

# EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal

Editor: Gene V Glass

College of Education

Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **Education Policy Analysis Archives**. EPAA is a project of the Education Policy Studies Laboratory. Articles are indexed in the Directory of Open Access Journals ([www.doaj.org](http://www.doaj.org)).

Volume 12 Number 52

September 24, 2004

ISSN 1068-2341

---

## School Size and the Influence of Socioeconomic Status on Student Achievement: Confronting the Threat of Size Bias in National Data Sets<sup>1</sup>

Craig B. Howley  
Ohio University  
ACCLAIM Research Initiative

Aimee A. Howley  
Ohio University

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

### Abstract

Most of the recent literature on the achievement effects of school size has examined school and district performance. These studies have demonstrated substantial benefits of smaller school and district size in impoverished settings. To date, however, no work has adequately examined the relationship of size and socioeconomic status (SES) with students as the unit of analysis. One study, however, came close (Lee &

---

<sup>1</sup> This research was partially supported by grants from the National Science Foundation (Grant No. 0119679) and the Rural School and Community Trust. Any opinions, findings, and conclusions or recommendations expressed in this report are exclusively those of the authors and as such do not necessarily reflect the views of the National Science Foundation or the Rural School and Community Trust.

Smith, 1997), but failed to adjust its analyses or conclusions to the substantial bias toward larger schools evident in the data set used. The present study, based on the same large data set, but with size issues in the rural circumstance clearly in focus, reaches rather different conclusions, extending previous work for the first time to a more adequate examination of size effects on individual students. Findings challenge assertions about *ideal* and *minimum* size. Analyses include comparison of means and multi-level modeling. Methodologically, the study illustrates the challenge of using nationally representative data sets of students to investigate second-level contextual phenomena, such as school size. When aggregated to schools attended by nationally representative students, the result cannot be a nationally representative set of schools. Adjustment with weights to simulate such a distribution, moreover, is inadequate to overcome this threat if one is interested in investigating size relationships among the smaller half of US schools, as one must be in seeking to generalize results to the nation as a whole. The present study finds that the smallest national decile of size maximizes the achievement of the poorest quartile of students. Moreover, *appropriate* size is shown to vary by student socioeconomic status.

Practical and scholarly interest in the relationship of school size to student achievement has one of the longest histories of any issue related to schooling, but the record of research accomplishment over that span is not impressive. Stemnock's review of the literature (Stemnock, 1974) traces interest to the turn of the 20<sup>th</sup> century, but even as recently as 30 years ago, when that review was written, most educational researchers held that larger size was preferable.

Since 1974, however, theoretical and empirical perspectives on size have gradually altered, so that scholarly judgment now generally favors smaller size.<sup>2</sup> In the current view, as well, an impressive array of benefits is often claimed for smaller size, just as an impressive array was once claimed for larger size (cf. Cotton, 1996). Much, however, remains to be learned about the influence of size, and a wide range of issues remains to be addressed.<sup>3</sup> Many of the claims currently made are difficult to warrant empirically. The study reported here addresses one of the many remaining issues.

The *urban* "small schools movement," moreover, comes inscribed with a variety of pedagogical and curricular reformist practices, and the upshot is that these pedagogical and curricular options have often been confounded, even in evaluative and research work, with the structural influence of size (e.g., Darling-Hammond, Aness, & Ort, 2002; Fine & Somerville,

---

<sup>2</sup> In fact, however, average school size is not declining in the US, and schools in rural and suburban areas probably continue to grow in size (DeYoung, Howley, & Theobald, 1995) due to consolidation efforts and the failure of policy makers to attend to rural concerns.

<sup>3</sup> The best evidence concerns participation in co-curricular activities; dropout rates; and achievement (level, growth, equity, and achievement cost-efficiency), with a strong preponderance of evidence favoring small schools.

1998; Lee & Smith 1995). Before turning to details about the present study, readers should note well the divide that exists between urban and rural concerns over size. Much of the current prescriptive literature on “small schools” (we prefer the comparative form, *smaller schools*) promotes administratively difficult reform initiatives to personalize huge city schools by creating “schools-within-schools” (AIR and SRI, 2003). These “small schools” are often—perhaps usually—more simulation than reality, because within-school “schools” are seldom free to behave like the (real) autonomous schools that have been studied by school size researchers (e.g., Raywid & Schmerler, 2003).<sup>4</sup> Some researchers believe that size itself is merely a somewhat tangential framework for practices, and it is practices, not structures per se, that produce results.

In rural areas, by contrast with the reformist intentions of urban “small schools,” the practical work centers on difficult efforts to forestall, in the various states, policies that structure wholesale closure and consolidation of *actually* smaller schools in which pedagogical and curricular changes are in no way tied to the issue of size per se.

Some leading urban educators nonetheless have a good grasp of the differences and the commonalities of the struggles in both rural and urban settings (e.g., Klonsky, 1995; Raywid, 1999). In general, these commonalities concern the way educational systems serve students from impoverished backgrounds. More typically, however, urban educators who strive to turn a big-city disaster of 3,000 students into four units of 750 students find it difficult to entertain the possibility that such a seemingly desirable (not to say *ideal*) size might be too large in a rural community. This inference is understandable, and understandably attributable to ignorance about rural places. If, however, researchers responsive to the urban challenge over-generalize their findings to valorize one particular reform package (i.e., smaller schools with constructivist pedagogies), they risk doing damage in rural places.

## Problem Statement

Most of the recent literature on school and district size rests on analyses using the school and the district as units of analysis.<sup>5</sup> In contemporary circumstances, the choice is a

---

<sup>4</sup> The SWAS phenomenon perhaps rests on the dubious assumption that small size harbors effective practices that can be used anywhere, even in large schools. This probable assumption likely emerges from the legacy of the school effectiveness movement, the aim of which was to identify the practices of effective schools and propagate them in other, not-so-effective schools. Some of that literature, indeed, identified size as detrimental to achievement. If one is implementing a range of reforms in a large school, and smaller size is among the recommendations, the natural (if illogical) response is some administrative arrangement to simulate smaller size. The dilemma is that smaller size actually requires the disestablishment of larger schools, and vested interests being what they are, the requirement is received as far too “radical.” In other words, resistance is very strong, and this resistance is probably as strong in rural as in big-city settings. The rural-urban difference is merely that mega-schools are less common in rural places than in big cities and their suburbs.

<sup>5</sup> Lee and Smith (1997) is one of the few exceptions, but it has never been replicated. As will be shown, moreover, applying the findings of this technically and conceptually fine study to policy decisions in the states is problematic, especially in the case of rural schools and districts.

logical one because test scores aggregated to schools and districts commonly serve as grist for the accountability ratings dispensed by the states,<sup>6</sup> and these studies tie the research conveniently to accountability issues, since they assess school and district performance, the focus of nearly all accountability judgments. These studies most commonly show that the direct relationship of size and achievement is zero or modestly negative, and that an interaction between size and SES mediates the direct relationship. This mediation often changes the negative relationship of size and achievement into a positive one for schools and districts exhibiting the highest levels of SES. In general, however, studies have not advocated any size as universally best (with one exception, Lee & Smith, 1997), and the poorer the community, the smaller should schools and districts be; see Howley (2002) for a thorough review of the literature relating size, SES, and achievement.

The findings about interaction effects are observable across 12 replications or near-replications and are comparatively robust. Nonetheless, no work to our knowledge has yet adequately examined the relationship of size and SES with students as the unit of analysis. The most prominent individual-level analysis is the aforementioned study by Lee and Smith. This study will be considered in detail in our review of the literature because its conclusions are flawed with respect to its own findings about smaller schools. In any case, none of the previously published work, even that conducted with students as the unit of analysis, including the first author's own previous work with individual students (Huang & Howley, 1993), represents an *adequate* investigation of SES as a mediator of the influence of size on the achievement of individual students.

Part of the difficulty of studying this issue is the challenge of securing appropriate data sets. Most data sets available from State Education Agencies represent schools and districts and not students. This situation is inevitable because confidentiality of individual students must be carefully protected. In the rare cases in which researchers can access achievement information about individual students within a state, however, individual-level SES measures are most often absent, and when present are likely to be dichotomous (eligibility for subsidized meals). The principal alternative for such study is one of the large, nationally representative databases on students. These data sets, as we will suggest, present rather predictable challenges to school size researchers concerned to study the interaction effects of size in rural places.

### Related Literature: Overview

For most of the twentieth century, small size was considered to be an impediment to school improvement, as suggested above. Beginning with the rise of the railroads as large

---

<sup>6</sup> Huang and Howley's 1993 study is an exception in this line of research. That study found that an interaction effect of size persisted in regression analysis after the introduction of several blocks of control variables related to student academic background, school climate, and school resources. The unique feature of this study was that it examined students in Alaska (the universe of students tested at the applicable grade levels) who had attended the same school for four consecutive years. This selection rule is arguably one that allows for a better test of the hypothesis. The weakness of the study is that it did not employ a multi-level or equivalent ANCOVA model, arguably deflating standard errors. The large number of control variables, as well, seems problematic as well in hindsight.

firms, the organization, per se, of an enterprise began to be understood as enhancing or detracting from productivity. Such efficiencies, the thinking ran then (as it still does), could be realized most immediately as economies of scale: the larger the firm (or school, or school district), the more opportunities existed for efficient organization and probable improvement in productivity (specialization, volume purchasing, supervision and staffing, and so forth).

School and district consolidations were thus linch-pins of the school improvement agenda of the first three-quarters of the 20<sup>th</sup> century. In rural areas professional educators believed that school reform was a literal hostage of consolidation (Cubberley, 1915/1922)—and that improvement was highly improbable or altogether impossible without it. For instance, writing in the early 1950s, Julian Butterworth and Howard Dawson asserted:

School consolidations and school district reorganizations, being made necessary and desirable by changes in agricultural production, new means of transportation, and new needs for education, have resulted in decreases in the number of schools and school districts and increases in the size of both. To the new rural school has come a new type of administrative leadership which has stimulated the improvement of the educational program and at the same time has been partly the product of the new conditions. (Butterworth & Dawson, 1952, p. 3)

Even at mid-century, however, nearly half of rural schools were one- or two-teacher organizations and not much *organization* for the purpose of efficiency could be accomplished with them, at least in the view of the rising class of professional educational managers.

Although consensus about consolidation was strong by mid-century, differences of opinion still prevailed, then as now, and research results also varied. In particular, the issue of what size might be appropriate had been variously determined in opinion and research. Ellwood Cubberley (1915/1922), working at Columbia Teachers College in New York City but writing about rural schools, argued, on one hand, for sizable graded schools and even larger high schools. Joseph Kennedy (1915), on the other hand, dean of education at the University of North Dakota, endorsed schools that were very small, then and now, by urban standards.<sup>7</sup>

As actual research began to accumulate in the field of education, however, this difference of perspective about appropriate size persisted. Stemnock (1974) summarized the results of nearly 120 studies conducted between about 1924 and 1974 for their recommendations about optimal school size. Most of the studies that focused on costs or curriculum (inputs) called for increases in school size. For example, at the high school level, Stemnock reported nine studies carried out between 1924 and 1974 that focused only on costs. Six specified an optimal high school size or size range, averaging about 1,100 students. Six that examined course offerings advised an optimal school size of about 1,200. These figures, in fact, became the mid- to late-20<sup>th</sup> century norm.

Eight studies, however, examined student achievement. Six found no difference in the achievement of students from small and large senior high schools, and only one recommended

---

<sup>7</sup> Kennedy wrote that a school with 25 children actively engaged by their teacher exhibited a fully adequate size.

an average high school size: 400-800 students, or about half the size recommended by studies focusing on input variables (see Stemnock, 1974).<sup>8</sup>

Howley (1996) described three episodes of size research since 1974: (1) studies confirming no significant differences, coming historically first and perhaps concluding around the time of Stemnock's review; (2) studies finding modest achievement benefits of smaller size, with such findings arguably emerging contingent on 50 years of school and district consolidation; and, most recently, (3) studies testing the hypothesis that the relationship of size and achievement is mediated

by SES. Since 1996, there have been at least 12 state-level investigations<sup>9</sup> of the interaction hypothesis (i.e., replicating and extending Friedkin and Necochea's 1988 analysis of California school- and district-level data). Most of these studies show effect sizes, under some conditions quite large, favoring smaller schools and smaller districts in impoverished communities, with effect sizes increasing to the extremes and with net negative effects over much of the SES range. They also show, more substantially and more consistently from state to state, that the relationship between SES and achievement is substantially weaker among smaller schools and districts as compared to larger schools and districts (again, this is a relationship among schools and districts, not among students).<sup>10</sup>

The interaction hypothesis has not received much attention from researchers studying student-level data. There are difficulties confronting those interested in such work. A data set carefully constructed to be nationally representative of students will not necessarily be representative of the size of schools. If smaller schools, for example, are defined as those enrolling less than the median enrollment, contemporary students will typically attend larger schools.<sup>11</sup> If the research focus is context but the data set was not constructed to be

---

<sup>8</sup> Lee and Smith (1997) commend a quite similar range (600-900 students). There are many unacknowledged conceptual and practical difficulties in the attempt to specify "one-best" solutions, not the least of which is the likelihood that the good is contingent, limited, contestable, and ultimately political, as Socrates discovered.

<sup>9</sup> Abbot, Joireman, & Stroh (2002); Alspaugh & Gao (2003); Bickel (1999a); Bickel (1999b); Bickel & Howley (2000); Bickel, Howley, Williams & Glascock (2001); FoutsFriedkin & Necochea, (1988); Howley (1995); Howley, (1996a); Howley (1996b); Howley (1999a); Howley (1999b); Howley & Bickel (1999); Huang & Howley (1993); Johnson (2004); Johnson, Howley & Howley (2002); Miley & Associates (2003). The states involved, and the number of works for each, are: Alaska (1); Arkansas (1); California (1); Georgia (2); Missouri (1); Montana (1); Nebraska (1); Ohio (1); South Carolina (1); Texas (3); Washington (1); West Virginia (3).

<sup>10</sup> Among these studies, the strength of the net size influence (the sum of direct and interaction effects), varies considerably by state, by grade level, by locale, and by unit of analysis (i.e., school versus district). Results of these replications have been repeatedly summarized in the studies cited in the preceding footnote (full citations appear in the reference list) and will not be detailed here, as it is the overall pattern of these 12 cases that is germane, and the relative absence of student-level analysis.

<sup>11</sup> In fact, if *any* population of schools is divided at the *median* of size, the larger half will always enroll more students than will the smaller half, so long as size is defined by number of students enrolled.

representative of context, or if the context of the average student is erroneously construed to constitute the representativeness of context, the findings and conclusions about the contexts can easily exhibit dubious validity. This threat will be a strong one if (as in the present study) one is interested in “unrepresentative” contexts that are disjunct with national patterns.<sup>12</sup> Analysts of school size who use national data sets representative of students, but not necessarily of contexts, should remain alert to this threat.

### Relevant Literature: The Lee and Smith Study

A commonly cited study in the school size literature examined mathematics and reading gain scores from the 8<sup>th</sup> to the 12<sup>th</sup> grades, using NELS data nationally representative of the cohort of students completing the 12<sup>th</sup> grade (Lee & Smith, 1997). We critique this important study at length because the accessible conclusions misrepresent the findings, and, indeed, because some of the methodological choices limit the generalizability of the findings. We begin by describing the conclusions and the findings.

The conclusions follow, taken verbatim from the study report:

1. High schools should be smaller than many are. (p. 216)
2. High schools can be too small. (p. 217)
3. Ideal size does not vary by the types of students who attend. (p. 217)

Before proceeding to a discussion of the reasons why these conclusions need to be viewed with caution, especially by those concerned with smaller schools in rural places, we summarize the findings (as distinguished from the conclusions) more closely. The findings that follow each reflect one of the figures used by the authors to display HLM results graphically and make them more accessible to a wider audience.<sup>13</sup>

1. The effect of size was apparently curvilinear. (Figure 1)
2. The magnitude of achievement gains from 8<sup>th</sup> to 12<sup>th</sup> grade in both mathematics and reading (all else equal) proved greatest in high schools enrolling 601-900 students. (Figure 2)
3. Achievement equity was maximized in high schools enrolling *fewer* than 300 students (math) and 301-600 (reading). High schools larger than 1,500 substantially increased achievement inequity. (Figure 3)

---

<sup>12</sup> This is a comparatively simple point, but it relates to the researcher’s conception of reality: it is an ontological question of some interest. Do we understand context as general or particular? What sort of sense does the construct “average context for an individual” make? We believe that it does not make much sense because context is indeed particular.

<sup>13</sup> These interpretations of the information in Figures 1-6 vary slightly from those given by the authors (cf. Lee & Smith, 1997), but our interpretations are defensible and conservative readings of the findings.

4. In comparing high-SES schools and low-SES schools, the achievement gap proved narrowest in *mathematics* in high schools enrolling 300-1,200 students. The gap between students in low-SES and high-SES high schools widened outside that broad band. *Interaction effects with SES were not significant for reading.* (Figure 4)
5. In comparing mathematics gains in high-minority (Black and Latino) and low-minority schools, gains were highest in schools enrolling 600-1,200 students. Achievement equity, however, was maximized in schools enrolling 300-600 students. Schools larger than 2,100 students showed both the smallest gains and the greatest inequity. (Figure 5)
6. In comparing reading gains in high- and low-minority schools, gains proved highest in schools enrolling 600-1,200 students. Achievement equity was maximized in schools enrolling fewer than 300 students. Gains and equity were sharply reduced in schools larger than 1,500 students. (Figure 6)

The study's three conclusions communicate these six findings very simply indeed. Both the complex findings and the simple conclusions have, however, proven useful to urban educators struggling to reverse the decades of educational damage arguably imposed by very large schools. Taken together, the findings rather clearly suggest that schools larger than 1,500 or 1,800 students do frightful harm to both educational equity and to the achievement growth of all their students and that schools this large substantially *compound* the threatened life-chances of the impoverished youth enrolled in them. The larger the school, in fact, the worse the damage. This study demonstrates the existence of upper limits of high school size in view of the desiderata of academic achievement.

Although this study is a very fine one, it has some limitations and shortcomings. The ones that concern us are those with policy implications for smaller size generally, and, more particularly, for school size in rural places. We turn next to the critique, which is centered on the three conclusions, as given above.

**High schools should be smaller than many are.** This broad statement exhibits sufficient caution and honors the complex findings. It seems evident from a conservative reading of the findings that schools enrolling more than 1,500 students, or perhaps 1,800, depress academic growth and simultaneously magnify academic inequity. The conclusion aligns nicely with the findings as restated by us. We can, however, use insight from our conservative reading of the findings to interrogate the conclusion further, and on objective terms.

Given the harm done by mega-schools of more than 1,500 students, we can ask how many high schools enrolled more than 1,500 or 1,800 students in 1988 (the year the NELS school-completers began high school). The answer is easily supplied with information available from the Common Core of Data (CCD): 2,062 public schools with 12<sup>th</sup> grade the highest grade enrolled more than 1,500 students in 1988, all but 300 of them in cities and suburban areas. Of schools with more than 1,800 students, 1,264 existed nationally at that time, all but 143 in cities or suburbs.<sup>14</sup> To call such proportions "many" seems an appropriate

---

<sup>14</sup> These 1988 data will predictably closely match the 1992 size distributions, the year many members of the cohort completed 12<sup>th</sup> grade and qualified for inclusion in the sample studies by Lee and Smith.

characterization of the situation in metropolitan areas. Elsewhere, however, mega-schools are not “many,” and one might jump to the conclusion that without a mega-school problem, achievement-related size issues do not exist outside of big cities and their suburbs.

In fact, however, school consolidation is an enduring issue in rural areas, and so adequate representation of the variation in school size there would seem necessary to the validity of any study for which generalization to rural areas is claimed. Unfortunately, the NELS data set does not capture this variation adequately and presents a sharp challenge to school-size researchers, and this brings us to a consideration of the second conclusion.

**Schools can be too small.** The relationship of this assertion to the findings is problematic for an important reason not discussed by Lee and Smith (1997): The schools of neither the study subsample nor the NELS data set as a whole reflect the national distribution of school size.<sup>15</sup> This non-representativeness is a specific threat to the findings, especially as they apply to small schools, and, therefore, as they apply to rural schools (which are smaller than non-rural schools).

We start with a discussion of the distribution of size among base year schools (when the NELS sample was in the 8<sup>th</sup> grade). The discussion is germane because any size bias in the original, base-year sample persisted into subsequent waves of data collection. Lee and Smith addressed this threat by (a) applying weights and using weighted z-scores to derive base-year school-level variables and (b) in the absence of NELS weights for high schools attended by NELS students in the 12<sup>th</sup> grade, developing their own weights.

This procedure addresses the threat, but by no means resolves issues for researchers interested in rural places and in smaller schools. For this segment of the population of schools, weighting still allows 77 smaller schools (those with eighth grade enrollments less than 84, the national median) in the NELS sample (n=77 of 801 cases with non-missing size data) to represent a very large number of smaller schools in the population (n=11,345). By comparison, of course, the upper half of the national size distribution in the NELS base year is represented by 724 schools—nearly ten times as many (cf. Jencks, 1985, for a similar observation related to High School and Beyond).<sup>16</sup> In other words, despite weighting and

---

<sup>15</sup> NELS *does* provide a base year weight (BYADMWT) for schools similar to that provided for students (BYQWT), but use of BYADMWT is recommended only for school-level analyses and not in any analyses, including multi-level analyses, where students are the unit of analysis (e.g., Spencer, Frankel, Ingels, Rasinski, & Tourangeau, 1990). Without some adjustment to the second-level data, however, this methodological choice means that nonrepresentative second-level data enter the analyses, because reality is being represented through the lens of student data (i.e., ontological priority is assigned to students over contexts). In any case, comparison of the unweighted distribution of size among NELS schools with those in the population (via the Common Core of Data) confirms the substantial size bias of the (unweighted) NELS data set. Weighting permits the estimation of national parameters, but cannot be relied on for analyses below the national level or for analyses responsive to issues of smaller size—in particular issues of size that include the smaller half of the universe of US schools.

<sup>16</sup> The bias at the high school level is probably quite similar, though we did not, however, examine it. Table 1 compares the unweighted NELS distribution of base year schools with the coeval national distribution; national data taken from the 1987-1988 Common Core of Data national school file. Christopher Jencks (1985) critiquing estimates of sectoral effects on student achievement estimated

adjustment of mean standard errors for design effects, much more error is embedded in findings, and therefore, in conclusions, about smaller schools than is acknowledged. The National Center for Education Statistics has in fact recommended against using school-level weights for any but school-level analyses (Spencer, Frankel, Ingels, Rasinski, & Tourangeau, 1990).<sup>17</sup> In any case, generalizations about smaller schools based on the findings of this study seem poorly justified, and this difficulty means that neither the related findings nor this conclusion (“schools can be too small”) should be accepted as an adequate characterization of the influence of school size on student achievement in rural areas. The threat posed by the NELS size bias also has implications for claims about “ideal” size.

**Ideal size does not vary by the type of students who attend.** At play in this conclusion are two issues: (1) the construct of “ideal size” and (2) the representation of findings about interaction effects. We begin with the construct of “ideal size.”

The determination of ideal size—nationally optimal size, as used here—embeds a judgment undisclosed in the conclusion (perhaps for the sake of brevity), but also almost entirely undefended in the narrative. The undisclosed judgment is that “ideal” should be determined only with reference to the findings about magnitude of achievement gain from grade 8 to 12, setting the findings about equity aside, and dismissing concern for students who drop out, or are pushed out, of high school. In fact, the equity findings are not explicitly represented in any of the three conclusions. They figure in this one implicitly, but only as a denial.<sup>18</sup> The issue is considered this way in the study report:

Obviously, the aims of effectiveness and equity are not completely parallel. In general terms and considering both outcomes, our results lead us to recommend an enrollment size of between 600 and 900 students as “ideal” [quotation marks in original] for a high school. (Lee & Smith, 1997, p. 217).

This passage is the entire consideration, not an excerpt. The judgment of “ideal” is asserted as authorial privilege and is not defended or explained. The judgment is literally (and surprisingly) unwarranted.

---

from High School and Beyond data noted a similar problem. According to Jencks (1985, p. 131), “The fundamental problem here is sample size” (p. 131) ; ... “The HSB sample may have yielded an inflated point estimate of the sector effect simply by chance” (p. 134). The HS&B base year sample included 894 public schools, 84 Catholic schools, and 27 “other” private schools (Coleman, Hoffer, & Kilgore, 1981, Table A.1.3).

<sup>17</sup> This recommendation, however, seems problematic, in the case of joint concern for context and for individuals within contexts and, like Lee and Smith, we have not followed it—except in the case of our size variable (see the “methods” section of this report for relevant details).

<sup>18</sup> The conclusion references equity by way of denial : it states that the “ideal size”—the one in which achievement gain is greatest—is identical from analysis to analysis, and is the same even in those analyses that examine parity of achievement gain; this conclusion does not, nor does any conclusion, summarize the findings about parity of achievement gain. Rather, it transmits the impression that the study’s equity findings are unimportant, whereas they *are* patently important (they are, after all, *reported*).

What might be the alternatives for determining the issue? There are several. One could *defend* the choice of excellence over equity—or the choice of equity over excellence; the case could be made in each instance. Another alternative would avoid privileging one value over another (equity over excellence, for instance), but recognize a need to represent the findings that encompass both values. One might, in short, make a *satisficing* recommendation (see Simon, 1979, on bounded rationality). As John Gardner (1961) wrote many decades ago, equity and excellence are not only related, they are arguably mutually dependent.

A “satisficing” recommendation in this case might have framed a conclusion to incorporate the findings about equity and excellence in specifying an appropriate size range. In particular, a size range of 300-900 students—or even 100 to 1,000 students—seems a suitably conservative interpretation of the findings (i.e., when both equity and excellence are considered valuable). Even this specification, however, might need to be adjusted for rural areas, given the size bias evident in the NELS sample (again, a threat not eliminated, but perhaps compounded, by weighting). Such a wide range can hardly be characterized as “ideal,” however, simply because it *is* so wide. In effect, this study’s chief virtue is the empirical warrant it provides for an upper limit of high school size—arguably about 1,000 students. The studies by Howley and colleagues suggest that within the wide range estimated from the Lee and Smith findings (say 100-1,000), schools in impoverished or mixed-SES communities should be proportionately much smaller, with the smallest schools sustained in the most impoverished communities.

**Summary critique.** In short, the applicability of this prominent national study to particular states, to rural schools, and certainly to the nation as a whole is problematic for several reasons, some methodological, some interpretive (the framing of conclusions), and some ontological (the nationally representative student does not exist in a nationally representative context). The study has made an important contribution to the size literature, but more in terms of methodology and actual findings than in terms of the conclusions, which, as we have argued, misrepresent the findings, especially in the case of schools outside metropolitan areas.

Furthermore, as an ideal or optimum, a range of 100-1,000 students, or even 300-900 students, seems problematic. A range this wide loses the specificity that seems essential to the concept of *ideal* or *optimal*. With respect to the pursuit of optimal (i.e., ideal) size, we rather agree with David Monk and Emil Haller (e.g., Monk & Haller, 1993) that such notions oversimplify the variability of size effects. Indeed, even if a defensible “optimal” size range is 300-900 students, one would like to know if any rule of thumb exists by which one might interpolate an appropriate size for a given situation. The interaction line of evidence seems to recommend SES as just such a metric.<sup>19</sup>

---

<sup>19</sup> Application of such a metric is subject to the knowledge that it is an estimation, subject to the revision of further study, theory, and evidence. The metric is not an ideal or an optimum, nor, we would caution the defenders of smaller schools, does it constitute “best practice.” It does constitute a basis of better-informed practice than conventional wisdom, unexamined assumptions, or unfounded illusion. State-level studies, moreover, can help those applying such a rule of thumb avoid basing their actions on national norms that cannot reflect circumstances that vary quite markedly from state to state.

Two final observations remain. First, the absence of early school leavers from the cohort of students graduating from high school likely mitigates the size effects reported from any study that excludes them (as does Lee and Smith's study). The state-level studies cited previously found interaction effects to be less strong at the 11<sup>th</sup> or 12<sup>th</sup> grade levels as compared to the 8<sup>th</sup>, 9<sup>th</sup>, or 10<sup>th</sup> grade levels. One might hypothesize that the absence of dropouts from the school- and district-level aggregations explains this rather consistent phenomenon, and that their absence in the subject study (Lee & Smith, 1997) has exhibited itself similarly, possibly contributing to the difficulties of drawing conclusions about the influence of size in the smaller half of public schools. Second, the issue of size is arguably confounded with sector. Smith and Lee (1997) included private school students in their study. Not only are private schools much smaller on average than public schools, but, once again, the smaller schools in the population of public schools are dramatically under-represented in NELS. The presence of smaller private schools in the data set compounds the existing threat from the public school size bias.

## Method

In this study, we ask if the sorts of influences observable when schools and districts are the unit of analysis (where such effects are rather firmly established) remain evident in a nationally representative data set in which students are the unit of analysis. Our use of the NELS data set is predicated on our interest in the issues of size in rural places and contingent on our understanding of the peculiar distribution of size among the NELS schools.

Our perspective on size honors structure over process, as previously intimated. Much work remains to be understood about the structural ramifications of size prior to attempts to *explain* them: such explanations confront an unspecified set of contingencies associated with the diverse contexts in which schools of different size exist. Thus, in this study, we are not keenly interested to explain hypothetical influences of size on process (e.g., achievement growth), and we therefore forego the use of gain scores as dependent variables.

Our principal analyses form a series of increasingly more narrowly focused comparisons of means of groups of students defined by school size and socioeconomic status. While we appreciate the usefulness of multi-level regression, in the present study it constitutes an ancillary method, which we deploy to help situate the findings in a structural perspective on the equity effects of size.

**Data set.** The study employs the restricted-use version of the NELS data base. Statistical analyses were conducted by the first author, who is licensed to access the data set. As noted previously, NELS harbors some shortcomings for addressing size issues salient to the rural context,<sup>20</sup> and our construction of size variables takes this fact into account. We selected all students who were enrolled in public schools during 1988 (n=19,396), the NELS

---

<sup>20</sup> We wish to clarify the point that NELS is an excellent data set; it is an aging data set, quite likely not designed with multi-level analysis in view. In any case, researchers are always under the injunction to exercise care in posing and answering questions with empirical data, and our critique is directed to that challenge, which is faced by all of us with varying degrees of success. The critique, then, is not in any way a critique of the data set or NCES. We are deeply appreciative of the availability of this data set; it is teaching us a lot.

base year, during which these students attended the 8<sup>th</sup> grade. All analyses are cross-sectional analyses of relevant cases (i.e., from the full public-school base year cohort, with the exception of a few cases eliminated for missing values on some variables).

**Table 1**  
**8<sup>th</sup> Grade Enrollment in CCD and NELS Schools**

percentile rank	Data set	
	CCD (g0887) (n=22,811)	NELS (bysc3) (n=816) <sup>a</sup>
minimum	1.00	1.00
10	12.00	62.00
20	24.00	106.80
30	39.00	150.00
40	58.00	184.00
50	84.00	219.00
60	121.00	253.00
70	168.00	300.40
80	221.00	353.00
90	296.00	434.80
maximum	1,361.00	1,327.00

<sup>a</sup> 15 of the 816 cases have missing data for bysc3

**Variables for comparisons of means.** The dependent variables in this study are the four standardized base-year achievement scores in NELS: math (BY2XMSTD), reading (BY2XRSTD), science (BY2XSSTD), and history-citizenship-geography (BY2XHSTD).

We created three independent categorical variables, two for school size and one for SES. First, since we wish to compare larger and smaller schools in the US as a whole, we created a dichotomous size variable based on the national median of 8<sup>th</sup> grade cohort enrollment for the nearly 22,811 schools with nonzero grade 8 enrollments in 1987-88 academic year (i.e., 84 students). Schools enrolling fewer than 84 students are designated “smaller” schools, and those enrolling 84 or more students are deemed “larger.”<sup>21</sup>

Second, we followed the same procedure to create a more nearly continuous size variable appropriate to our purposes. We classified schools by national deciles of 8<sup>th</sup> grade

<sup>21</sup> Because of our appreciation of rural schools, we understand the variability in grade-span configurations exhibited across American schools, even at the high school level. In nearly all our previous studies, we use enrollment by grade level (or average enrollment by grade level—much the same thing) as our metric of size. This choice eliminates the possible confounding effects of grade span configuration on size. Dichotomizing the size variable at the national median parallels decisions in previous analyses with schools and districts as the unit of analysis.

enrollment. We used this variable in our analyses comparing means and in our multi-level analysis.

Third, we classified the 19,062 students with data on both school size (BYSC3) and SES (BYSES) into SES quartiles. Quartiles were determined in analyses weighted by the base-year weight (BYQWT). See Table 2 for the quartile cut-points.

**Table 2**  
**Quartiles of students' socioeconomic status<sup>a</sup>**

Percentile rank	
minimum	- 2.970
25	- 0.769
50	- 0.197
75	+ 0.329
maximum	+ 1.907

<sup>a</sup> n=19,062 (all cases valid on bysc3 and byses); mean = -.20939

For prior achievement both in our comparison of means and in multi-level analysis, we used base year composite prior grades (BYGRADS), which are based on student self-reports. This variable shows quite strong correlations with the dependent variables, averaging approximately  $r = .50$ .

For locale, the study uses the NCES "Johnson codes," which in the iteration newly available in the 1987-88 CCD records employed 7 categories.<sup>22</sup> We created a salient locale variable, rural vs. other, by coding small-town (Johnson code 6) and rural (Johnson code 7) into one category, rural (1), and all else "non-rural" (0).<sup>23</sup> This decision was based on the insight that most small towns exist as trading centers within rural areas. We imported locale codes from the 1987-1988 CCD file. The procedure results in missing data for locale on about 18% of cases. Among the valid cases, 39% of students attended rural and small-town schools versus 61% attending nonrural schools. Table 3 compares locale for the valid NELS cases with added CCD locale data to the distribution in the original CCD source file for all schools with nonzero grade 8 enrollments.<sup>24</sup>

<sup>22</sup> See Johnson (1989). The locales and their associated CCD codes follow: 1 = large city; 2 = mid-size city; 3 = urban fringe of large city; 4 = urban fringe of mid-size city; 5 = large town; 6 = small town; and 7 = rural. Currently the CCD deploys an 8-category scheme. The 8<sup>th</sup> category, newly added, separates metropolitan rural locales (from locales 1-5). Locales 6 and 7 (currently as well as previously) are the non-metropolitan locales. The change acknowledges the existence of rural locales within metropolitan areas. Of the 816 public schools in the NELS data set, 666 were successfully matched with CCD information.

<sup>23</sup> Small town schools are, by definition, located in nonmetropolitan areas.

<sup>24</sup> Although we must be cautious about this conclusion due to the 18% missing cases resulting from the addition of CCD data, it nonetheless appears that rural schools are under-represented in the

**Table 3**  
**School Locale in CCD and NELS Schools**

Locale	Data set			
	CCD (Locale87)		NELS (schlocal)	
	(n=22,811) <sup>a</sup>		(n=816) <sup>b</sup>	
	N	%	N	%
1	1,825	8.1	98	14.7
2	2,196	9.8	115	17.3
3	2,297	10.2	103	15.5
4	1,694	7.6	73	11.0
5	359	1.6	13	2.0
6	5,100	22.7	153	23.0
7	8,966	40.0	111	16.7

<sup>a.</sup> minus 374 missing cases    <sup>b.</sup> minus 150 missing cases

Details about variables used in our multi-level analyses are more fully described in the next section, “methods of analysis,” in that part of the discussion devoted to describing the theoretical multi-level model. Briefly, in addition to our decile measure of size and our measure of prior achievement, the multi-level analyses employ a dichotomous measure of ethnicity, school-level aggregations of specified student-level variables, and the cross-level interaction terms implicit in the theoretical model. Please see that section for details about centering and weighting.

**Methods of analysis.** Using our size and SES variables we can compare the performance of the poorest students (and the most affluent students) in smaller versus larger schools. We can also inspect the pattern of achievement differences between smaller and larger schools across the SES quartiles. All analyses use the student-level weight variable BYQWT, with standard errors adjusted for design effects, as advised by NCES (Curtin, Ingels, Wu, & Heuer, 2000).

First, we report achievement means for students grouped into SES quartiles in smaller versus larger schools (divided at the CCD national median of size) for all four tests: mathematics, reading, science, and history. In addition, to gauge the influence of prior achievement on such means, we compute estimated marginal means with a measure of prior achievement, composite base year grades, as covariate.

Second, we investigate the trends observable in the display of such means via a planned comparison of means (t-tests) for students in the top and bottom SES quartiles in smaller and larger schools, for the entire sample and for students attending rural and small-town schools only.

Third, to provide a finer-grained picture of school size effects among individual students, we investigate the achievement of the lowest SES quartile of students across the 10 national size deciles. We report means in all tests across the decile groups, and follow this

---

(unweighted) NELS data set. This conclusion is at least consistent with the evident (unweighted) size-bias previously discussed (rural schools being markedly smaller than other schools).

report with a comparison of the means for students (lowest-SES-quartile students) in the smallest size decile for which statistics are reportable as compared to the largest decile.

Fourth, and parallel to analyses in previous state-level reports, we examine the relationship between individual-level SES and individual-level achievement in the extreme size deciles. This analysis includes all students attending school in the relevant deciles, not only those in the lowest SES quartile.

Fifth, because these analyses show quite consistent effects of school size with respect to bona fide national norms of size, we use multi-level modeling to situate school size effects within a regression model. The theoretical model is based on the most potent individual-level predictors of achievement (i.e., SES and prior achievement). In OLS regression, these two independent variables account for 35% of the variance in the dependent variable (mathematics achievement for this analysis).<sup>25</sup>

The theoretical model posits 10 independent variables (at both levels and including cross-level interaction terms, but excluding the intercept and the four error terms) as hypothetically efficient predictors of the achievement of students attending the 801 schools for which valid data exists across cases. The theoretical model follows:

$$\begin{aligned} \Gamma_{ij} &= \beta_{0j} + \beta_{1j}\text{SES1} + \beta_{2j}\text{PRIOR1} + (u_{0j} + u_{1j} + u_{2j} + \epsilon_{1j}) \\ \beta_{0j} &= \gamma_{00} + \gamma_{01}\text{SES2} + \gamma_{04}\text{SCHSZ} + \gamma_{03}\text{PRIOR2} + \gamma_{04}\text{WHITE2} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}\text{SCHSZ} + \gamma_{12}\text{WHITE2} + \gamma_{13}\text{SES2} + u_{1j} \\ \beta_{2j} &= \gamma_{20} + \gamma_{21}\text{SCHSZ} + \gamma_{22}\text{SES2} + u_{2j} \end{aligned}$$

In particular, the model theorizes that school-level SES (SES2), school size (SCHSZ), school-level composite prior achievement (PRIOR2) and school-level proportion of nonhispanic white students (WHITE2) will influence the school-level mean of the dependent variable (mathematics achievement), which is given by the expression for the random intercept,  $\beta_{0j}$ . The model further suggests that the school-level gradient of individual-level SES ( $\beta_{1j}$ ) is mediated by school size, school-level proportion of white students, and school-level SES. It suggests that the school-level gradient of individual-level prior achievement ( $\beta_{2j}$ ) is mediated by school size and school-level SES.

The expanded theoretical model, with a maximum of 15 parameters available for estimation is given by the following equation:

$$\begin{aligned} \Gamma_{ij} &= \gamma_{00} + \gamma_{01}\text{SES2} + \gamma_{04}\text{SCHSZ} + \gamma_{03}\text{PRIOR2} + \gamma_{04}\text{WHITE2} \\ &+ \gamma_{10}\text{SES1} + \gamma_{11}\text{SES1}*\text{SCHSZ} + \gamma_{12}\text{WHITE2}*\text{SES1} + \gamma_{13}\text{SES2}*\text{SES1} \\ &+ \gamma_{20}\text{PRIOR1} + \gamma_{21}\text{PRIOR1}*\text{SCHSZ} + \gamma_{22}\text{SES2}*\text{PRIOR1} \\ &+ u_{0j} + u_{1j}*\text{SES1} + u_{2j}*\text{PRIOR1} + \epsilon_{1j} \end{aligned}$$

---

<sup>25</sup> Including other student-level predictors (i.e., ethnicity, gender, homework effort) increases explained variance only a little (raising  $r^2$  proportion to 39%).

School-level variables are weighted z-scores, with the exception of the size variable, which is already based on the national distribution of 8<sup>th</sup> grade cohort enrollment. We adopt this method of centering and school-level weighting following the example of Lee and Smith. Our individual-level variables are also calculated as z-scores and centered on the weighted school values.

Our multi-level analysis was conducted with the SPSS mixed-models module following the method outlined by Bickel (forthcoming). Design effects were estimated using a shareware software package (American Institutes for Research, 2003). Theoretical models must be fitted to the data (Kreft & de Leeuw, 1998), and experience suggests that some of the theoretically salient predictors will not appear in the fitted model. We began the process of fitting the model by computing the null model (predicting only the intercept—the grand mean, and without random coefficients). Next we entered the two student-level predictor variables (SES1 and PRIOR1), but with fixed slopes. We then introduced all four school-level predictors of the random intercept. Next, in sequence, we added school-level predictors of student-level SES, adding one school-level predictor variable at a time; and then followed suit with by fitting the two predictors of prior achievement. At each step we assessed model fit by inspecting the minus-2 log likelihood, allowing for the Kreft and de Leeuw rule of thumb (a reduction in minus-2 log likelihood of at least twice the number of parameters estimated by the equation, see Kreft & de Leeuw, 1998, p. 65). The process of fitting stopped when model fit was maximized and all available independent variables (both levels and cross-level interactions) had been examined.

## Results

Our first analysis examined means from two analyses of variance: one two-way ANOVA of achievement scores in all four subjects tested (smaller or larger school size by SES quartile), and the same analysis, but with composite prior grades the covariate. Table 4 reports the mean achievement (with and without the covariate; estimated marginal means in parentheses). We are interested here only to observe the pattern of estimated marginal means; we do not yet report significance tests for particular comparisons at this step, either for unadjusted or adjusted means. The interaction of size and SES, however, exhibits adequate statistical significance to warrant planned tests of mean differences (varying from  $p < .001$  for mathematics to  $p < .05$  for science and history). Again, all analyses in this study were weighted by the base year weight variable (BYQWT).

What evident patterns do we observe in Table 4? First, the predictable differences in achievement by SES quartile are evident. Each cell in the table gives the achievement mean in a subject for a given SES quartile in smaller schools, larger schools, and as the total for the quartile. In each subject, quartile total scores increase to a similar degree, with the observed difference between lowest and highest SES quartile equivalent to about .65 standard deviations—a substantial effect size.

**Table 4**  
**Achievement means for quartiles of socioeconomic status, smaller versus larger schools**

SES Quartile <sup>a</sup>	Size <sup>b,c</sup>	mathematics		reading		science		history	
		mean <sup>d</sup>	SD						
1	smaller n=902	45.73 (47.29)	8.04	46.08 (47.51)	8.66	46.24 (47.52)	8.83	45.83 (47.16)	9.12
	larger n=3786	44.17 (45.81)	7.21	44.40 (45.90)	7.89	44.67 (46.02)	7.86	44.32 (45.72)	8.37
	total n=4688	44.49 (46.56)	7.42	44.74 (46.70)	8.08	44.99 (46.77)	8.10	44.63 (46.44)	8.55
2	smaller n=826	49.06 (49.38)	8.68	49.11 (49.40)	9.18	49.80 (50.06)	9.55	49.29 (49.56)	9.13
	larger n=3627	47.66 (48.38)	8.82	48.11 (48.77)	9.20	48.01 (48.60)	9.10	47.83 (48.45)	8.98
	total n=4453	47.94 (48.88)	8.81	48.31 (49.09)	9.21	48.37 (49.33)	9.22	48.13 (49.01)	9.03
3	smaller n=568	51.56 (50.59)	9.25	51.36 (50.49)	9.38	52.12 (51.38)	9.56	51.53 (50.71)	9.16
	larger n=3826	50.31 (50.15)	9.50	50.29 (50.14)	9.57	50.52 (50.39)	9.60	50.44 (50.30)	9.26
	total n=4394	50.49 (50.37)	9.48	50.45 (50.31)	9.55	50.75 (50.86)	9.61	50.60 (50.50)	9.26
4	smaller n=395	54.72 (52.65)	9.80	54.35 (52.45)	10.27	54.88 (53.18)	9.99	54.64 (52.89)	10.12
	larger n=3994	55.90 (53.98)	10.31	54.81 (53.05)	9.88	55.02 (53.45)	10.06	54.92 (53.28)	9.85
	total n=4389	55.78 (53.31)	10.27	54.76 (52.75)	9.92	55.01 (53.31)	10.06	54.89 (53.08)	9.88
all	smaller n=2691	49.47 (49.98)	9.34	49.50 (49.96)	9.68	50.01 (50.52)	9.89	49.56 (50.08)	9.78
	larger n=15233	49.76 (49.58)	10.07	49.62 (49.46)	9.95	49.77 (49.61)	9.99	49.60 (49.47)	9.96
	total n=17924	49.71 (49.78)	9.96	49.60 (49.71)	9.91	49.81 (50.07)	9.97	49.59 (49.76)	9.93

Notes.

- BYQWT-weighted quartile of BYSES for 19,062 cases valid on byses (SES) and bysc3 (8<sup>th</sup> grade enrollment)
- BYSC3 divided at the national median of 8<sup>th</sup> grade enrollment for 22,811 schools in the Common Core of Data with nonzero grade 8 enrollment in 1987-1988
- unweighted N across all achievement measures
- estimated marginal means in parentheses (covariate: composite prior grades [BYGRADS]; analyses weighted by base year weight [BYQWT])

Second, in Table 4, we see that there seem to be observable patterns related to differences in size, as portrayed by our categories of smaller and larger. In every subject, with and without control of prior achievement, the observed means for the lowest SES quartile are higher in smaller than in larger schools. This pattern of difference, however, holds true also for the second and third SES quartiles: observed means for smaller schools are higher than for larger schools. Only in the fourth SES quartile does this pattern seem to reverse. The differences in magnitude in the observed means in the fourth quartile, however, are very modest, not even one-tenth the magnitude observed in the first quartile—largest in mathematics at .12 and much smaller in the others (.05 in reading, .03 in science, and smallest in history at .02). Even in a large data set, differences this small can be of doubtful statistical significance. In any case, Table 4 suggests the existence of an interaction effect of size at the level of individual students, and, as noted above, the overall tests of significance support the inference.

We conducted planned tests of significance on some of the observed differences reported in Table 4. We compared the performance of students in the lowest SES quartile enrolled in smaller schools to the performance of other students in the same SES quartile enrolled in larger schools; and we also compared the performance of highest-SES students enrolled in smaller versus larger schools.

Results are presented in Table 5. Again, analyses were weighted with the base-year weight variable and standard errors were adjusted for design effects. As anticipated, in all four subjects the difference in test scores between those lowest-SES-quartile students enrolled in smaller schools and those enrolled in larger schools exhibits strong statistical significance ( $p < .01$  in all cases). Using within-quartile pooled variance to estimate effect size, we assess the practical significance of these differences. Effect sizes are about  $+ .20$  in favor of smaller schools for these students. The situation is different for students in the highest SES quartile, however. As we noted previously, means for larger schools exceed those for smaller schools among students in the highest SES quartile. Table 5 reports the statistical non-significance of all these observed differences in the fourth SES quartile.<sup>26</sup>

Because we are particularly concerned with the implications of school size in rural places, we repeated the previous analysis for students attending rural and small town schools only. Again, we compared lowest and highest SES students attending larger versus smaller schools (split, again, at the national median of 8<sup>th</sup> grade enrollment), and the results are reported in Table 6. The pattern of differences within SES quartiles for smaller as compared to larger schools is similar to that reported in Table 5, but the effect sizes are consistently about 50% larger for lowest SES rural and small-town students attending smaller schools than are those calculated for the sample as a whole (i.e.,  $es \cong +.30$  vs.  $+.20$ ). The differences favoring larger schools in the highest SES quartile, however, are nonsignificant, as in Table 5.

Given the foregoing results, the study was poised to examine the continuity or discontinuity in the evident influence. The poor representation of smaller schools in this data set, however, has rendered such an inquiry somewhat difficult. We addressed this difficulty, as noted previously, through use of a measure of size based on national deciles of 8<sup>th</sup> grade

---

<sup>26</sup> We report an effect size only for mathematics achievement, despite the lack of significance. Just 395 highest-SES public-sector NELS students are enrolled in smaller schools, and the difference in means reported in Table 5 approaches statistical significance. It is possible, though perhaps unlikely, that statistical significance would be reached with a larger sample size.

enrollment. This measure avoids the problem of multiplying error through weighting, but caution is still needed, especially in inspecting the sample sizes of cells within the analyses. Data from cells with fewer than 30 cases should not be reported according to NCES (Curtin, Ingels, Wu, & Heuer, 2002). Although an unknown degree of error lurks in the alternative measure of size, the alternative seems appropriate, given the available options.

**Table 5**  
t-tests of mean achievement differences<sup>a</sup>, students in smaller as compared to larger schools (schools in all locales)

**Panel 1: lowest SES quartile of students<sup>b</sup>**

	mean in smaller	mean in larger	difference	SE <sub>diff</sub> <sup>c</sup>	t-value	P	effect size <sup>d</sup>
mathematics	45.730	44.168	1.562	.46790	3.3828	.00089	+ .21
reading	46.082	44.398	1.684	.42613	3.9526	.00009	+ .21
science	46.239	44.667	1.572	.53435	2.9418	.00338	+ .19
history	45.833	44.320	1.512	.55083	2.7464	.00619	+ .18

**Panel 2: highest SES quartile of students**

	mean in smaller	mean in larger	difference	SE <sub>diff</sub> <sup>c</sup>	t-value	p <sup>e</sup>	effect size <sup>d,f</sup>
mathematics	54.717	55.901	-1.184	.75847	1.5608	.12 (ns)	(-.12) ns
reading	54.346	54.808	-0.461	.66295	0.6967	.49 (ns)	----
science	54.881	55.025	-0.144	.66342	0.2172	.83 (ns)	----
history	54.642	54.923	-0.281	.72866	0.3857	.70 (ns)	----

Notes.

- means unadjusted for covariate (BYGRADS); sample sizes are those reported in Table 4
- sample selected for valid cases on BYGRADS, to yield the same unadjusted parameter estimates as reported in Table 4
- standard errors adjusted for design effects with with AM Version 0.06.01 Beta (released October 27, 2003); software available gratis from the following URL: <http://am.air.org/about2.asp>
- based on within-quartile pooled variance; see Table 4 for the related standard deviations
- see Table 4 for cell sizes; relatively few NELS students from the highest SES quartile ( $n = 395$ ) attended smaller schools (see note); with a larger sample size attending smaller schools, the mathematics difference might have proven statistically significant (the other observed nonsignificant differences are quite small and would not likely be rendered significant with larger sample size)
- the effect size associated with the observed non-significant values (see item *e*) is reported parenthetically for mathematics because marginal significance may be an artifact of small sample size

**Table 6**  
**t-tests of mean achievement differences, students in smaller as compared to larger schools (rural and small-town schools only)**

**Panel 1: lowest SES quartile of students<sup>a</sup>**

	mean in smaller	mean in larger	SD (pooled)	diff.	SE <sub>diff</sub> <sup>b</sup>	t-value	p	effect size <sup>c</sup>
mathematics	46.438	44.253	7.443	2.184	0.62771	3.4793	.00060	+ .29
reading	46.833	44.639	8.278	2.194	0.57491	3.8163	.00017	+ .27
science	47.600	45.058	8.305	2.542	0.62983	3.6404	.00033	+ .31
history	47.069	44.372	8.633	2.697	0.69011	3.9084	.00012	+ .31

**Panel 2: highest SES quartile of students<sup>e,d</sup>**

	mean in smaller	mean in larger	difference	SE <sub>diff</sub> <sup>b</sup>	t-value	p	effect size
mathematics	55.083	55.942	-0.859	1.0189	0.8432	.40 (ns)	----
reading	55.229	55.246	-.017	0.8908	0.0196	.98 (ns)	----
science	55.801	55.895	-0.094	0.7538	0.1247	.90 (ns)	----
history	55.030	55.332	-0.302	0.8955	0.3376	.74 (ns)	----

Notes.

- unweighted n: 462 (students attending smaller schools); 1,247 (students attending larger schools)
- standard errors adjusted for design effects with AM Version 0.06.01 Beta
- unweighted n: 213 (in smaller schools), 1,037 (in larger schools)
- based on within-quartile pooled variance (SD pooled); pooled-variance standard deviations not reported in panel because calculation of effect sizes is not warranted based on nonsignificance of difference scores.

Our particular interest for this analysis is the achievement of students in the lowest SES quartile. We computed achievement means for the four tests for these students in each size decile. Table 7 reports the results. In addition to the means for the four achievement measures, Table 7 reports the 8<sup>th</sup> grade enrollment band defining each decile, and the number of cases in each cell. The Table notes provide further information on the number of schools and of states represented in each size decile. Curtin and colleagues' suppression rule (Curtin et al., 2000) applies only to the means for the first decile. Deciles 2 through 10, however, exhibit adequate cell sizes and we report the achievement means for those cells.

The data in Table 7 are *suggestive* of a linear relationship across the deciles in all subjects (see multi-level results for unique estimates of the influence of school size and the interaction of size and SES based on this size variable). The lowest reportable decile exhibits the highest means, and the tenth decile exhibits the lowest means, across all achievement measures. Scores, moreover, tend to decline from one cell to the next as one ascends from lower to higher deciles. One might, indeed, characterize the center (deciles 4 through 8) as "lumpy," and, of course, the difference between adjacent cells is not likely to reach statistical significance in many cases. Sampling artifacts related to the NELS size bias may be responsible for the lumpiness, or size effects in the population might be nonlinear, or some other condition might be responsible. One cannot judge definitively, but Table 7 would seem in any case to put the

lower threshold of size for such students quite low indeed. The threshold might even be lower than the 12-23 enrollment band (8<sup>th</sup> grade cohort enrollment) that defines the second decile. This judgment, too, cannot be given because the NELS data do not permit an adequate test. (The note to Table 7 gives readers a sense of the various states represented by students in the analysis. The varied sample sizes and other sampling issues, we believe, make formal tests of significance and derivation of effect sizes somewhat problematic here. Nonetheless, conservative estimates of effect sizes for observed differences can be derived by using the NELS tests' theoretical standard deviation of 10 as the denominator for observed differences.)

**Table 7**  
**Mean achievement for students in lowest**  
**SES quartile by national deciles of school size**

Size decile <sup>a</sup>	8 <sup>th</sup> grade enrollment	listwise n (unweighted)	Math	Reading	Science	History
1	<12	10 <sup>b</sup>	-----	-----	-----	-----
2	12-23	63 <sup>c</sup>	47.871	47.315	48.998	47.402
3	24-38	206 <sup>d</sup>	46.078	46.032	46.591	45.797
4	39-57	297 <sup>e</sup>	44.870	45.717	45.807	46.122
5	58-83	320 <sup>f</sup>	45.682	46.049	45.581	44.956
6	84-120	421	45.124	45.882	45.739	45.822
7	121-167	613	43.474	44.269	44.003	43.901
8	168-220	719	44.341	44.320	45.226	44.596
9	221-295	714	44.261	43.961	44.473	44.057
10	296+	1416 <sup>g</sup>	43.731	43.949	44.064	43.719

Notes.

- a. based on 1987-1988 CCD national deciles for 22,811 schools with nonzero 8<sup>th</sup> grade enrollment; analyses weighted by BYQWT
- b. unweighted n < 30 in all cells; values not reported, per instructions of NELS base year student manual; the 10 students in this cell attended 3 schools, each in a different state
- c. these 63 students attended 14 schools in 12 states (i.e., AK, AR, IA, ID, IL, KS, MN, MO, ND, NE, OK, TN)
- d. these 206 attended 21 schools in 17 states (i.e., AL, AR, AZ, CA, IL, KS, LA, MI, MO, MS, NJ, ND, NM, NY, OK, TN, WI)
- e. these 297 students attended 31 schools in 21 states (i.e., AL, AR, CA, IA, IL, IN, KY, LA, ME, MI, MO, ND, NH, NJ, NY, OH, OR, TN, TX, WA, WI)
- f. these 320 students attended 44 schools in 25 states (i.e., AL, AZ, CA, CO, FL, IA, IL, KY, LA, MA, ME, MI, MN, MO, NC, NJ, NY, OH, OK, OR, PA, SC, TN, TX, WI)
- g. these 1416 students attended 228 schools in 36 states (i.e., AL, AR, AZ, CA, CT, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, MI, MO, MT, NC, NE, NJ, NM, NV, NY, OH, OK, PA, SC, TN, TX, UT, VA, WA, WI)

The significance and effect sizes of the difference in mean achievement *are* reported, however, in Table 8, for the second and tenth deciles. Again, this comparison is made for students in the lowest quartile of SES. Despite the comparatively small number of students representing the second decile, the t-values have remained large and the effect sizes are substantial, averaging about .50 standard deviations. We also drew a small random sample of students from the 10<sup>th</sup> decile (n=91) to perform the same analysis as reported in Table 8, but

with balanced cells. These unreported results proved nearly identical to those reported in Table 8 and are available from the authors. All differences were statistically significant ( $p < .03$ ).

**Table 8**  
**Achievement differences among students in lowest quartile of SES, in second<sup>a</sup> versus tenth deciles of 8<sup>th</sup> grade enrollment**

size decile <sup>b</sup>	math		reading		science		history	
	2 (12-23)	10 (296+)	2 (12-23)	10 (296+)	2 (12-23)	10 (296+)	2 (12-23)	10 (296+)
means	<b>47.871</b>	<b>43.731</b>	<b>47.315</b>	<b>43.963</b>	<b>48.998</b>	<b>44.096</b>	<b>47.402</b>	<b>43.711</b>
N <sub>unw</sub>	63	1440	63	1432	63	1438	63	1427
SE <sub>mean</sub> <sup>c</sup>	1.150	0.258	0.787	0.333	1.052	0.301	0.724	0.357
SD	9.119	7.276	9.190	8.073	9.999	7.850	10.072	8.598
SD <sub>pooled</sub> <sup>d</sup>	7.400		8.050		8.072		8.548	
diff	4.110		3.352		4.901		3.691	
SE <sub>diff</sub>	1.178		0.854		1.094		0.807	
t-value	3.489		3.923		4.480		4.573	
p	.00058		.00012		.00001		.00001	
effect size	<b>+ .56</b>		<b>+ .42</b>		<b>+ .61</b>		<b>+ .43</b>	

Notes.

- too few cases in lowest size decile ( $n < 30$  all cells) to produce reliable estimates; next lowest quartile used instead
- national deciles of size based on 22,811 schools with nonzero enrollment in grade 8, 1987-1988
- standard errors adjusted for design effects with AM Version 0.06.01 Beta
- standard deviation based on pooled variance of full SES quartile (i.e., across all deciles of size)

Previous investigations of the interaction effects of size have hypothesized that the higher test scores in smaller schools serving impoverished communities may be partly the result of a weakening of the usually strong bond that exists between socioeconomic status and achievement. In those studies with schools and districts, the overall relationship between SES and achievement (i.e., across all SES groups) were reported to be consistently weaker among smaller schools and districts. Those findings were, moreover, very consistent from study to study. Table 9 reports a similar analysis on all the students attending schools in the second, third, and tenth deciles of national size (i.e., from all SES quartiles). The Table provides three contrast groups: the second, the tenth, and the combined second and third deciles.

As in previous studies, Table 9 reports the proportion of variance shared by SES and the various achievement measures. Before computing the shared variance (the squared correlation coefficient), we examined the SES distributions in each decile. The second decile exhibited a narrower range than the tenth; the range among students classified in that decile was modestly restricted at both ends. We therefore constrained the range for all analyses to that exhibited by the second decile.<sup>27</sup>

<sup>27</sup> That is, we imposed the SES maxima and minima evident among the group of smaller schools to the group of larger schools, eliminating some cases. We ran analyses with and without the restriction; the reduction in  $r^2$  among students in the tenth decile was very modest (cf., for

**Table 9**  
**Percentage of variance shared by achievement and**  
**SES<sup>a</sup> for selected deciles of 8<sup>th</sup> grade enrollment**

decile groups <sup>b</sup>	$r^2_{\text{math, SES}}$	$r^2_{\text{reading, SES}}$	$r^2_{\text{science, SES}}$	$r^2_{\text{history, SES}}$
2	7.6%	9.2%	8.4%	9.5%
2+3	12%	11.7%	10.6%	9.9%
10	18.9%	15.2%	16.7%	17.2%

Notes.

- SES range for all deciles constrained to that of the second decile ( $-1.851 < \text{byes} < +1.369$ )
- second decile unweighted  $n = 227$ ; second plus third decile unweighted  $n = 577$ ; tenth decile unweighted  $n = 1440$  (all listwise)

In comparison to the relationships between achievement and SES that prevail among students attending the largest schools, those attending the smallest schools experience a 60% reduction in the influence of SES on mathematics performance, a 39% reduction on reading performance, a 50% reduction for science, and a 45% reduction for history. The results for comparisons with the combined second and third deciles show predictably weaker, but still substantial, reductions (37% in mathematics, 23% in reading, 37% in science, and 42% in history). It is well to remember that in the present study these are *student-level* associations of SES and achievement in schools enrolling 8<sup>th</sup> grade cohorts of 12-23, 24-38, and 296 or more students.

**Multi-level analysis.** The SPSS mixed-models parameter estimates for the fitted equity model are reported in Table 10. The fitted equity model includes the intercept plus 8 predictors (i.e., instead of the 10 in the theoretical model); all predictors of the random intercept proved very highly significant (SES2, PRIOR2, WHITE2, SCHSZ, for all of which  $p < .000$ ). Not surprisingly, prior achievement at the student level and at the school level together exerted the strongest influence on mathematics achievement, in about equal measure, and amounting to about 8.5 standard score units—equivalent to about 85% of the overall individual-level standard deviation if each variable were increased by one standard deviation. This finding is consistent with a century of study on student academic ability,<sup>28</sup> except for the somewhat surprising fact that the multi-level analysis partitions some of the influence of this variable to the context of prior achievement (i.e., PRIOR2) in which students find themselves.

---

mathematics,  $r^2 = 20.8\%$  without the imposed restriction). This procedure cannot disclose the “true” association, but it does suggest that the effects of restricted range are comparatively minor.

<sup>28</sup> Composite prior achievement is a very arguable proxy for IQ. IQ is seldom used any longer in educational research, for many reasons. Here, we wish to observe that the differences between general achievement (as in composite prior grades) and ability and aptitude and intelligence are theoretical. Statistically, they behave in nearly identical ways, for the most part. The construct (whatever one calls it and however much one may object to its misuse as prop to bigotry) retains explanatory power in a statistical sense, and, indeed, other researchers (e.g., Lee & Smith, 1997) have sometimes used a related construct (composite prior test scores).

**Table 10**  
**Fitted multi-level equity model<sup>a</sup>**

	estimates <sup>b</sup>				SE	df	t	p
	model 0	model 1	model 2	model 3				
Intercept	49.334	49.721	50.186	50.186	.097	771.176	517.381	.000
SES1		1.716	1.719	1.433	.094	775.853	15.245	.000
PRIOR1		4.518	4.493	4.362	.069	775.132	67.130	.000
SES2			3.699	3.699	.059	773.966	62.695	.000
WHITE2			0.519	0.519	.061	772.296	8.492	.000
PRIOR2			4.046	4.046	.043	774.853	94.070	.000
SCHSIZE <sup>c</sup>			-0.199	-0.199	.021	771.460	9.476	.000
SIZE X SES1				0.146	.030	775.904	4.867	.000
SES2 X PRIOR1				0.944	.055	776.756	17.164	.000
-2 log likelihood	135792.7	126412.1	124417.6	124261.5				

Dependent variable: BY2XMSTD; intra-class correlation = .237

$R_1^2 = 44.2\%$ <sup>d</sup>

Notes.

- weighed analysis (BYQWT) with all variables centered (individual variables on school means; school variables as weighted school-level z-scores [cf. Lee & Smith, 1997], except for SCHSIZE); standard errors of the estimates computed with AM Version 0.06.01 Beta; t and p calculated by authors from corrected standard errors
- estimates show model improvement from null model (0) to fitted model (3); model 1 without random intercept or random slopes; model 2 with random intercept added; model 3 with random slopes; significance data pertain to model 3
- SCHSIZE gives deciles 2-10 of national school size, centered on decile 6 (decile 1 excluded, see discussion in text) and reflects the same measure of size used for the analyses reported in Table 8 and Table 9
- computed from covariance estimates by authors (cf. Snijders & Bosker, 1999)

Similarly, SES1 and SES2 together exert the second strongest influence on mathematics achievement, except that for SES as compared to PRIOR, the unique magnitude of contextual influence (SES2) is substantially stronger than the individual-level attribute (SES1). Socioeconomic status, unlike prior achievement, operates most strongly as a contextual variable. Together, the two SES measures produce an adjustment to the mean of about 5 standard-score units (i.e., equivalent to half a standard deviation if each independent variable were increased by one standard deviation).

Net of these extremely strong influences (jointly of individuals and contexts), the magnitude of other influences is predictably weaker. In the fitted model, and as theorized, school size directly influences school-level achievement (via its influence on  $\beta_{0j}$ ) but also indirectly mitigates the influence of student socioeconomic status (via its influence on  $\beta_{1j}$ ) on student-level achievement. The interaction takes on positive values (recall that all variables except the dependent variable are centered) for students with SES that is below average in their smaller schools and for students with SES that is above average in their larger schools.

School size does not remain in the model as an influence of the school-level gradient for student-level prior achievement ( $\beta_{2j}$ ), but school-level socioeconomic status (SES2) does. All else equal (a particularly potent idea in multi-level analysis), students with prior achievement that is lower than average in their schools will perform *better* in schools with below-average aggregate SES. This interaction, in other words, tends to weaken the school-level gradient of student-level prior achievement.

## Discussion

Overall, these findings, based on national student-level data, appear consistent with those previously reported in state studies with schools and districts as the units of analysis: (1) smaller school size confers an achievement advantage on all but the highest-SES students, (2) smaller size mediates the powerful association between SES and achievement, (3) the relationship between school size and achievement is predominantly linear, and (4) size effects are at least as robust in rural schools as compared with schools overall.

Three findings also modify and extend previously reported findings. First, contrary to the assertion of Lee and Smith (1997), these results do not disclose any lower limits for school size. Second, contrary to our own previous work, this study suggests that larger size does *not* significantly improve performance among affluent students. Third, the multi-level results show that even in complex relationship with the most potent influences on student achievement, both school size and the interaction of school size and student SES retain a significant influence student achievement.

**Warranted policy options.** Efforts to build and sustain smaller schools are warranted on the basis of these findings. Urban areas, in particular, need new and autonomous smaller schools (rather than non-autonomous ‘schools’-within-schools). This observation applies in some rural areas, as well. More particularly, however, these findings provide further warrant for sustaining the existing small schools so common in rural America.

In contemplating school size, citizens and educators should attend carefully to grade-level cohort size as well as to overall school size.<sup>29</sup> A ninth grade academy with 300 students will likely embed the sorts of structural disadvantages associated with 9-12 schools enrolling 1,200 students.<sup>30</sup> A ninth-grade “academy” with 300 students, is not, in other words, a “smaller school.”

**Unwarranted policy options.** Educators and policy makers sometimes make unwise inferences from research findings. We want to warn them against one such inference in the case of this study. We strongly caution readers not to infer from our findings support for any plan that would allocate smaller schools only to lower-SES students. In fact, the results of the multi-level analyses suggest that a subtle and difficult-to-achieve SES balance is productive of improved equity and “excellence,” not segregation. Intentional isolation of impoverished

---

<sup>29</sup> Findings from this study and from the state-by-state studies of the interaction hypothesis construe size in this way.

<sup>30</sup> Research on school-to-school transitions (e.g., Alspaugh, 1998) provides some evidence, moreover, that narrow gradespan configurations increase the number of school-to-school transitions and thereby affect school performance adversely.

students already living in mixed-SES communities into segregated smaller schools would constitute a particularly insidious form of tracking. Indeed, in some communities such class-based segregation would inevitably segregate students by race. Such de facto policies (whether race- or class-based) already work strongly to subvert the U.S. democracy. They should not be re-made *de jure*. The inference against which we caution would show a dramatic—an *unethical*—lack of circumspection.

Findings from this study obviously offer no support for arrangements that work to increase the size of already small schools, especially those that serve impoverished communities (the common proposal in rural areas). Certainly, rural areas confront the challenges of out-migration, deterioration of school facilities, temporary stress on state revenues, and population increases in rural suburbs and small towns. In light of the findings from this and other studies, concern for achievement and for reducing achievement gaps means that educators and policy makers must search for ways to meet these challenges without closing schools that are already appropriately small. How should policy makers balance competing considerations? We advise it be done in ways that enable impoverished rural communities not only to retain but to improve existing smaller schools. How? Policy makers can change the rules under which state systems operate, from big-school to small-school norms. They might, in other words, un-rig the game that requires schools to be large. This study and others show that large size is not the criterion of “excellence” it was once thought to be. And smaller schools have now been shown to exert an evidently robust effect on equity. It is interesting to observe that at the time large-school norms were instantiated—the early and mid-twentieth century—few educators or legislators worried about equity. Such norms seem to have outlived their utility.

### Recommendations for Further Study

Many participants in the small schools movement, as previously suggested, confound smaller size not only with the opportunity for certain reforms, but also with the reforms themselves. Although we understand this move, we regard it as a misstep. Addressing the connection between smaller size and reform efforts requires quasi-experimental studies comparing schools similar in size and contextual location and different only with regard to the curricular and instructional strategies they deploy. No work of this type has to our knowledge been reported. Such studies would prove an extremely valuable contribution to the literatures on policy and curriculum and instruction, but they simply must take substantive rural and non-rural contextual differences into account. Otherwise they are likely to obscure relevant issues instead of clarifying them. This is a sharp challenge for most educational researchers, who harbor little experience of, interest in, or concern for rural communities.<sup>31</sup>

Studies that examine the nature of relationships and communication in smaller schools are also needed (see Lee & Loeb, 2003, for a relevant study conducted in the context of urban reform). Arguably, participation in a smaller enterprise enables a more sure-footed construction of self in the company of trusted others as well as a greater investment by all

---

<sup>31</sup> The general tendency among educational researchers is to excuse their own lack of concern by invoking (most commonly only to themselves) the familiar stereotypes that render rural places as simultaneously inadequate, idyllic, and idiotic (Karl Marx famously dismissed “the idiocy of rural life.”)

participants in the construction of common purpose (see e.g., Meier, 2003). Such benefits might persist regardless of the particular instructional and curricular arrangements preferred by participants, and research efforts must provide for that hypothesis to ensure their objectivity.

In the absence of studies unpacking the dynamics of school size, however, viewing smaller size as a proxy measure for transportable practices or processes seems misdirected at best. At worst it deflects effort away from relatively simple policy initiatives, such as resisting the temptation to close smaller schools where they already exist and establishing autonomous smaller schools where they do not exist, toward far more ambitious and difficult ones, such as changing the practices of school teachers and administrators across the nation. One might, from a systems perspective, say that by comparison to the influence of smaller size, such reform efforts are distinctly unpromising. Change of this sort has, at any rate, so far proven remarkably costly and ineffective, at least in the estimation of such astute critics as Richard Gibboney (1994) and David Tyack and Larry Cuban (1995). We ourselves favor certain of these reforms, but it seems that the purposes of and expectations for such efforts ought to be substantively reconceived.

Finally, one finding from our multi-level analyses raises a momentous issue worthy of much more study in our view. That finding suggests that students achieving below the mean in their schools perform better if the schools they attend are below the SES mean for all schools. This intriguing finding relates to our assertion that a subtle social-class balance may help maximize achievement level and equity. Engineering this balance in any comprehensive way seems quite improbable. This finding, instead of prompting precipitous actions, suggests to us the need to examine more closely the ways in which class prejudice operates, especially in schools serving mixed-SES and affluent communities. It also suggests the need for much more extensive analyses of the relationship between inequality in civil society (especially as related to residence) and class-related achievement inequality in schools. Such work is particularly needed if the nation is ever to take seriously the challenge of closing the achievement gap (sometimes known as “leaving no child behind”). The proposition that closing the achievement gap is the lever by which society will become more equitable (the seeming logic behind much current legislation) strikes us unsupportable and ultimately deceptive. The absence of such research, moreover, is arguably a sign of its critical importance in an age of increasing income and wealth inequality.

## **Ten Practical Recommendations**

We conclude this report with recommendations intended to be as unambiguous as the conclusions developed by Lee and Smith. The difference between those conclusions and these recommendations is that the recommendations are framed to remain debatable by a lay audience precisely because they are *recommendations*. We explicitly disavow the privileged status that conclusions of research so often enjoy, whereby conclusions of research appear to lay audiences as unassailable. Our recommendations are partially warranted by the findings from this and related studies, but they are also informed by our own experience, by the experience of some of our colleagues and students, and by our reading and other writing. We present these recommendations as our *considered judgments*:

1. Sustain the smallest schools in the poorest communities.
2. In communities that serve all social classes, do not build large schools.

3. In affluent communities (or attendance zones), do not build high schools larger than 1,000 students.
4. Keep elementary and middle schools proportionately smaller than high schools.
5. When building new, keep schools everywhere smaller than recommended in the 20<sup>th</sup> century.
6. Provide appropriate and adequate support to smaller schools: small size improves the odds of success, it does not guarantee it.
7. Attend to rural and urban issues of size with equal care.
8. Regard smaller school size and reform as distinct issues, but do not hesitate to innovate in smaller schools.
9. Base smaller schools in extant communities so as to avoid the intentional concentration of impoverished students from mixed-SES communities into smaller schools (e.g., as in contemporary “alternative schools,” so-called).
10. Doubt that an educationally-relevant lower limit of school size exists. Much depends on context, and even in the contemporary world, dedicated parents educate very small groups of children with remarkable success at home.

**Note:** This research was partially supported by grants from the National Science Foundation (Grant No. 0119679) and the Rural School and Community Trust. Any opinions, findings, and conclusions or recommendations expressed in this report are exclusively those of the authors and as such do not necessarily reflect the views of the National Science Foundation or the Rural School and Community Trust. This paper was presented at the annual meeting of the American Educational Research Association, San Diego, CA, April 2004.

## References

- Abbott, M., Joireman, J., & Stroh, J. (2002). *The influence of district size, school size, and socioeconomic status on student achievement in Washington: A replication study using hierarchical linear modeling* (Technical Report No. 3). Lynnwood, WA: Washington School Research Center. (Retrieved January 14, 2004 from <http://www.spu.edu/orgs/research/districtschoolsizechreport3.html>)
- Adams, H. (1907/1943). *The education of Henry Adams*. New York: Modern Library.
- Alspaugh, J. (1998, February/March). The relationship of School-to-school transitions and school size to high school dropout rates. *The High School Journal*, 154-160.
- Alspaugh, J., & Gao, R. (2003). *School size as a factor in elementary school achievement*. Unpublished manuscript, University of Missouri, Columbia.
- American Institutes for Research. (2003). *AM statistical software* (Version 0.06.01 Beta, released October 27, 2003). Washington, DC: Author. Software available online from: <http://am.air.org/about2.asp>

- American Institutes for Research and SRI, Inc. (2003). *High time for high school reform: Early findings from the evaluation of the national school district and network grants program*. Washington, DC, and Menlo Park, CA: Authors. (Retrieved January 15, 2004 from [http://www.gatesfoundation.org/nr/downloads/ed/smallschools/Small\\_schools\\_eval\\_ExecSumm\\_2003.pdf](http://www.gatesfoundation.org/nr/downloads/ed/smallschools/Small_schools_eval_ExecSumm_2003.pdf))
- Bickel, R. (manuscript submitted for publication). *Multi-level analysis for policy analysts: It's just regression!* Huntington, WV: Marshall University.
- Bickel, R. (1999b). *School size, socioeconomic status, and achievement: A Georgia replication of inequity in education*. Randolph, VT: Rural Challenge Policy Program. (ERIC Document Reproduction Service No. ED 433 985)
- Bickel, R. (1999b). *School size, socioeconomic status, and achievement: A Texas replication of inequity in education*. Randolph, VT: Rural Challenge Policy Program. (ERIC Document Reproduction Service No. ED 433 986)
- Bickel, R., & Howley, C. (2000). The influence of scale on student performance: A multi-level extension of the Matthew principle. *Education Policy Analysis Archives* (Online), 8(22). Retrieved February 11, 2002 from: <http://epaa.asu.edu/epaa/v3n18.html>.
- Bickel, R., Howley, C., Williams, T. and Glascock, C. (2001, October 8). High school size, achievement equity, and cost: Robust interaction effects and tentative results. *Education Policy Analysis Archives*, 9(40). Retrieved October 8, 2001 from <http://epaa.asu.edu/epaa/v9n40.html>.
- Boyer, E. (1983). *High school: A report on secondary education in America*. Philadelphia: Harper & Row.
- Coleman, J., Hoffer, T., & Kilgore, S. (1981). Public and private schools (Contractor Draft Report). Washington, DC: National Center for Education Statistics. Retrieved April 10, 2004 from <http://www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid=82230D>
- Cotton, K. (1996). *School size, school climate, and student performance* (Close-Up No. 20). Portland, OR: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED397476)
- Cubberley, E. (1922). *Rural life and education: A study of the rural-school problem as a phase of the rural-life problem* (Revised Edition). NY: Houghton-Mifflin. (Original work published 1915; ERIC Document Reproduction Service No. ED392559)
- Curtin, T., Ingels, S., Wu, S., & Heuer, R. (2002). *National Educational Longitudinal Study of 1988: Base year to fourth follow-up data user's manual*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

- Darling-Hammond, L., Aneess, J., & Ort, S. (2002). Reinventing high school: Outcomes of the Coalition Campus Schools Project. *American Educational Research Journal*, 39(3), 639-673.
- Friedkin, N., & Necochea, J. (1988). School system size and performance: A contingency perspective. *Educational Evaluation and Policy Analysis*, 10(3), 237-249.
- Gibboney, R. (1994). *The stone trumpet: A story of practical school reform, 1960-1990*. Albany, NY: State University of New York Press.
- Goodlad, J. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill.
- Hennen, J. (1998). *The Americanization of West Virginia: Creating a modern industrial state, 1916-1925*. Lexington, KY: University of Kentucky Press.
- Howley, C. (1995). The Matthew principle: A West Virginia replication? *Education Policy Analysis Archives*, 3(18). Retrieved February 13, 2002, from <http://seamonkey.ed.asu.edu/epaa/v3n18.html>.
- Howley, C. (1996a). 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. (1996b). *Sizing up schooling: A West Virginia analysis and critique*. Dissertation Abstracts International(A), 57(3), 940. (University Microfilms No. AAT 9622575)
- Howley, C. (1999a). *The Matthew Project: State report for Montana*. Randolph, VT: Rural Challenge Policy Program. (ERIC Document Reproduction Service No. ED 433 173)
- Howley, C. (1999b). *The Matthew Project: State report for Ohio*. Randolph, VT: Rural Challenge Policy Program. (ERIC Document Reproduction Service No. ED 433 175)
- Howley, C., & Bickel, R. (1999). *The Mathew Project: National report*. Randolph, VT: The Rural School and Community Trust. (ERIC Document Reproduction Service No. ED 433 174)
- Howley, C. (2001). *Research on smaller schools: What education leaders need to know to make better decisions*. Arlington, VA: Educational Research Service.
- Howley, C. (2002). Small schools. In A. Molnar (Ed.), *School reform proposals: The research evidence* (pp. 49-77). Greenwich, CT: Information Age Publishing.
- Huang, G., & Howley, C. (1993). Mitigating disadvantage: Effects of small-scale schooling on student achievement in Alaska. *Journal of Research in Rural Education*, 9(3), 137-149.
- Jencks, C. (1985). How much do high school students learn? *Sociology of Education*, 58(2), 128-135.

- Johnson, F. (1989). *Assigning type of locale codes to the 1987-88 CCD public school universe*. Paper presented to the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED312113)
- Johnson, J. (2003). *Nebraska's smaller school systems counter the harmful effects of poverty on student achievement*. Washington, DC: Rural School and Community Trust. (Retrieved January 14, 2004 from [http://ruraledu.org/docs/nebraska/Nebraska\\_Matthew12-03.pdf](http://ruraledu.org/docs/nebraska/Nebraska_Matthew12-03.pdf))
- Johnson, J., Howley, C., & Howley, A. (2002). *Size, excellence, and equity: A report on Arkansas schools and districts*. Athens, OH: Educational Studies Department, Ohio University.
- Kennedy, J. (1915). *Rural life and the rural school*. New York: American Book Company. (ERIC Document Reproduction Service No. ED392561)
- Klonsky, M. (1995). *Small schools: the numbers tell a story, a review of the research and current experiences*. Chicago: The Small Schools Workshop. (ERIC Document Reproduction Service No. ED386517)
- Lee, V. E., & Loeb, S. (2000). School size in Chicago elementary schools: Effects on teachers' attitudes and students' achievement. *American Educational Research Journal*, 37(1), 3-31.
- Lee, V., & Smith, J. (1995). Effects of high school restructuring and size on early gains in achievement and engagement. *Sociology of Education*, 68(4), 241-270.
- Lee, V., & Smith, J. (1997). High school size: Which works best, and for whom? *Educational Evaluation and Policy Analysis*, 19(3), 205-227.
- Mathews, D. (1996). *Is there a public for public schools?* Dayton, OH: Kettering Foundation Press.
- Meier, D. (2003). *In schools we trust: Creating communities of learning in an era of testing and standardization*. Boston: Beacon Press.
- Miley and Associates. (2003). *School district organization in South Carolina: Evaluating performance and fiscal efficiency* (Prepared for The Education Oversight Committee). Columbia, SC: Author. Retrieved March 23, 2003 from <http://www.state.sc.us/eoc/PDF/MileyReportFinalDraft010903A.doc>
- Monk, D., & Haller, E. (1993). Predictors of high school academic course offerings: The role of school size. *American Educational Research Journal*, 30(1), 3-21.
- Raywid, M. (1999). *Current literature on small schools* (ERIC Digest). Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools. (ERIC Document Reproduction Service No. ED425049)
- Raywid, M., & Schmerler, G. (2003). *Not so easy going: The policy environments of small urban schools and schools-within-schools*. Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools.

- Simon, H. (1979). Rational decision making in business organizations. *The American Economic Review*, 69(4), 493-513.
- Spencer, B., Frankel, M., Ingels, S., Rasinski, K., & Tourangeau, R. (1990). *National Educational Longitudinal Study of 1988: Base year sample design report* (Technical Report). Washington, DC: National Center for Education Statistics. (Retrieved January 17, 2004 from <http://nces.ed.gov/pubs90/90463.pdf>)
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. Thousand Oaks, CA: SAGE.
- Stemnock, S. (1974). *Summary of research on size of schools and school districts*. Arlington, VA: Educational Research Service. (ERIC Document Reproduction Service No. ED140458).
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

## About the Authors

### **Craig Howley**

Ohio University

Email: [howleyc@ael.org](mailto:howleyc@ael.org)

Craig Howley has researched rural education and published widely in that field. He has taught mathematics at the University of Charleston and has evaluated mathematics professional development projects in rural schools. He is coeditor of *Small High Schools That Flourish* (AEL, 2000) and coauthor *Out of Our Minds: Anti-intellectualism in American Schooling* (TC Press, 1995). Howley co-directs the Research Initiative of the Appalachian Collaborative Center for Learning, Assessment, and Instruction in Mathematics (ACCLAIM) and is an adjunct professor in the Educational Studies Department at Ohio University. His most recent empirical work investigates the mediating influence of school and district size on academic achievement, construction costs of smaller schools, rural school busing, rural principals' perspectives on planning. His work for ACCLAIM entails cultivation of research capacity in rural mathematics education. Previously, he directed the ERIC Clearinghouse on Rural Education and Small Schools.

### **Aimee Howley**

Ohio University

Email: [howley@ohio.edu](mailto:howley@ohio.edu)

**Aimee Howley** is professor of educational studies and coordinator of the educational administration program at Ohio University. Her recent research focuses on leadership and policy issues, particularly as they pertain to rural schools and communities.

***Education Policy Analysis Archives***

<http://epaa.asu.edu>

**Editor: Gene V Glass, Arizona State University**

Production Assistant: Chris Murrell, Arizona State University

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

***EPAA Editorial Board***

**Michael W. Apple**  
University of Wisconsin

**Greg Camilli**  
Rutgers University

**Sherman Dorn**  
University of South Florida

**Gustavo E. Fischman**  
Arizona State University

**Thomas F. Green**  
Syracuse University

**Craig B. Howley**  
Appalachia Educational Laboratory

**Patricia Fey Jarvis**  
Seattle, Washington

**Benjamin Levin**  
University of Manitoba

**Les McLean**  
University of Toronto

**Michele Moses**  
Arizona State University

**Anthony G. Rud Jr.**  
Purdue University

**Michael Scriven**  
University of Auckland

**Robert E. Stake**  
University of Illinois—UC

**Terrence G. Wiley**  
Arizona State University

**David C. Berliner**  
Arizona State University

**Linda Darling-Hammond**  
Stanford University

**Mark E. Fetler**  
California Commission on Teacher  
Credentialing

**Richard Garlikov**  
Birmingham, Alabama

**Aimee Howley**  
Ohio University

**William Hunter**  
University of Ontario Institute of  
Technology

**Daniel Kallós**  
Umeå University

**Thomas Mauhs-Pugh**  
Green Mountain College

**Heinrich Mintrop**  
University of California, Berkeley

**Gary Orfield**  
Harvard University

**Jay Paredes Scribner**  
University of Missouri

**Lorrie A. Shepard**  
University of Colorado, Boulder

**Kevin Welner**  
University of Colorado, Boulder

**John Willinsky**  
University of British Columbia

## **AAPE Editorial Board**

### **Associate Editors**

**Gustavo E. Fischman & Pablo Gentili**

**Arizona State University & Universidade do Estado do Rio de Janeiro**

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez

**Hugo Aboites**

Universidad Autónoma  
Metropolitana-Xochimilco

**Dalila Andrade de Oliveira**

Universidade Federal de Minas  
Gerais, Belo Horizonte, Brasil

**Alejandro Canales**

Universidad Nacional  
Autónoma de México

**Erwin Epstein**

Loyola University, Chicago,  
Illinois

**Rollin Kent**

Universidad Autónoma de  
Puebla. Puebla, México

**Daniel C. Levy**

University at Albany, SUNY,  
Albany, New York

**María Loreto Egaña**

Programa Interdisciplinario de  
Investigación en Educación, Chile

**Grover Pango**

Foro Latinoamericano de  
Políticas Educativas, Perú

**Angel Ignacio Pérez Gómez**

Universidad de Málaga

**Diana Rhoten**

Social Science Research  
Council, New York, New York

**Susan Street**

Centro de Investigaciones y  
Estudios Superiores en  
Antropología Social Occidente,  
Guadalajara, México

**Antonio Teodoro**

Universidade Lusófona Lisboa,

**Lilian do Valle**

Universidade Estadual do Rio  
de Janeiro, Brasil

**Adrián Acosta**

Universidad de Guadalajara  
México

**Alejandra Birgin**

Ministerio de Educación,  
Argentina

**Ursula Casanova**

Arizona State University,  
Tempe, Arizona

**Mariano Fernández Enguita**

Universidad de Salamanca.  
España

**Walter Kohan**

Universidade Estadual do Rio  
de Janeiro, Brasil

**Nilma Limo Gomes**

Universidade Federal de Minas  
Gerais, Belo Horizonte

**Mariano Narodowski**

Universidad Torcuato Di Tella,  
Argentina

Vanilda Paiva

**Universidade Estadual do**

**Rio de Janeiro, Brasil**

**Mónica Pini**

Universidad Nacional de San  
Martín, Argentina

**José Gimeno Sacristán**

Universidad de Valencia,  
España

**Nelly P. Stromquist**

University of Southern  
California, Los Angeles,  
California

**Carlos A. Torres**

University of California, Los  
Angeles

**Claudio Almonacid Avila**

Universidad Metropolitana de  
Ciencias de la Educación, Chile

**Teresa Bracho**

Centro de Investigación y  
Docencia Económica-CIDE

**Sigfredo Chiroque**

Instituto de Pedagogía Popular,  
Perú

**Gaudêncio Frigotto**

Universidade Estadual do Rio  
de Janeiro, Brasil

**Roberto Leher**

Universidade Estadual do Rio  
de Janeiro, Brasil

**Pia Lindquist Wong**

California State University,  
Sacramento, California

**Iolanda de Oliveira**

Universidade Federal  
Fluminense, Brasil

**Miguel Pereira**

Catedrático Universidad de  
Granada, España

**Romualdo Portella do**

**Oliveira**

Universidade de São Paulo

**Daniel Schugurensky**

Ontario Institute for Studies in  
Education, Canada

**Daniel Suarez**

Laboratorio de Políticas  
Públicas-Universidad de  
Buenos Aires, Argentina

**Jurjo Torres Santomé**

Universidad de la Coruña,  
España