable is mathematics achievement.

#### Discussion

The celebrated interaction of socioeconomic status and school size clearly stands with respect to eighth-grade reading achievement in these Maine schools. Here, the statistical-artifact hypothesis fails its test. In contrast, the statistical-artifact hypothesis is supported when the dependent variable is mathematics achievement. For eighth-grade mathematics achievement, then, poor reliability appears to be a plausible explanation of the reduced power rating of poverty among these smaller schools.

Ironically, the latter conclusion is complicated by a possible statistical artifact of its own. Specifically, by conducting analyses on successively less-volatile collections of schools, I successively compromised the full representation of small schools as well. This, of course, is because achievement volatility is more pronounced among smaller schools (Figure 1). In short, I am excluding some of the very schools required for a fair test of my statistical-artifact hypothesis. Yet this second problem—the successive under representation of small schools—had no effect on 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 under representation of small schools, then the challenge is to explain why I did not obtain a similar outcome for reading achievement. That is, what is it about reading achievement that makes the poverty-size interaction immune to the successive under representation of small schools in these analyses? On the other hand, for the reader whose confidence in the statistical-artifact results for mathematics achievement is unshaken by this under representation problem, the challenge is to explain why the statistical-artifact hypothesis did not prevail for reading achievement. That is, what is it about reading achievement that makes the poverty-size interaction immune to the volatility of achievement among smaller schools?

Because I cannot explain a statistical-artifact finding that is specific to mathematics achievement, I am inclined to attach greater significance to the successive under representation of small schools in these analyses than I had at the outset of this investigation. I fail to understand why this under representation does not affect the poverty-size interaction for reading achievement, but this confounds me less than does a mathematics-specific statistical artifact. Moreover, it is only in the most restrictive analysis—where a sizeable number of small 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 unequivocally support the statistical-artifact hypothesis with respect to mathematics achievement. Although this conclusion is far less straightforward than that for reading achievement, it is the reasonable conclusion all things considered.

In planning this study, I was not motivated by a desire to debunk popular opinion regarding the ameliorative effect that smaller schools have on the achievement-related disadvantages traditionally associated with poverty. Rather, I simply wished to determine whether a celebrated proposition in the rural education literature could withstand a constructive attempt to falsify it. And it did. Consequently, we can have greater confidence—greater warranted confidence—in the poverty-size interaction than we were entitled to before.

### 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. Available at <http://files.ruraledu.org/docs/nclb/coladarci.pdf>.
- 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., & 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*. (ERIC Document Reproduction Service No. ED433174).
- 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*. (ERIC Document Reproduction Service No. ED459987).
- Kane, T. J., Staiger, D. O., & Geppert, J. (2002). Randomly accountable. *Education Next*, 2(1), 57-61. Retrieved January 15, 2003, from <http://www.educationnext.org/20021/56.html>
- 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/>
- Rural School and Community Trust. (2005). *Maine's smaller schools cut poverty's power over student achievement*. [January 11, 2005 news release.]
- 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 February 22, 2006, from <http://www.demos.org/aroundthekitchentable/article.cfm?edition=014&article=education>

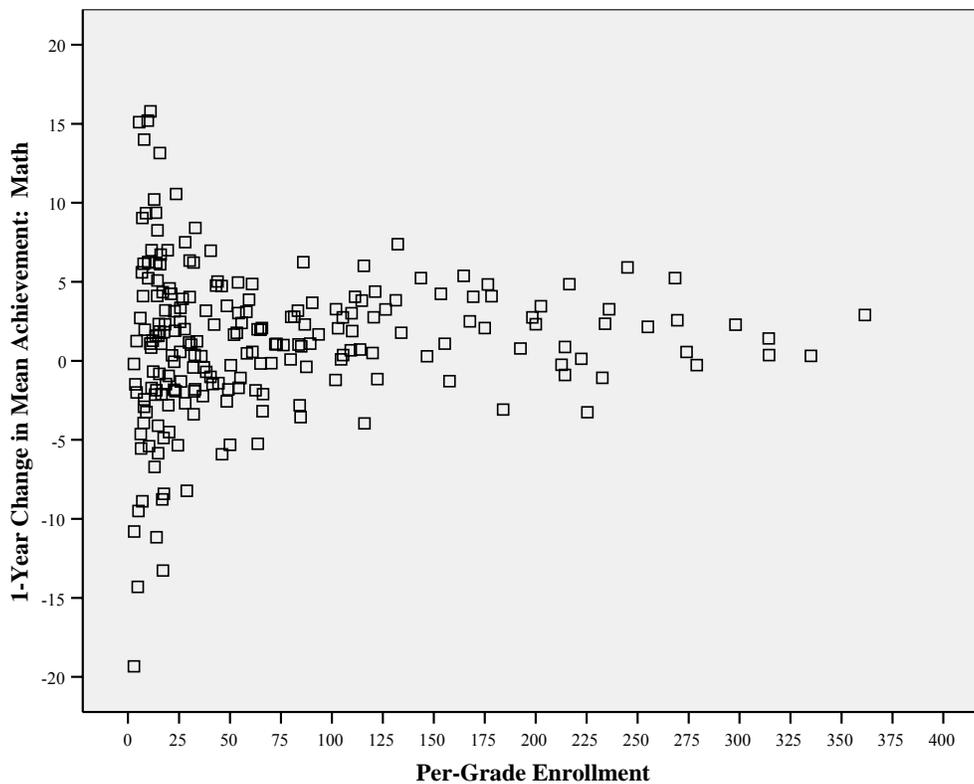
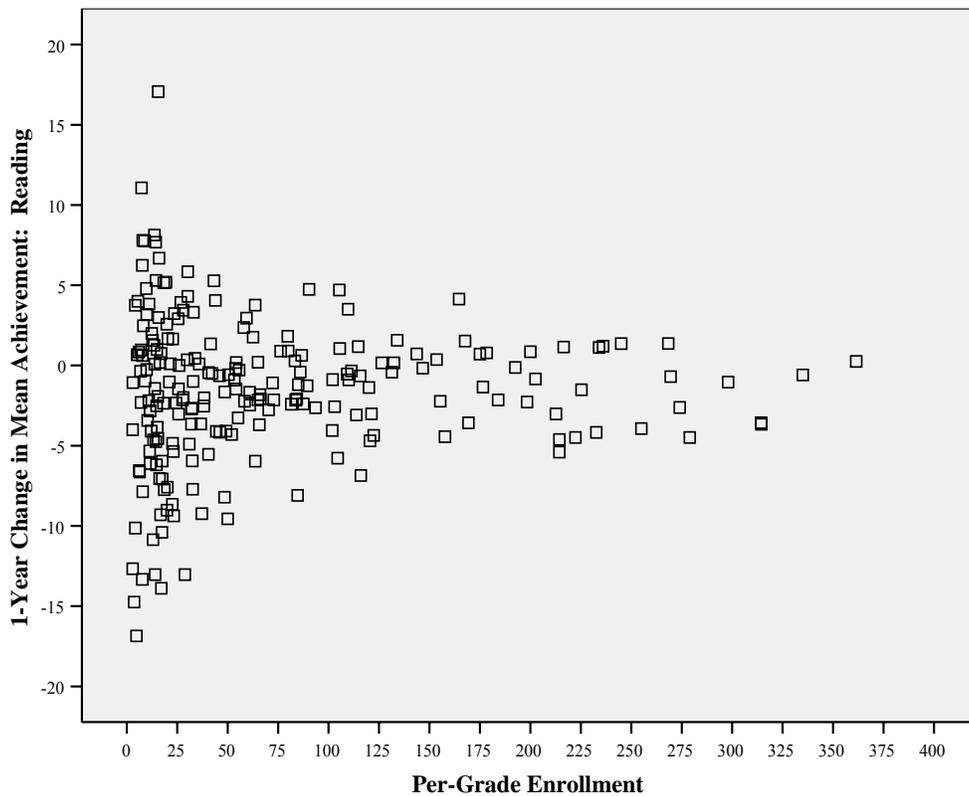


Figure 1. The relationship between school size and the one-year change in achievement for reading (top) and for mathematics (bottom).

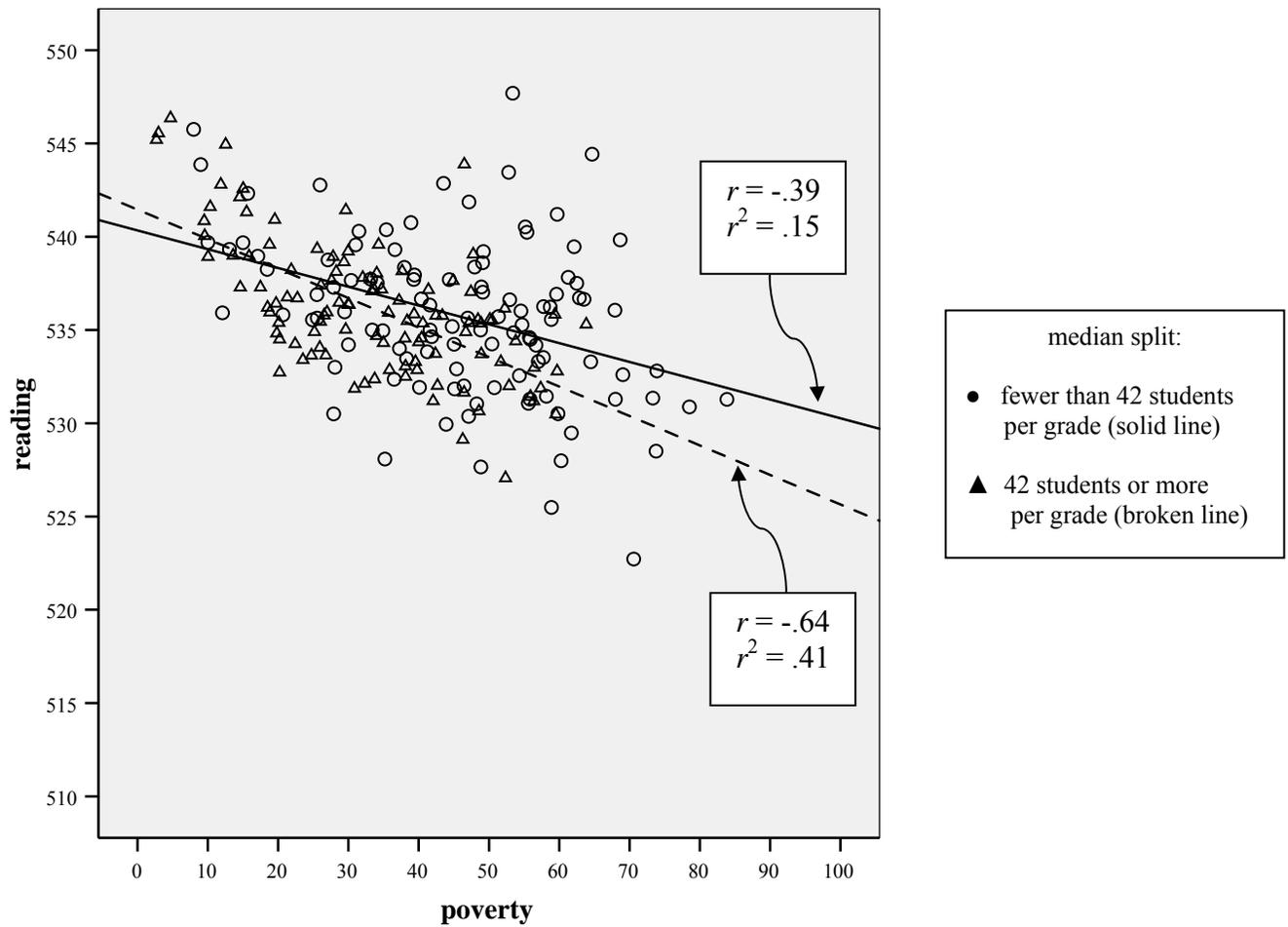


Figure 2. Within-group regression lines (all schools,  $n = 216$ ): reading.

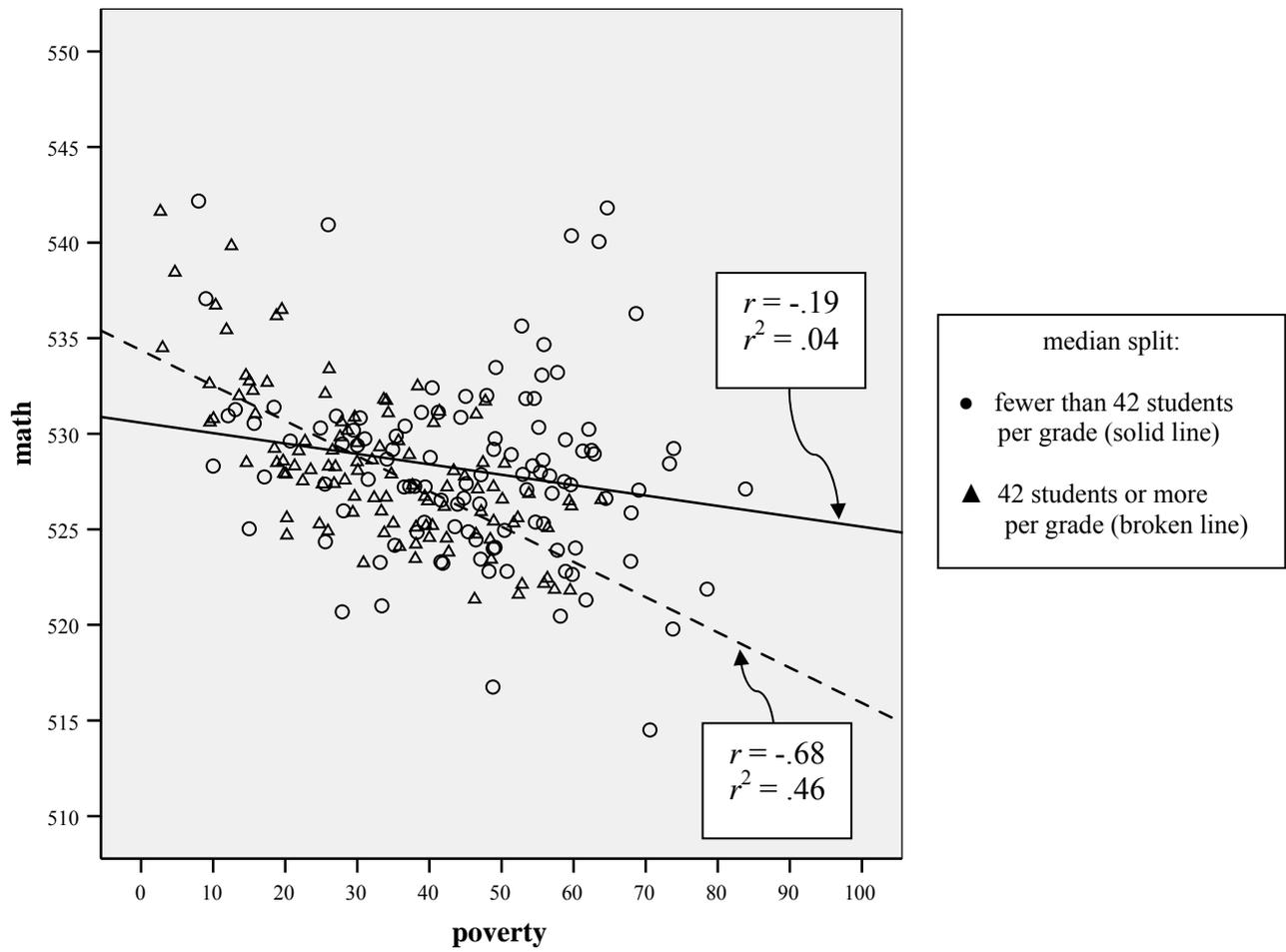


Figure 3. Within-group regression lines (all schools,  $n = 216$ ): math.

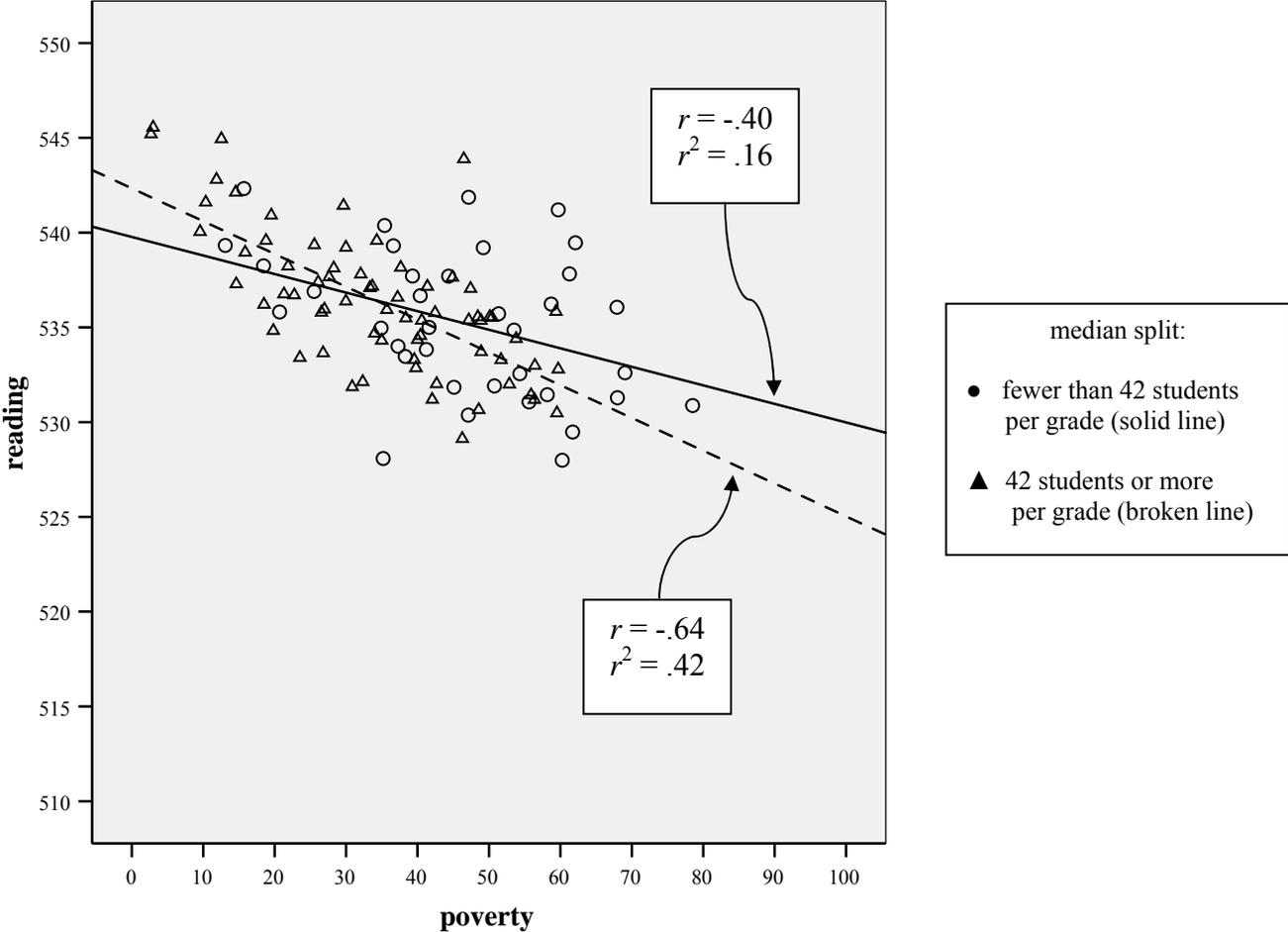


Figure 4. Within-group regression lines (least volatile schools,  $n = 104$ ): reading.