*Research in Rural Education, Volume 2, Number 2, 1983*

# Paradise Regained: An Apodictic Analysis of the Relationship Between School Size and Public Achievement

ARTHUR COLADARCI[1]

Approximately four score score and tweny years of research on the effects of institution size on pupil progress have produced only a literature of disagreement. Young theory builders seeking tenure and old administrators seeking guidance find no comfort in the conflicting conclusions and ambiguities. A critical examination of this confusing body of research, however, reveals evidentiary and inferential errors, naivité, intellectual puritanism, and rational extravagance. Revised and reinterpreted, this literature confesses a clear and near-unanimous finding: the smaller unit definitely is superior to the larger in pupil achievement. Two new studies add unnecessary confirmation; a startling afterthought will interest doctoral candidates seeking significant but painless dissertation topics.

The controversy about the relative merits of large and small schools has a long history. It was heard among the dying gasps of the Roman Empire [2], and the debate probably is as old as the droll assumption that many people want to learn and some can teach. Interest in this matter grew at an exponential rate following Conant's avuncular advice that smallness is no virtue [3]; if became somewhat strident when Coleman's [4] computers entered the fray. "The old battle," as Dunne [5] calls it with an undertone of affection, continues. Her identification of the opposing armies as "consolidators vs. conservators" neatly connotes that the central issues go beyond those that can be resolved merely by measuring and counting. Many observers have noted the nature and variety of considerations here—e.g., Sher [6] amiably, Gump and Barker [7] grimly, and Matthews [8] tersely. The question of whether "small" or "large" is the path to salvation appears to be complex.

I devoutly believe that complexity, like Italian cuisine, has a bright future. The literature on this topic, however, tempts one to agnosticism. There are many difficulties with it. At the most general level, as is so often and sorrowfully the case in educational research, the analyses tend to be more interesting than instructive. Good questions pursue and frequently catch bad data; many scholars ignore the Talmudic warning that "for example" is not evidence; the explanatory power of some findings come close to John Bradbury's conclusion that "light is darkness, lit up" [9]; and there are those who apparently take their cue from Sir Thomas Browne's 17th century discovery that proverbs are more powerful than demonstrations [10]. To be sure, there is no dearth of facts in the presentations, but the oft-times implication is that Morgan and Langford [11] were right when they wrote that "facts are trifling things compared with the immense weight of what one knows 'instinctively'" [p. 11]. We may add that some of the proffered information, although of virginal validity, has as much explicit relevance to anything as does the finding that the 11th century Archbishop Adalbert of Breman did not bathe [12]. It seems also (to close this woeful recital) that not a few of the research tribe ("the tough minded") base their designs and methodologies on the curious assumption that scientific inquiry must be rational. They give no ear to Feyerbend's *caveat* [13] that excluding the irrational and the non-science is detrimental to knowledge-building. They, unlike Broad and Wade [14], have not learned that "[w]here the conventional [scientific] ideology goes most seriously astray is in its focusing on the process of sciences instead of on the motives and needs of scientists" [p. 19]. In short, a generous portion of published research on "small vs. large" is like the peace of God: it passes all understanding.

However, I come not to excoriate sinners, but to save them—and salvation is at hand. A closer examination of the literature in this domain shows that the research and analyses addressed to a specific subset of the general controversy contain clear signals amongst the noise, once they are correctly and properly interpreted. To wit: is there an association between the size of a school and the achievement of its pupils? I modestly propose here to lay that question to rest (a) by rectifying earlier findings on this question and (b), in an uncontrollable fit of empiricism, by producing heretofore unpublished and unambiguous data. For those who remain rapt or patient to the end, there is a closing lagniappe.

## Relevant Research, Heuristically Interpreted

Published research on the relationship between school size and pupil achievement, like Caesar's Gaul, comprises three parts: (a) a few studies that find the relationship

[1]Professor of Education and Psychology, from the School of Education, Stanford University, Stanford, California 94305.

©1984 College of Education, University of Maine at Orono

to be inverse (i.e., smallness is associated with high achievement); (b) a few that find the converse of the inverse; and, (c) the largest number, those that mournfully sigh, "no significant difference." The causal browser will conclude that small schools have no clear advantage over their larger counterparts with respect to the effectiveness of instruction. However, as careful observers know full well, the small school *really is* much more effective. Something clearly is amiss in most of the investigations reported thus far. The following critical consideration of each of the three categories of findings identifies and corrects the errors. When these are remedied, the reports yield acceptable conclusions.

*Studies Favoring Small Schools*

Studies that reveal an inverse relationship between school size and level of pupil achievement are almost invariably well designed, carefully conducted, and correct in their inferences. They are, it is true, infected by a flaw that is endemic in research on rural education: demographic observor bias—which is explained below. However, this flaw does not attenuate the findings. Indeed, as you will see later, its correction reveals even stronger association between school size and pupil progress.

*Studies Favoring Large Schools*

Several research reports suggest that pupils learn faster in the large schools. A few of these are disingenuous; some are fraught with invincible error. The remainder can be dismissed easily on the application of Maier's Law [15].

*Studies Finding No Difference*

The overwhelming majority of published work on the general question concludes that there is no demonstrable relationship between school size and pupil achievement. Given what we know to be the truth of the matter, such results must be puzzling and confusing to graduate students and assistant professors [16]. These "no difference" studies, however and fortunately, are plagued by the following errors, oversights, and misunderstandings. Remedied, they provide more acceptable conclusions.

*Researcher Demographic Bias.* I find that, with only a few exceptions, those who have conducted and reported research on the relative effectiveness of large and small schools are or have been urban dwellers or are or have been employed by institutions located in or near an urban area [17]. Because of this urban enculturation, these investigators understandably may be biased against rural schools [18], however unconscious the biases may be. The inference from this previously unpublished fact is easily drawn: The achievement level of rural school pupils tends to be underestimated, and achievement of the favored urban pupils is overestimated. This bias, of course, results in a spurious difference between observations of the two groups, with the pupils in small schools on the losing side. A recently unearthed formula [19] corrects for demographic bias in urban-lovers. I took the liberty of applying it to the "no difference" studies and found that most of the corrected findings are in favor of the small schools, with differences of welcome size and friendly statistical significance.

*Naive Use of Two-Tailed Tests.* A very small number of our "no difference" reporters—contrary to the general urban bias—imply a personal hope that the small school would emerge victorious in their studies. With only tolerable feelings of guilt, they could have followed modal practice and stated these hopes as formal directional hypotheses, thereby justifying one-tailed tests and producing more significant differences. It is estimated [20] that most of the otherwise ambiguous published studies will demonstrate the pleasures of smallness if the rejection region in only one tail of the sampling distribution is used to evaluate the null hypothesis.

*Narrow Interpretation of the Null Hypothesis.* Research in this domain, like that in the social sciences and the professions generally, suggests only limited understanding of the meaning of the failure to reject the null hypothesis. The "no difference" studies under review here conclude that "the difference between large and small schools is not significant"—or some congener of that language informing us that, at the stipulated level of assurance, the sample difference is not inconsistent with a zero population difference. Fair enough. Why not, however, use different language? For example: "the large schools are not significantly better than the small schools." The latter language is denotatively as correct as the former, but the *connotation* is more reinforcing to those who know what the conclusion should be.

*Anal Choice of Fiduciary Levels.* Generations of students have learned that significance levels for rejecting the null hypothesis must be set initially and, for reasons enscribed in tablets yet unearthed, must be no larger than .05. How much knowledge has been tossed away by studies in which differences and coefficients were significant only at the "oh shucks!" [21] level? It has been found [22] that the "no difference" reports not covered under the categories used above clearly favor small schools when larger rejection regions are employed—and dramatically so when we also substitute one-tailed tests for the unproductive alternative. Verily, the bounty of creative inference is rich and knows no end.

We have seen that previous research, corrected and properly interpreted [23], clearly demonstrates that the small educational unit is superior, at least against the achievement criterion. *Quod erat demonstratum*— or so say most of us. Those who remain unconvinced may be compelled by two heretofore unreported studies. These investigations, very briefly summarized below, were designed and executed by me in collaboration with K. Busciardo [24], my assistant, who needs and welcomes most of the credit and to whom all queries should be directed.

## New Data: On the Cutting Edge

*Evaluations by School Superintendents*

We identified a very large random sample of all American public school superintendents with the

assistance of the American Association of School Administrators and Rotary International. Each superintendent received a postcard containing the following directions:

> The more thoughtful and competent American educators conclude that the regression of pupil achievement on school size is linear, homoscedastic, and inverse. If you *disagree,* please sign this card, place it in an envelope, affix a stamp, and address it to us. No response is necessary if you agree.

The non-response rate was 96%! – an emphatic agreement that small is best. That percent, tested by *Chi-cube* [25], is significantly different from zero at awesomely less than a ridiculously small probability. This finding cannot be ignored; the most important professionals in public schools judge that children learn best in smaller schools.

So much for well-informed opinion. Now for an experimental demonstration.

*Experimental Study of Achievement Differences*

We drew a random sample of modest size from a list of all public school principals, with the assistance of AASA and Kiwanis. The chosen individuals were asked if they would cooperate in a study of school size and pupil achievement. Two-thirds agreed (the other principal was in the midst of a grievance proceeding and begged off). The New Universal Test Schedule, a test of basic subjects, was administered to all upper-elementary pupils in the cooperating schools. The distribution of NUTS scores for all pupils was perfectly symmetrical – a marked skewness can be ignored because it occurs only on one side of the distribution. Scores were divided into two groups: those from larger and those from smaller schools. No attempt was made to control for school or pupil characteristics because we wished the study to reflect "the real world." Our first computations produced a disappointing, untoward result: The mean for the larger schools was higher than that for the smaller units. We then eagerly made four modifications: (a) The scores were corrected for Demographic Observor Bias; (b) very high (i.e., least reliable) scores were deleted – these, of course were usually from large schools; (c) the mean for large schools was reduced slightly, a modification that has unimpeachable authority [26]; and (d) scores were carefully adjusted by the "Trimming" procedure [27]. After these modifications, we found that the mean score for the smaller schools was gratifyingly and significantly higher than that for the larger schools – validating Coase's inquisitorial but insightful suggestion that "if you torture data long enough, Nature will confess" [28].

## A Summation, Quickly

In view of the foregoing analyses, can there be much question about what the proper conclusion should be? Other than home, there is no place like the small school. "*Roma locuta, causa finita*" – as the familiar expression goes.

## A *Post Scriptum,* Serendipitously

Plot the means from our experimental study on a figure in which the vertical axis is school size (from zero upwards), and the horizontal axis is pupil achievement (increasing from left to right). Observe that a line connecting the two means slopes from upper-left to lower-right (graphing the negative relationship we noted above). With Euclid's permission to extend a straight line infinitely, take it to the horizontal axis. What do you find at the intercept? Eureka! *Maximum pupil achievement occurs at zero school size.* Nonsense? Not at all. Although liars figure, figures do not lie. Furthermore, the possibility of zero size is acknowledged in all research in which school size is a variable. The lowest category of size invariably uses zero as its lower bound. See Pladson and Lewin [29], for example, whose lowest category is "an enrollment of 0-100 pupils," tacitly accepting the theoretical existence of zero-size, and presumably prepared to deal with it should it occur. The realization that pupils learn best when there aren't any is momentous. Consider, as one example among many, the salubrious effect of zero-size on the costs of operating schools.

I close abruptly, for fear of succumbing to the temptation to point out that the line need not merely stop at the horizontal axis; it may extend below that point, confronting us with the possible reality of *negative* school size and its interesting probable correlates.

## Notes and References

1. The title reflects the recent finding that the title of a competent scholarly report usually is divided by a colon, with the post-colonic clause longer than the pre-colonic clause. (Dillon, J.T. The emergence of the colon: an empirical correlate of scholarship. *The American Psychologist,* 1981, *8,* 879-884.) In addition to the title colonicity, my presentation contains the usual hallmarks of good scholarship: an occasional use of foreign language and long words, a sprinkling of arcane and recondite terms (like "arcane" and "recondite"), and a high ratio of notes and references to paragraphs of exposition.

2. del Gomito, Nervo. Questo, in grand parte, non è vero. *Roma Rurale,* CCCXXIX, *XC,* MCCC-MCCCV.

3. Conant, J.B. *The American high school today.* New York: McGraw-Hill, 1959.

4. Coleman, James, *et al. Equality of opportunity.* Washington, D.C.: The Government Printing Office, 1966.

5. Dunne, Faith. Choosing smallness: An examination of the small school experience in rural America. In Sher, J. P. (Ed.). *Education in rural America: A reassessment of conventional wisdom.* Boulder, Colorado: The Westview Press, 1977. Pp. 81-124.

6. Sher, J.P. (Ed.). *Op. cit.*

7. Gump, Paul, and Barker, Roger. *Big School, small School.* Stanford, California: Stanford University Press,

1964.

8. Matthews, Walter, M. Rural education. *Encyclopedia of educational research.* New York: The Free Press, 1982. Pp. 1627-1635.

9. Morgan, Chris, and Langford, David. *Facts and fallacies.* New York: St. Martin's Press, 1981.

10. Browne, Sir Thomas. *Pseudodoxia epiemica.* Volume 1. London, 1646.

11. Morgan, Chris, and Langford, David. *Op. cit.*

12. Wright, Laurence. *Clear and decent: The unruffled history of the bathroom and the w.c.* New York: Viking Press, 1960. P. 24.

13. Feyerbend, Paul. *Against method.* London: Verso, 1975.

14. Broad, William, and Wade, Nicholas. *Betrayers of the truth.* New York: Simon and Shuster, 1982.

15. "If facts do not conform to the theory, they must be disposed of." (Maier, N.R.F. Maier's law. *The American Psychologist.* 1960, *15,* 208-212.) Maier gives instructions on how this helpful surgery is performed with little or no risk to the reputation of the surgeon.

16. No denigration intended here. The specification merely acknowledges the primary and perhaps exclusive readers of research literature.

17. "Urban" is defined as "non-rural."

18. "Rural" is defined as "non-urban."

19. Menteur, J. Suis. Galimatias. *Nulle Part.* (Unpublished ms.) The correction formula, too involved for explanation here, is a serendipitous by-product of the discovery that the median of a distribution can be located by finding the point that divides it in half.

20. The use of the passive voice here, as in common practice, makes it unnecessary to identify the actor.

21. I regret that I cannot take credit for "oh shucks!" The invention is George Kelly's. He, equally honest, confesses that it is an expurgated version of a fiducial limit used by R.C. Jones in a conversation. (Kelly, George A. The theory and technique of assessment. *Annual Review of Psychology,* 1958, *9,* 323-227. P. 325.)

22. See note 20, above.

23. The sins are more venial than mortal, I hasten to note. There is no culpable ignorance and no room for scorn.

24. Busciardo expired as this went to press, leaving no forwarding address. The editor of this journal, who has highly placed contacts in both locations, is making inquiries.

25. The better known Chi-square distorts information because negative values are lost in the squaring operation. Chi-cube is honest. It's computation is quite straightforward: find Chi-square, extract its square root, and cube the result. To interpret this statistic, use any sampling distribution at hand; it does not have one of its own, preferring the parasitic life.

26. Isaiah, 5.15: "and the mean . . . shall be brought down . . ." Those who prefer secular authority can cite some distinguished scholars who have used the broader Scholar's Sanction, with at least short-term success: "Facts needed to support a preferred conclusion may be created."

27. "Trimming consists of clipping off little bits here and there from those observations which differ most in excess from the mean, and in sticking them on to those which are too small." Babbage, Charles. *Reflections on the decline of science in England.* New York: Augustus M. Kelley, 1970.

28. Quoted in: Leamer, Edward E. Let's take the con out of econometrics. *The American Economic Review,* 1983, *73,* 31-43. P. 37.

29. Pladson, Janet L., and Lewin, Donald K. Elementary principals and their school districts in three categories of ruralness. *Research in Rural Education,* 1982, *1,* 11-14.