Aggregation and the Estimated Effects of School Resources

Hanushek, Eric A., Steven G. Rivkin and Lori L. Taylor

Working Paper No. 397
February 1995

Aggregation and the Estimated Effects of School Resources

Eric A. Hanushek, Steven G. Rivkin and Lori L. Taylor

Rochester Center for Economic Research
Working Paper No. 397

February 1995

Aggregation and the Estimated

Effects of School Resources

by

Eric A. Hanushek*

Steven G. Rivkin**

and

Lori L. Taylor***

February 1995

abstract

This paper attempts to reconcile the contradictory findings in the debate over school resources and school effectiveness by highlighting the role of aggregation in the presence of omitted variables bias. While data aggregation for well-specified linear models yields unbiased parameter estimates, aggregation alters the magnitude of any omitted variables bias. The theoretical impact of aggregation is ambiguous, but analysis of High School and Beyond data strongly suggests that aggregation inflates the coefficients on school characteristics. Moreover, the pattern of results is not consistent with an errors-in-variables explanation, the alternative explanation for the larger estimated impact with aggregate estimates. Since studies using aggregate data are much more likely to find positive school resource effects on achievement, these results provide strong evidence against the view that additional expenditures alone are likely to improve student outcomes.

Revised version of the paper prepared for the meeting of the
American Economic Association, Washington, DC, January 6-8, 1995.

---

*University of Rochester

**Amherst College

***Federal Reserve Bank of Dallas

Aggregation and the Estimated Effects of School Resources

by Eric A. Hanushek, Steven G. Rivkin, and Lori L. Taylor*

I. Introduction

A key element of the policy discussion surrounding schools has been the effect of additional resources on student performance. In simplest terms, if schools effectively turn added resources into higher student achievement, policy makers can concentrate on the appropriate level and distribution of resources and can let the local school districts concentrate on uses of those resources. However, if schools do not effectively turn resources into performance, then education policy that is aimed at either the level or distribution of outcomes becomes much more complicated. Policy makers either must concentrate on picking good approaches and processes—something that school districts themselves may not be able to do—or they must turn to different incentive mechanisms that might alter the way districts spend their available resources.[1]

A growing body of research casts doubt on the effectiveness of local school districts at turning added resources into higher student achievement. In comprehensive summaries of empirical evidence, Hanushek (1986, 1989) found that there was no consistent or systematic relationship between achievement and either pupil/teacher ratios, teacher salaries, years of teacher schooling, years of teacher experience or per-student expenditure. The inefficacy of smaller pupil/teacher ratios is particularly noteworthy, given that the widely held belief that lower pupil/teacher ratios improve educational outcomes has played a prominent role in the crafting of educational policies. Between 1940 and 1990 the average pupil/teacher ratio in the United States public schools has fallen from 28

---

*We would like to thank Julian Betts, Jeff Grogger, Robert Willis, Geoffrey Woglom, and participants at the NSF/Review of Economics and Statistics conference on *School Quality and Educational Outcomes* (Harvard University, December 1994) for helpful comments.

[1]A discussion of alternative policy approaches in the absence of clear relationships between resources and policies is found in Hanushek with others (1994).

to less than 16.[2] Though the expansion of special education accounts for a portion of the decline, the vast majority of the change reflects a fall in mainstream class sizes.[3]

The conclusion that higher expenditures (including lower pupil/teacher ratios) have not raised achievement has been far from universally accepted. Proponents of meta-analysis[4] point out that weighted averages of the individual estimates of school resource effects are positive, however the vast differences among specifications raises serious questions about the validity of using weights based upon the estimated standard errors. Others argue that the use of standardized test scores as the measure of achievement in most research ignores the potential impact of smaller class sizes or higher teacher salaries on other aspects of educational performance. Two recent studies by Card and Krueger (1992a, 1992b) provide evidence that smaller classes and higher teacher salaries increase the wage premium associated with an additional year of schooling.

This paper attempts to reconcile the contradictory findings in the debate over the effectiveness of additional school resources in raising achievement. Two important differences among the various studies of educational performance are the level at which school characteristics are aggregated and the number and type of controls for differences in student backgrounds, opportunities and academic preparation.[5] Some studies analyze the relationship between student outcomes and individual school and school district characteristics, while others, including Card and Krueger, use state average school

---

[2]These calculations include school principals in addition to teachers.

[3]Hanushek and Rivkin (1994) analyze the contribution of special education programs to the pupil/teacher ratio decline. Between 1980 and 1990, the period most often cited for the influence of special education, less than one third of the fall in the pupil/teacher ratio can be attributed to the expansion of special education.

[4]See Hedges et al. (1994) and the discussion below.

[5]Betts (forthcoming) examines specific aspects of the Card and Krueger methodology in order to understand why their findings contradict much of the previous literature. See also Heckman, Layne-Farrar, and Todd (1994) and Speakman and Welch (1995) for an examination of the sensitivity of the estimates to sample and specification.

characteristics. There is even greater variation in the variables used to control for differences in socio-economic backgrounds and opportunities.

The probability that studies report a positive and statistically significant relationship between achievement and school resources rises dramatically along with the level of aggregation (see below). This pattern holds when achievement is measured by test scores or other outcomes.[6] While most researchers suggest that disaggregated analysis is generally superior, some argue that aggregation of the relationships may actually have beneficial effects through reduction in measurement error. Thus a crucial issue is identifying whether aggregation reduces a downward bias present in school-level studies (perhaps by reducing measurement error) or introduces a positive bias which inflates school resource coefficients.

We present an empirical model demonstrating that aggregation alters the degree of omitted variables bias, even when the true marginal impacts of included variables are constant across observations. Omitted variables have their clearest effects on estimates when the data are aggregated to the level of the omitted factors (such as when the data are aggregated to the state level and state-level determinants of students performance are neglected). Because data limitations inevitably lead to the exclusion of variables that influence both school outcomes and school expenditures, the model suggests that aggregation-induced changes in the magnitude of omitted variables bias could explain the pattern of results produced by school resource research.

Using the High School and Beyond (HSB) data set, we estimate aggregated and disaggregated relationships between the primary determinants of school expenditures (teacher/pupil ratios and teacher salaries) and two measures of educational outcomes (standardized test scores and years of post-secondary schooling). Consistent with prior research, our analysis finds that aggregation inflates

[6]Betts (1995) finds a very similar pattern for studies in which earnings are used to measure achievement.

the coefficients on school characteristics for both outcome measures. More importantly, the pattern of results in not consistent with an errors-in-variables explanation, rather aggregation appears to exacerbate problems of omitted variables bias and produce incorrectly large school resource coefficients.

II. Model

Following the literature on estimating the value added by schools (see, e.g., Hanushek (1979), Aitkin and Longford (1986) or Hanushek and Taylor (1990)), we model the relationship between educational attainment and student and family characteristics as:

$$(1) \qquad A_{ij} = \mathscr{F}(T_{ij}, F_{ij}, C_{ij}, S_{ij}).$$

where A_{ij} is the level of educational achievement for individual i in school j, T_{ij} is a standardized pretest score, F_{ij} is a vector of individual and family characteristics, C_{ij} is a vector of community variables, and S_{ij} is a vector of school characteristics.

Adequate controls for differences in family background, community environment, and student preparation are needed in order to isolate the effects of school characteristics. Education goes on both inside and outside schools. The performance of any specific student will combine the influences of the school and of the outside environment, particularly the family. Moreover, parents may systematically select school districts through migration in accordance with their preferences (Tiebout 1956) or otherwise attempt to secure good school resources for their children. In such a case, unmeasured parental inputs could be correlated with measured school resources.[7]

[7]It is often argued that aggregation eliminates the endogeneity problem because students are likely to choose among schools within a limited geographic area. Yet there is no reason to believe that people limit their choices in such a way.

Accounting for pre-existing differences in academic preparation is necessary in order to capture the impact of school factors during a given period. Education occurs over time, so that the achievement, say, of a ninth grader is determined in part by schools (and family) in the ninth grade and in part by these inputs in prior years. Since data on the past history of educational inputs are frequently unavailable, strategies that will isolate the achievement gains that might be related to specific measured inputs such as through estimation of value-added models of achievement are frequently employed. In addition, schools may use past performance in determining input allocations (e.g., lower class size for poorly performing students). Such a possibility reinforces the need to control for past performance in analyzing the effects of school resources.

These general issues in the estimation of educational production functions are set out at the beginning because they pervade the discussion here. Clearly, inadequate controls for academic preparation, family inputs, and the like will bias the estimated effects of school characteristics on achievement. At the same time, virtually no attention has been given to how aggregation of data might interact with such specification bias. We develop a conceptual model that shows how the level of data aggregation will affect the magnitude of any omitted variables bias and then subject it to empirical analysis.

III Aggregation, Omitted Variable Bias and Measurement Error

Virtually all discussion of aggregation of models, measurement errors, and omitted variables is conducted in the context of a simple linear model. While some have suggested that such models are inappropriate in many of the educational circumstances investigated here, we neglect consideration of nonlinearities and concentrate on the other issues.[8]

[8]For example, both Summers and Wolfe (1977) and Ferguson (1992) suggest significant nonlinearities in the effects of class size on student achievement. Summers and Wolfe further suggest that class size may have differential effects depending on the level of achievement of individual

The level of aggregation can influence the estimated relationship between educational outcomes and specific school characteristics in a number of ways.[9] This section examines aggregation related issues in the simplest form using regression equation 2:

$$A_{ijs} = \alpha_{ijs} + \beta_{ijs}T_{ijs} + \eta_{ijs}F_{ijs} + \theta_{ijs}C_{ijs} + \psi_{ijs}S_{ijs} + \epsilon_{ijs} \tag{2}$$

where the subscript s indexes state of residence and ϵ_{ijs} is an i.i.d. random error. To fix ideas, assume that S_{ijs} represents a single measure of school quality, say, per-student educational expenditure.

Most studies of educational production functions implicitly assume that the coefficient ψ_{ijs} is invariant across individuals, schools and state of residence.[10] In other words, the marginal impact of a change in school expenditure is the same regardless of socio-economic background, academic skill, or even the value of school expenditure: $\psi_{ijs} = \psi$ for all i, j, and s. Under such conditions, aggregation to the district or state level will not alter the estimated relationship between attainment and school expenditure (although the form of the data will generally affect the efficiency of the estimates).[11]

---

students. Bryk and Raudenbush (1992) examine of a variety of specialized nonlinear models in the context of schools.

[9]See Theil (1971) for a comprehensive discussion of aggregation bias.

[10]Theil (1971) shows that this assumption is a sufficient condition for perfect aggregation for a linear specification.

[11]If this assumption is not valid, the estimate of ψ can no longer be interpreted as the marginal impact of a change in school expenditure, because there is no single marginal impact. In specific instances it can it be meaningfully interpreted as an average marginal impact as derived from a random coefficients model (see, for example, Swamy 1970). However, depending on the distribution of the underlying parameters, changes in student composition or the influences of other variables could alter the coefficient estimate and interpretation. Proponents of hierarchical linear modelling (e.g., Bryk and Raudenbush 1992) and others argue that the impacts of school characteristics vary by socio-economic background. However, exploratory analysis in our empirical modeling indicated that the estimated impacts of school characteristics on attainment did not vary significantly by either

Unfortunately, equal marginal effects for all students is a sufficient condition for perfect aggregation only when the empirical model is correctly specified. In practice, information for certain relevant variables might be unavailable. Such data limitations are frequently most severe when aggregate data are employed, as opposed to more detailed survey or individual record information. If the omitted variables are correlated with school expenditure, then the estimated school expenditure coefficient will be biased regardless of the level of aggregation.

We consider omission of an important community factor, C_{ijs}. The magnitude of omitted variable bias depends upon both the coefficients on the omitted variables θ and the coefficients on the included school characteristics from auxiliary regressions of the omitted variables on the included school characteristics. If aggregation alters any of these coefficients, it will change the size of the bias and thus the estimated relationship between attainment and school expenditure. As the following discussion illustrates, there is no general rule that indicates whether aggregation will increase or reduce the size of the bias. Even in the simple two variable case, the direction of the bias cannot always be determined a priori.

A. Aggregation and Omitted Variables in a Simple Two-Variable Case

Equation 3 presents a simplified version of equation 2 that ignores all variation in pretest scores and family backgrounds and subsumes the influences of all community factors into a single measure of community environment. By assumption, the marginal impacts of both school expenditure and community environment are identical for all students.

$$\textbf{(3)} \qquad A_{js} = \theta C_{js} + \psi S_{js} + \epsilon_{js}.$$

If there is no information on the relevant aspects of community environment, a regression of

---

student pretest score or race. Therefore, we will maintain the assumption that the marginal effects of school characteristics are equal for all students throughout the analysis.

academic attainment on school expenditure will produce the following biased estimate of the school expenditure coefficient:

$$(4) \quad \hat{\psi} = \psi + \theta\phi,$$

where ϕ equals the school expenditure coefficient in an auxiliary regression of community environment on school expenditure. Even if equation 3 satisfies the conditions for perfect aggregation, aggregation will alter the bias in the estimate of ψ through its impact on the coefficient ϕ.

The effects of aggregation on the coefficient ϕ are analyzed assuming that both C and S are driven by a common underlying factor, Λ, which indexes, say, tastes for education. With a linear specification,

$$(5) \quad C_{js} = \mu(\Lambda_{js} - \bar{\Lambda}_s) + \xi\bar{\Lambda}_s + U_{js},$$

$$(6) \quad S_{js} = \gamma(\Lambda_{js} - \bar{\Lambda}_s) + \delta\bar{\Lambda}_s + V_{js}.$$

We allow both local and state values of Λ to effect school expenditure, reflecting the influence of both the state and local political processes in determining school revenues and expenditure.[12] We define C_{js}, Λ_{js} and Λ_s so that μ, ξ, γ and δ are all non-negative, and restrict δ to be greater than γ so that school expenditures do not fall as Λ_s rises. U_{js} and V_{js} are i.i.d. random errors that are independent of each other and the variable Λ.

We first examine the impact of aggregation when a local community characteristic is omitted. Following that, we turn to the case were a state-level variable is excluded.

Omitted Local Factors

The coefficient ϕ equals the covariance of C and S divided by the variance of S. Using

---

[12]State mean expenditure levels reflect both local and state expenditures. Therefore, within state variation results both from differences in local values of Λ and the method by which states allocate money to districts.

equations 5 and 6, and assuming that only local values of Λ effect the community environment ($\mu = \xi$), this equals

$$\textbf{(7)} \qquad \phi = \frac{\xi\,(\gamma\sigma^2_{\Lambda_w} + \delta\sigma^2_{\Lambda_b})}{\gamma^2\sigma^2_{\Lambda_w} + \delta^2\sigma^2_{\Lambda_b} + \sigma^2_{V_w} + \sigma^2_{V_b}}$$

where the variances of Λ and V (σ^2) have been partitioned into within state (subscript w) and between state (subscript b) components.

When the data are aggregated to the state level, equations 5 and 6 are rewritten, with bars indicating state average values, as:

$$\textbf{(8)} \qquad \bar{C}_s = \xi\bar{\Lambda}_s + \bar{U}_s$$

$$\textbf{(9)} \qquad \bar{S}_s = \delta\bar{\Lambda}_s + \bar{V}_s$$

and $\bar{\phi}$, the aggregate auxiliary regression coefficient, equals

$$(10) \qquad \bar{\phi} = \frac{\xi\delta\sigma^2_{\Lambda_b}}{\delta^2\sigma^2_{\Lambda_b} + \sigma^2_{V_b}}$$

The difference in the auxiliary regression coefficients, $\phi - \bar{\phi}$, equals

$$(11) \qquad \phi - \bar{\phi} = \frac{\xi\,(\gamma\sigma^2_{\Lambda_w} + \delta\sigma^2_{\Lambda_b})}{\gamma^2\sigma^2_{\Lambda_w} + \delta^2\sigma^2_{\Lambda_b} + \sigma^2_{V_w} + \sigma^2_{V_b}} - \frac{\xi\delta\sigma^2_{\Lambda_b}}{\delta^2\sigma^2_{\Lambda_b} + \sigma^2_{V_b}}$$

Aggregation increases the omitted variable bias when $\phi - \bar{\phi} < 0$, and decreases the bias when $\phi - \bar{\phi} > 0$. Let Γ be the positive term

(12) $$\Gamma = \xi \delta \gamma \sigma^2_{\Lambda_w} \sigma^2_{\Lambda_b}$$

Dividing equation 11 by Γ does not change the sign of $\phi - \bar{\phi}$:

(13) $$\frac{\phi - \bar{\phi}}{\Gamma} = \delta - \gamma - \frac{\sigma^2_{v_w}}{\gamma \sigma^2_{\Lambda_w}} + \frac{\sigma^2_{v_b}}{\delta \sigma^2_{\Lambda_b}}.$$

Equation 13 indicates that a multitude of factors determine the direction of aggregation effects even in the case where C and S are single variable representations of community and school characteristics. The following partial derivatives describe the influences of the variance components and regression coefficients on the sign of $(\phi - \bar{\phi}) / \Gamma$ (abbreviated as D):

$$
(14) \qquad
\begin{aligned}
\frac{\partial D}{\partial \sigma^2_{\Lambda_b}} &= -\frac{\sigma^2_{V_b}}{\delta(\sigma^2_{\Lambda_b})^2} &&<0, \\
\frac{\partial D}{\partial \sigma^2_{\Lambda_w}} &= \frac{\sigma^2_{V_w}}{\gamma(\sigma^2_{\Lambda_w})^2} &&>0, \\
\frac{\partial D}{\partial \sigma^2_{V_b}} &= \frac{1}{\delta\sigma^2_{\Lambda_b}} &&>0, \\
\frac{\partial D}{\partial \sigma^2_{V_w}} &= -\frac{1}{\gamma\sigma^2_{\Lambda_w}} &&<0, \\
\frac{\partial D}{\partial \delta} &= 1 - \frac{\sigma^2_{V_b}}{\delta^2\sigma^2_{\Lambda_b}} &&? \quad \text{and} \\
\frac{\partial D}{\partial \gamma} &= -1 + \frac{\sigma^2_{V_w}}{\gamma^2\sigma^2_{\Lambda_w}} &&?.
\end{aligned}
$$

Increases in the between-state variance of Λ (the underlying factor determining both attainment and school expenditure) and the within-state variance of V (the error in the auxiliary school expenditure equation) raise the probability that aggregation exacerbates omitted variable bias. On the other hand, increases in the within-state variance of Λ and the between-state variance of V lower the probability that aggregation worsens omitted variable bias. The relative sizes of these variance components thus determine the impact of aggregation.

Omitted State Effects

The ambiguity of the aggregation bias disappears, however, if the omitted community effects are determined at the state level with no within state variation (μ=0). This might occur if overall state regulations and policies directly affected the character of student learning. In such a case, the relevant community factor for both the individual and aggregate relationship is determined by equation 8. In this situation, ϕ becomes:

$$\phi = \frac{\xi \delta \sigma^2_{\Lambda_b}}{\gamma^2 \sigma^2_{\Lambda_w} + \delta^2 \sigma^2_{\Lambda_b} + \sigma^2_{v_w} + \sigma^2_{v_b}} \tag{15}$$

Compare this with the aggregate auxiliary regression coefficient, $\bar{\phi}$, from equation 10. Since the numerators are identical and the denominator in equation 15 is unambiguously larger than that in equation 10, the aggregate estimate of ψ is biased upward from the microlevel estimates. With schools, where the key policies are made at the state level, this model structure appears very relevant.

B. Measurement Error

To this point, we have assumed that the school characteristics are measured without error. An alternative concern about the estimation of school performance models is the possibility of measurement error in the school variable. If data are collected by surveys, their quality may be low when the local respondent is uncertain about the precise values, say, of expenditure or even number of students. On a related issue, if the educational models are aggregated over time instead of employing the basic value-added formulation sketched in equations 1 and 2, year-to-year fluctuations in the data may provide a misleading picture of the relevant historical data.[13] In other words, measurement error is likely to be particularly important as empirical specifications diverge from the ideal described in equation 1.

---

[13]Card and Krueger (1994) highlight fluctuations in capital expenditure or inexplicable annual movements in pupil-teacher ratios as evidence of this sort of error. While annual movements in pupil-teacher ratios or other inputs have little impact on value-added models, they will obviously introduce serious measurement errors in the cumulative inputs.

The effect of measurement error in simple linear models is well-known (see, e.g., Maddala 1977). If the observed school input is:

$$\text{(18)}\quad \tilde{S}_{js} = S_{js} + \nu_{js}$$

then the estimate of ψ will be inconsistent and biased toward zero, even when ν is i.i.d. with mean zero. The coefficients for other variables in the model (measured without error) will also be biased, but the direction of bias is ambiguous. (As discussed below, this conclusion about downward bias only holds for a single mismeasured variable, and the direction is no longer clear with multiple variables measured with error).

An important aspect of random measurement error is that aggregation within defined groups (e.g. states) can lessen the bias from measurement error under certain circumstances (Maddala 1977). Grouping effectively reduces bias as long as the grouping strategy preserves between group differences in the value of the regressor. Employing state average differences in school characteristics suggests that aggregation by state would reduce the downward bias of any measurement error and potentially increase the parameter estimates. Therefore random error in measuring the school characteristics offers a plausible explanation for the fact that state aggregate studies are far more likely to find a positive relationship between achievement and school expenditures than disaggregated school-level studies (see below). However, as discussed, the presence of omitted variables could also produce this identical pattern of inflated state-level estimates. Thus, the source of the effects of aggregation is not identified when aggregation preserves between group differences in an omitted factor which is itself related to both achievement and school expenditures.

C. Complications

The previous analyses concluded that omitted local community factors have an ambiguous effect on the estimated schooling parameter when aggregated to the state level. If, however, the omitted factor applies at the state level, aggregation to the state level will unambiguously bias the schooling parameter upward. Finally, with classical errors-in-variables, aggregation will tend to increase the magnitude of the schooling parameter by virtue of reducing the downward bias imparted by measurement error.

All of these derivations were, nonetheless, conducted within a very simplistic model. First, none of this analysis generalizes in any simple way to nonlinear models. Second, just a single variable was omitted and only a single behavioral parameter was being estimated. The omitted variable bias becomes much more complicated when multiple included and excluded variables are considered, and the effects of aggregation can no longer in general be ascertained. Third, in the simple case of measurement error, the error is assumed independent of the included variables, and systematic measurement error will not have the clear effects previously derived. Fourth, the measurement error model no longer yields any simple predictions when more than one variable is measured with error (see, e.g., Maddala (1977)). With multiple measurement errors, the coefficients are not necessarily biased toward zero, but instead depend on the entire pattern of covariances among the exogenous variables.

The dependence of the theoretical conclusions about aggregation on the specifics of the model and data lead us to examine aggregation empirically within the context of state variations in educational performance. The effects of school resources on both cognitive achievement and schooling attainment are considered at both the school level and the state level to investigate aggregation within a consistent model specification and data source.

IV. Data

Data for the empirical analysis come from the High School and Beyond (HSB) longitudinal survey. The survey is administrated by the National Opinion Research Center under contract with the

Department of Education. The base year of the survey is 1980, at which time approximately 36 high school sophomores and 36 high school seniors from close to 1,000 high schools were interviewed. Follow-up surveys were completed in 1982, 1984, and 1986, yielding six years of information for individuals retained in all follow-up surveys. The base year survey reports parental schooling and family income levels and contains information on the high schools. Six years of schooling histories are contained in the follow-up surveys. In addition, a battery of standardized mathematics, verbal and science tests are administered along with the base year and first follow-up surveys.

Two separate measures of student achievement (A_{ij}) are employed: composite 12th grade test scores and years of schooling attained. The 12th grade composite test score is a linear combination of mathematics and reading scores, where the weights equal the estimated parameters for the two test scores in a regression of the probability of continuing high school on the test scores and other explanatory variables.[14] The rankings of students by this achievement measure is thus not arbitrary, but related to their academic preparation for continuing school. Similar weights are also produced when samples are divided by race and gender.

The second attainment measure is years of post-secondary schooling for high school graduates. It varies from zero (no post-secondary schooling) to eight (a Ph.D., M.D., etc.)[15]. We measure school attainment at roughly age 24 (6 years following high school graduation). Approximately 20 percent of individuals are still in school, and recent trends suggest many nonstudents will return to school at some point in the future. Thus, we are measuring educational attainment 6 years after high school, not total years of schooling.

---

[14]The weight on the mathematics test score is three times greater than the weight on the reading test score, a result that is consistent with other work (e.g., Bishop 1992) suggesting that mathematical skill is disproportionately valued.

[15]Students might complete a degree associated with eight years of post-secondary schooling in fewer than eight calendar years.

The value added framework uses a pretest score as a regressor in order to isolate the contribution of high schools. We use the 10th grade composite test score as the pretest in the analysis of 12th grade test scores, and the 12th grade test score as the pretest in the analysis of post-secondary schooling. Therefore any impact of high schools on post-secondary schooling is in addition to their affects on cognitive achievement.

The vector F_{ij} includes race, gender, parental schooling and family income[16]. C_{ij} includes the percentage of community residents who have a college degree, the local unemployment and wage rates for high school graduates, the resident tuition at state universities, and indicator variables if the school is located in the south or in a rural area. Both local and state aggregates of community factors are analyzed. The percentage of college educated residents captures environmental effects on educational expectations and achievement. Local wages and unemployment rates reflect variations in the opportunity cost of attending college, while the state resident tuition at public universities indexes the monetary cost of college attendance.[17]

Because they are the primary determinants of educational expenditures, we use teacher-pupil ratios and relative teacher salaries to measure school characteristics. We divide starting teacher salaries by the average earnings of college-educated residents in the community (normalized to 40-week salaries) to derive relative salaries. Scaling salaries by the local wage level controls for any differences in the alternative wage opportunities available to teachers.

---

[16]Because of missing data problems, the family income variable is not directly included in the analyses of 12th grade test scores. Instead, the high school family income distribution for same race students who report family income is added as a school characteristic. Individual parental education is included to capture both family education and income effects.

[17]The High School and Beyond survey contains very little information on community environment, therefore we use U.S. Census data to construct measures of local unemployment, wage and college completion rates. The information on university tuition is taken from Peterson's Guide to Four Year Colleges (1984). Rivkin (1991) describes the construction of the community characteristics.

We restrict our attention to non-Hispanic Blacks and Whites attending public high schools, and omit all observations in schools that have fewer than five observations with nonmissing data. We also exclude states with only a single high school in the final sample. We use the sophomore cohort to analyze test scores, and the senior cohort to analyze post-secondary schooling. The sophomore cohort sample includes 11,386 observations for students who attended 627 schools in 46 states, and the senior cohort sample has 2,309 observations for students who attended 307 schools in 38 states.

V. Empirical Results

We use a two stage estimation framework to analyze the impact of aggregation on schooling coefficients. Because we used the same procedures in both the school- and state-level analyses, we will only describe the school-level specifications.

In the first stage, we regressed academic attainment on the pretest score, student and family characteristics, and a series of indicator variable for each high school, E_{ij}:

$$A_{ij} = \beta T_{ij} + \eta F_{ij} + \sum j \alpha_j E_{ij} + \mu_{ij}, \tag{19}$$

where E_{ij} equals 1 if student i attends school j and 0 otherwise and where μ_{ij} is an i.i.d. random error. The stage one regression produces estimates of the average school effects, α_j, controlling for differences in student characteristics.

The second stage regression estimates the following relationship:

$$\alpha_j = \lambda + \theta C_j + \psi S_j + \epsilon_j. \tag{20}$$

However, the actual school intercepts are not observed. Instead, the first stage regressions generate fitted values for the school intercepts. Using the first stage predicted values implies an additional error to the second stage regression because

(21) $\hat{\alpha}_j = \alpha_j + \eta_j$

Thus, the second stage becomes a random components model.

(22) $\hat{\alpha}_j = \delta + \lambda C_j + \psi S_j + \epsilon_j + \eta_j.$

Because the sampling variances of the predicted values differ across schools, η_j is heteroskedastic. We assume that the variance of η_j is proportional to the stage 1 sampling variance of α_j, which suggests the use of weighted least squares in the second stage regression. But as Hanushek (1974) points out, this implicitly assumes that the other component of the error term, ϵ_j, has a variance that is either proportional to η_j or zero. In order to allow for the more general random effects specification, we use specialized form of generalized least squares in the second stage. We first estimate equation 22 using ordinary least squares.[18] Next, we regress the square of the residuals on the sampling variance of the school intercepts. Finally, we use the predicted square of the residuals from this auxiliary regression as the weight in the GLS estimation of equation 22.

Table A1 presents the coefficient estimates from the first-stage regression. As expected, both measures of academic attainment are positively related to the pretest score, parental education and family income. Whether school or state dummy variables are included appears to have little impact on the coefficient estimates. There are also significant between-school attainment differences. The hypotheses that the high school dummy variables do not add to the explanatory power of the regressions can be rejected at any conventional level of significance.[19]

A. Basic Aggregation Effects

We turn now to the second stage regressions which contain information about the effects of

[18]Borjas (1987) describes this procedure.

[19]The F Test statistics equal 12.52 (625, 10752 degrees of freedom) for the test score regression and 2.32 (306, 1993 degrees of freedom) for the post-secondary schooling regression.

school resources. Two sets of regressions are computed for each attainment measure. The first set uses estimates of school value added as the dependent variable, while the second set uses estimates of the state value added. Tables 1 and 2 contain the generalized least square regression results for the test score specifications, and Tables 3 and 4 contain the results for the educational attainment specifications. Each table reports the results of regressions which contain both the teacher-pupil ratio and teacher salary variables as well as regressions in which the two school characteristics are entered separately.

Consistent with most prior research, there is no evidence that either increasing the teacher-pupil ratio or raising teacher salaries increases the composite test score in any of the school-level regressions that are reported in Table 1. The coefficients for the teacher-pupil ratio are all very small and insignificantly different from zero, and the teacher salary coefficients consistently have negative signs. The level of college completion at the state level is strongly related to gains in student achievement—substantively suggesting that state differences in educational policies are important and analytically suggesting the strong possibility of upward biases through aggregation that ignores such state differences. Local community variations appear less important for student achievement once state community variation is considered. Though the community education level is positively related to 12th grade test scores, there is little evidence that omitting this variable substantially alters the estimated school characteristic effects in these school-level regressions. In other words, bias introduced by the omission of these community characteristics does not appear to be a major problem at the school level of aggregation (although these may not accurately reflect relevant omitted factors).

The state-level regressions reported in Table 2 produce very different results. Columns 2 and 4 in Table 2 show that the teacher-pupil coefficient estimate is large and closer to conventional statistical significance in specifications that exclude the average community education level. The inclusion of community differences, however, substantially reduces the magnitude of the teacher-pupil

Table 1. Generalized Least Squares Estimates of School-Specific Test Performance (n=627) [t-statistics in parentheses][a]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Local Community Factor									
Local college completion rate	1.05 (1.62)	1.28 (2.04)	1.05 (1.65)	0.60 (0.87)		0.84 (1.25)		0.59 (0.87)	
State Community Factor									
State college completion rate				1.94 (1.83)		1.92 (1.81)		1.93 (1.84)	
School Factors									
Teacher-pupil ratio	0.17 (0.08)	-.01 (-.01)		-.08 (0.04)	.66 (0.31)	-.26 (-0.12)	.56 (0.27)		
Teacher salary	-.67 (- 1.55)		-.66 (- 1.55)	-.68 (- 1.57)	-.82 (-1.97)			-.68 (- 1.58)	-.82 (-1.97)

[a]All models include a dummy variable for being in a rural area, a dummy variable for being in the South, and the average income of families in the school.

Table 2. Generalized Least Squares Estimates of State-Specific Test Performance (n=46) [t-statistics in parentheses]

	(1)	(2)	(3)	(4)	(5)	(6)
Community Factor						
State college completion rate	4.73 (2.18)		4.60 (2.01)		5.07 (2.57)	
School Factors						
Teacher-pupil ratio	2.95 (0.39)	9.93 (1.34)	5.29 (0.66)	12.30 (1.65)		
Teacher salary	-3.04 (-2.02)	-2.83 (-1.82)			-3.16 (-2.15)	-3.28 (-2.12

coefficient estimate and leaves it far below its standard error. The sensitivity of the teacher-pupil ratio coefficient estimate to the inclusion of the community education level is evidence that aggregation worsens problems caused by the exclusion of relevant variables. As in the school-level regressions, the teacher salary coefficient estimates remain negative regardless of whether the community education level is included.

Results for the educational attainment specifications in Tables 3 and 4 provide additional evidence that aggregation exacerbates the problem of omitted variable bias. The school fixed effect terms for years of attainment clearly incorporate many factors outside of the secondary school, because college attendance will be strongly affected by labor market conditions (opportunity costs) in addition to community preferences and to direct costs of schooling. Table 3 demonstrates that local variations in schooling costs (including opportunity costs) are very important, although the addition of state community factors add little. At the same time, the table shows that school-level coefficient estimates for the teacher-pupil ratio and teacher salary are statistically insignificant in all specifications. While some people have suggested that school resources have important indirect effects on students through influencing school completion, these results do not support such a notion. Though the exclusion of the community variables does increase the magnitude and significance of the estimated effects of changes in the teacher-pupil ratio and teacher salaries, the estimates remain statistically insignificant.

Table 4 vividly demonstrates, however, how the apparent picture changes when aggregated to the state level. Aggregation to the state level substantially increases the magnitudes and statistical significance of the estimated relationship between post-secondary schooling and both the teacher-pupil ratio and teacher salary. The teacher-pupil ratio has a statistically significant impact in the expected direction regardless of whether the community variables are included in the regression and its magnitude is approximately five times that found in the school-specific estimates of Table 3.

Table 3. Generalized Least Squares Estimates of School-Specific Educational Attainment Models (n=307) [t-statistics in parentheses]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Local Community Factors									
Local college completion rate	2.11 (3.19)		2.00 (3.10)		2.20 (3.34)		2.46 (3.45)	2.29 (3.45)	2.46 (3.45)
Local unemployment rate	3.62 (2.50)		3.79 (2.70)		3.48 (2.40)		4.18 (2.51)	4.25 (2.59)	4.18 (2.51)
Local wage	-.0030 (-2.68)		-.0032 (-2.92)		-.0031 (-2.80)		-.0009 (-.42)	-.0016 (-.84)	-.0009 (-.42)
State Community Factors									
State College completion rate							-1.50 (-1.17)	-1.68 (-1.17)	-1.50 (-1.17)
State unemployment rate							-2.17 (-.84)	-1.75 (-.68)	-2.17 (-.84)
State wage							-.0019 (-.78)	-.0012 (-.52)	-.0019 (-.78)
In-State Tuition	-.0001 (-.84)		-.0001 (-.97)		-.0001 (-.88)		-.0001 (-.91)	-.0001 (-1.06)	-.0001 (-.91)
School Factors									
Teacher-pupil ratio	2.31 (1.22)	2.86 (1.50)	2.44 (1.30)	3.01 (1.59)			2.39 (1.26)	2.55 (1.35)	
Teacher salary	0.33 (0.80)	0.47 (1.24)			0.38 (0.92)	0.51 (1.34)	0.45 (1.03)		0.51 (1.16)

Table 4. Generalized Least Squares Estimates of State-Specific Educational Attainment t-statistics in parentheses (n=38)

	(1)	(2)	(3)	(4)	(5)	(6)
Community Factors						
State college completion rate	4.89 (2.64)		4.89 (2.65)		5.54 (2.91)	
State unemployment rate	8.51 (2.95)		9.40 (3.50)		7.72 (2.51)	
State wage	-.0041 (-1.71)		-.0049 (-2.29)		-.0063 (-2.36)	
In-State Public Tuition	.00001 (0.47)		.000 (0.25)		.000 (0.24)	
School Factors						
Teacher-pupil ratio	11.66 (2.53)	11.63 (2.50)	11.94 (2.61)	12.33 (2.59)		
Teacher salary	0.94 (0.88)	1.91 (1.91)			1.21 (1.04)	2.21 (2.03

Similarly, the estimated effect of teacher salary is 3-4 times as large in the aggregate as in the school specific regressions. Moreover, excluding the community variables has a dramatic impact on the teacher salary coefficient: The magnitudes of the coefficient and the t-statistic are roughly twice as high in specifications where the community variables are excluded.

These results are broadly consistent with the notion that aggregation exacerbates the problem of omitted variable bias. The exclusion of the community variables usually had a larger impact on the magnitude of the school characteristic coefficients at the state level as opposed to the school level of aggregation. Because aggregation tended to exacerbate the bias from excluding the observed community factors, it is also likely that aggregation increased the bias due to the omission of other relevant factors.

There were substantial differences in the sensitivity of the school characteristic estimates to both the omission of the community variables and to aggregation depending upon which measure of academic performance was used. These differences across the two attainment measures indicate that the excluded determinants of test score growth differ from the factors that impact post-secondary schooling. This might be one reason for why there is little correlation between the test score growth and post-secondary schooling estimates of school value added. The correlation coefficient between these two measures of school value added equals -0.07. At first glance the lack of correlation coefficient suggests that schools choose between a variety of objectives and that schools which are good at raising test scores might not have characteristics which are conducive to post-secondary school attainment. Yet because the estimates of school value added capture both school and nonschool influences, differential impacts of nonschool variables on the different attainment measures might potentially conceal the fact that higher quality schools increase most types of attainment.

B. Measurement Error

While the prior results are consistent with expectations about aggregate measurement error and omitted variables, they are also somewhat consistent with a simple measurement error story for school-level resources. Aggregation moved the school resource and community environment coefficients away from zero, as would be predicted with classical measurement error. Thus the prior results do not conclusively demonstrate that aggregation aggravated omitted variable bias is the sole problem.

A test of the competing hypotheses is based on traditional grouping methods applied in a way so as to not confound grouping with possible omitted state or community factors.[20] Grouping methods, which can be thought of as variants of instrumental variables, are based on aggregating observations within groups where group membership is determined by a variable that is correlated with the true explanatory variable but uncorrelated with the measurement error. Here, however, we add another component to creation of the groups because we also wish to find a grouping variable that is uncorrelated with omitted variables that also enter into the determination of school outcomes. Thus the choice of grouping variables, or instrumental variables, is more difficult than the standard situation where only measurement error is relevant.

The implicit use of state of residence as the grouping variable in the aggregate school resource studies likely violates the criterion that the grouping variable is uncorrelated with relevant omitted factors. Schools are not randomly sorted into states, and a number of state policies likely influence both achievement and school resource decisions. Therefore the finding that aggregation by state inflates coefficients does not by itself permit a distinction between omitted variables bias and measurement error explanations.

A way of distinguishing between the omitted variables and measurement error explanations is to regroup the schools into "pseudo-states" equal in number to the actual states in the test score and

[20]A similar approach of looking for between-jurisdiction differences has been independently proposed in Heckman[1994]. His discussion provides a public choice rationale for the existence of such differences.

post-secondary schooling analyses. One plausible set of grouping variables simply uses the ordering of the school resources themselves—i.e., the ordering by the magnitude of the teacher-pupil ratio or the teacher salary.[21] Grouping across teacher-pupil ratios will lead to consistent estimates if schools are correctly categorized on the basis of true (error-free) teacher-pupil ratios or true salaries, and there are no relevant omitted factors correlated with the true values of school resources. Because errors in measurement can lead to misclassification, an approach that balances efficiency and misclassification concerns is to omit observations at the boundaries between groups, i.e., to produce trimmed group means. Here we divide all schools into the "state" groups and omit two schools at each group boundary, since these schools are the most likely to be wrongly classified.

A second approach orders states according to a state characteristic which is presumed to be correlated with school resource decisions but not directly correlated with achievement or with measurement errors. To implement this, the states are subsequently divided into four "divisions" on the basis of the grouping variable. Within these divisions, schools are randomly assigned to pseudo-states. Schools in each pseudo-state share a similar state attribute, but the random assignment substantially weakens the link between schools and states.[22] Because any single random allocation many give misleading point estimates, this random allocation of schools to pseudo-states is repeated thirty times for each state characteristic grouping variable, and the reported parameter estimates and standard errors are averages over the thirty replications.

---

[21]Early proposals for grouping and instrumental variables concentrated on bivariate models and analyzed the trade-off between bias and efficiency from aggregating to two groups (Wald), three with an omitted center category (Bartlett), and multiple groups (Durbin); see Maddala[1977]. The approach here is an extension of these. For a general consideration of instrumental variables in cross-sections, see White[1982].

[22]A completely random assignment of schools to pseudo-states without first grouping by divisions based on some underlying factor would fully break the link between schools and states. However, this strategy would also eliminate all differences between pseudo-states in the expected values of all variables, thus making aggregate estimates meaningless.

Three state characteristics are used as instruments through creation of pseudo-divisions defined by rank ordering: 1) the state per capita assessed property value; 2) the state poverty rate; and 3) the state percent of workers unionized. The assumption is that controlling for observed family and community differences, each of these grouping variables will be correlated with true school resources but uncorrelated with both measurement error in the school resources and with any state- or local-specific omitted factors that explain student achievement. Of course, the last condition is difficult to verify and, indeed, is suspect when grouping is based on income or wealth—which might well be correlated with influences on school performance.

If aggregation by state inflated the school resource coefficients by reducing measurement error, we would expect that aggregation by pseudo-state should produce estimates quite similar to the aggregate coefficients reported in Tables 2 and 4. On the other hand, if aggregation by state increased the school resource coefficients by exacerbating omitted variables bias, we would expect that the new pseudo-state aggregate estimates would more closely resemble the school-level coefficients in Tables 1 and 3. Each of these five grouping variables substantially weakens the link between schools and states.

Table 5 presents the aggregate test score results; Table 6 presents the aggregate school attainment results. The results are clearest in the case of school attainment, the outcome measure most strongly related to the school resources in the aggregate state-level specifications. The estimated effects of a higher teacher salary and a higher teacher-pupil ratio in Table 6 are quantitatively very close to those in the school-level estimates (Table 3) but very much smaller than those in the state aggregates (Table 4). The only teacher salary coefficients that exceed the school-level estimates are produced by the teacher/pupil ratio and state unionization rate groupings, and even these estimates are much closer to the school level than to the actual state aggregate coefficients. State groupings by the percent unionized and the poverty rate produce teacher/pupil coefficients slightly above those

Table 5. Alternative Grouped Estimates (IV) of Test Performance (n=46 pseudo-states) [t-statistics in parentheses]

	Pseudo-states Grouped by Teacher/Pupil Ratio		Pseudo-states Grouped by Teacher Salary		Random Pseudo-States within "Divisions" Grouped by:					
					per capita assessed value		% unioned manufacturing workers		% poverty	
Community Factor										
College completion rate	-6.26 (2.7)		-2.96 (1.4)		0.07 (0.1)		068 (0.3)		0.82 (0.3)	
School Factors										
Teacher-pupil ratio	-.73 (0.5)	.30 (0.2)	-2.13 (-2.5)	-1.65 (-2.1)	-2.00 (1.2)	-1.95 (-1.3)	-1.24 (-.8)	-1.41 (-1.0)	-.81 (-.5)	-.93 (-.6)
Teacher salary	2.32 (0.7)	-.98 (-0.3)	-9.64 (-.8)	-13.8 (-1.1)	6.69 (0.8)	6.62 (0.9)	-2.00 (0.1)	-1.41 (-.1)	14.7 (1.7)	15.6 (1.8)

Table 6 Alternative Grouped Estimates (IV) of School Attainment
(n=38 pseudo-states) [t-statistics in parentheses]

	Pseudo-states Grouped by Teacher/Pupil Ratio		Pseudo-states Grouped by Teacher Salary		Random Pseudo-States within "Divisions" Grouped by:					
					per capita assessed value		% unioned manufacturing workers		% poverty	
Community Factors										
University Tuition	-.0003 (-.9)		.0006 (2.0)		.0003 (1.1)		.00008 (0.2)		-.00001 (-0.1)	
Wage Rate	-.0064 (-1.7)		-.0096 (3.7)		-.0043 (1.3)		-.0029 (1.0)		-.0041 (1.2)	
Unemployment Rate	8.51 (1.9)		10.16 (2.4)		4.33 (1.0)		6.12 (1.7)		4.6 (1.2)	
College Completion	5.48 (2.9)		3.25 (1.9)		2.35 (1.2)		2.02 (1.1)		1.64 (0.8)	
School Factors										
Teacher Salary	.54 (0.5)	.48 (0.4)	-.20 (0.3)	.35 ().7)	.02 (.04)	.34 (0.3)	.50 (0.4)	.70 (0.7)	.03 (.01)	.44 (0.4)
Teacher-Pupil Ratio	1.67 (0.6)	4.04 (1.5)	-2.72 (0.4)	-2.12 (0.3)	1.28 (0.1)	1.62 (0.2)	5.04 (0.8)	5.45 (0.9)	2.64 (0.3)	2.19 (0.4)

produced by the school-level specifications, but again these estimates lie much closer to the school estimates in Table 3 than to the aggregate state estimates in Table 4.

The test score results in Table 5 also offer very little evidence in favor of the measurement error explanation. Similar to both the school and state-level results, all teacher salary coefficients are negative. The teacher/pupil ratio estimates fluctuate noisily depending upon the grouping variable: three grouping variables produce positive coefficients and two produce negative coefficients. There is no clear pattern in support of the measurement error explanation, particularly in light of the fact that the actual state aggregate coefficients in Table 2 are small and statistically insignificant at any conventional level once the observed community factors are included.

There is no confirmation that the inflated coefficients generated by the actual state-level aggregate specifications were simply the result of correcting measurement errors at the individual school level. Specific measurement error corrections that break the correlation with omitted state-level variables yield resource results that are consistent with the school-level estimates. Thus the overall evidence strongly supports the position that the difference between school-level and aggregate-level results in the apparent importance of school resources arises from omitted variables bias that is aggravated by aggregation as opposed to a simple measurement error problem.

VI. Linkage to Prior Work

These results provide evidence that is crucial in interpreting past analyses of the effects of school resources. In past work positive resource findings are directly associated with aggregation. Our analysis indicates that the discrepancy of findings between state- and school-level analyses can be reconciled by consideration of aggregation biases which inflate the apparent impact of resources at the state level. To see the effects of aggregation bias, we return to the educational production function studies that have been conducted. Past summaries (e.g., Hanushek 1986, 1989) treat all studies symmetrically. Individual level analyses are combined with aggregate analyses. Here we expand on previous reviews and consider the potential effects of data aggregation.

Since the Coleman Report (Coleman *et al.*, 1966) was published, a substantial number of analyses have been directed at uncovering the effects of different resources on student performance. A previous summary of these results (Hanushek 1989) found 187 separate estimates of educational production functions, spread across 38 separate publications.[23] An attempt to uncover both more recent studies and any studies missed in previous searches expanded this to 367 studies spread across 85 publications.

For the purposes here we focus on two aspects of school resources: teacher-pupil ratios and expenditure per student.[24] These resources are central to much of the policy making in education, as discussed previously. Moreover, they are important determinants of the growth in schooling costs (see Hanushek and Rivkin 1994), and they enter centrally into recent debates about the effectiveness of resources. Because data are seldom collected with analyses of student performance in mind, however, many less-than-perfect sources of data have been employed in an effort to tease out the

---

[23]For these purposes, a "study" is a separate estimate of an educational production function. An individual publication may include several studies, pertaining to different grade levels or measurement of outcomes. Alternative specifications of the same basic model were not double counted; nor was publication of the same basic results in different sources. The attempt was to record the information from all published studies that included one of the central measures of resources (either of real resources of pupil-teacher ratios, teacher experience, teacher education, facilities, or administrator characteristics or monetary resources of expenditures per student or salaries), recorded information about the statistical significance of the estimated relationships, and included some measure of family and nonschool inputs. The 1989 summary missed a few estimates that were available prior to mid-1988, and these have been incorporated in the update.

[24]The best way to tabulate the results across different studies has been the subject of lively debate; see Hedges *et al.* (1994) and Hanushek (1994). Under certain circumstances (which do not appear applicable here), the simple tabulations presented below may lead to misleading conclusions. The requirements to do anything beyond these simple counts are, however, unlikely to be met in the studies reviewed.

Technical issues aside, there is complete agreement in the debate that resources do not consistently matter regardless of how they are employed. Further, no consensus exists on when resources are likely to matter and when not, thus precluding any efficacious use of pure resource policies.

effects of the key school resources. These analyses often involve merging data from different sources, and this merging frequently implies that specific resources cannot be directly matched with the students receiving them. We provide information on how aggregation of school resources corresponds to the results that are obtained.

Table 7 displays the overall results for teacher-pupil ratios. Conventional policy arguments suggest that the teacher-pupil ratio should be positively related to student performance. This first thing to note from the table is that the combined estimates over the 268 separate investigations give no reason to believe that smaller classes are related to higher student performance.[25] Fifteen percent (39/268) of the estimates are statistically significant and positive, while thirteen percent (36/268) are also statistically significant and negative. Disregarding statistical significance does not change the picture. The second thing to note is that the positive estimates (both statistically significant and total) come disproportionately from studies which aggregate the school data to the state or even district level. While relatively few studies include state differences (11), almost two-thirds find positive and statistically significant effects of teacher-pupil ratios. At the district, 21 percent find positive and statistically significant results and 39 percent find positive but statistically significant effects. When the resources are measured closer to the students, at the classroom or school level, any hint of disproportionate impact of smaller classes goes away. This story reinforces the previous discussion of aggregation that includes systematic measurement error and, quite possibly, errors correlated with omitted characteristics of students.

Table 8 repeats the same consideration of aggregation for estimates of expenditure on performance. For the 160 estimates of school expenditure effects, 85 come from relating spending in

---

[25]Not all studies contained information on each specific resource. Of the 363 studies looking at any of the identified resources, 264 analyzed either teacher-pupil ratios, pupil-teacher ratios, or class size. (All analyses of pupil-teacher ratios are put in terms of teacher-pupil ratios by reversing the signs).

Table 7. Percentage Distribution of Estimated Effect of Teacher-Pupil Ratio on Student Performance (268 estimates)

Level of Aggregation of Resources	number of estimates	Statistically significant		Statistically insignificant		Insignificant, unknown sign
		+	-	+	-	
Total	268	15%	14%	26%	25%	21%
Classroom	77	12	8	18	26	36
School	119	9	18	25	28	20
District	56	21	16	39	20	4
County	5	0	0	40	40	20
State	11	64	0	27	9	0

Note: Rows may not add to 100 because of rounding.

individual schools to the performance of the students in those schools while the remaining studies are aggregated to the district level or above. At the school level, fifteen percent find positive and significant effects of spending on performance, while 7 percent incredibly find a statistically significant *negative* effect of spending. But this picture again changes when the resources are no longer related to specific students through aggregation to higher levels. At the state level, 64 percent of the estimates suggest that higher spending is associated with student performance in a statistically significant manner. Only one of the 28 estimates state-level resources is negative. Again, this pattern of results—where findings of positive resource effects come disproportionately from highly aggregated measures—is precisely what was found in our own analysis of school resource effects.[26]

Some previous speculation about why the production function results appear to differ from the results for earnings have centered on the measurement of outcomes. Table 9 divides the results into those related to test score measurement of outcomes and all other measures. Two-thirds of the studies consider test score measures with the remainder evaluating school attainment, dropout behavior, earnings, and other measures. This table makes it clear, however, that aggregation produces disproportionate numbers of estimates indicating that expenditure affects outcomes, *regardless of how outcomes are measured.* Seventy-five percent of the state-level estimates of expenditure on test performance and sixty percent of the state-level estimates of expenditure on nontest performance are positive and statistically significant, compared to only 14 percent of the estimates of classroom and school expenditure.[27]

The aggregation explanation that has been developed in this paper clearly has direct bearing

---

[26]The separate analysis of resource effects by Hedges *et al.*[1994] concentrates on expenditure studies without regard to the quality of the underlying analysis. Of the significant positive expenditure results, 43 percent come from state-level aggregate estimates and another 29 percent come from aggregation to district levels.

[27]Betts (1995) finds a very similar pattern for wage and school attainment studies.

Table 8. Percentage Distribution of Estimated Effect of Expenditure per Pupil on Student Performance (160 estimates)

Level of Aggregation of Resources	number of estimates	Statistically significant		Statistically insignificant		Insignificant, unknown sign
		+	-	+	-	
Total	160	26%	7%	34%	19%	13%
Classroom	4	0	0	0	0	100
School	81	15	7	36	23	19
District	42	29	10	36	26	19
County	5	0	0	40	20	40
State	28	64	4	32	0	0

Note: Rows may not add to 100 because of rounding.

Table 9. Percentage Distribution of Estimated Effect of Expenditure per Pupil on Student Performance by Outcome Measure and Aggregation of Resource Effects (160 estimates)

Measure of Outcome and Aggregation of Resources	number of estimates	Statistically significant		Statistically insignificant		Insignificant, unknown sign
		+	-	+	-	
A. Test Score Outcomes[a]						
Total	107	24%	9%	28%	22%	17%
Classroom	4	0	0	0	0	100
School	56	18	9	29	21	23
District	37	27	11	35	27	0
County	2	0	0	0	50	50
State	8	75	13	13	0	0
B. Other (Nontest) Outcomes[b]						
Total	53	30%	2%	47%	15%	6%
School	25	8	4	52	28	8
District	5	40	0	40	20	0
County	3	0	0	67	0	33
State	20	60	0	40	0	0

Note: Rows may not add to 100 because of rounding.

a. All studies measure student performance by some form of standardized test score.

b. All studies employ some outcome measure (such as income or school attainment) other than a standardized test score.

on previous analyses of resource effects. Aggregate analyses of student performance, particularly at the state level, typically have very crude measures of school and family factors. They never employ value-added models. In short, they are subject to extensive specification problems. This situation is exactly where aggregation bias is most important, and the review of past analyses shows a pattern of results that is entirely consistent with the presence of substantial specification error in aggregate studies of school resource effects.

VII. CONCLUSION

School funding levels are determined by many of the same factors that directly influence student attainment. Therefore, it is quite difficult to isolate the effects of school expenditure on student performance. Variations in school expenditure across time or place are likely to be accompanied by differences in other related variables. It is generally not possible to account for the impacts of all of these related variables in empirical analyses; consequently, omitted variable bias is a common flaw in education production function estimates. Under a broad range of empirical specifications, omitted factors might be expected to bias estimated school resource parameters upwards.

The results in this paper provide evidence that problems of omitted variable bias tend to increase along with the level of aggregation. This suggests that analyses that use more aggregated data tend to systematically over-estimate the influence of school expenditure related characteristics on student attainment. Studies which contain more information about community characteristics and which use less aggregated data are likely to produce more reliable estimates of the true impact of school expenditure on attainment. Investigation of the competing hypothesis that aggregation is beneficial because it reduces biases from measurement error provides no support for the alternative.

Our findings are consistent with the view that increases in school expenditure used to reduce the teacher-pupil ratio and raise teacher salaries have had little impact on student attainment. Therefore, further reductions in the teacher-pupil ratio or further increases in teacher salary are unlikely to generate improvements in the performance of students who attend United States public elementary and secondary schools. Additionally, nothing suggests that resources are more important in assuring the students attain more schooling than they are in determining the pattern of cognitive achievement.

Significant differences in school quality exist. These differences, however, are not systematically related to school resources—a finding that introduces added complexity into the development of educational policies. Policies aimed at altering the incentive structure and, thus, the ways in which resources are used appear much more likely to succeed than policies aimed simply at adding more resources to schools (Hanushek with others, 1994).

References

Aitkin, M., and N. Longford. "Statistical modelling issues in school effectiveness studies." *Journal of the Royal Statistical Society A* 149, pt. 1 (1986): 1-26.

Betts, Julian R. "Does school quality matter? Evidence from the National Longitudinal Survey of Youth." *Review of Economics and Statistics* (forthcoming).

Bishop, John. "The impact of academic competencies of wages, unemployment, and job performance." *Carnegie-Rochester Conference Series on Public Policy* 37 (December 1992): 127-94.

Borjas, George J. "Self-selection and the earnings of immigrants." *American Economic Review* 77, no. 4 (September 1987): 531-53.

Bryk, Anthony S., and Stephen W. Raudenbush. *Hierarchical Linear Models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications, 1992.

Card, David, and Alan B. Krueger. "Does school quality matter? Returns to education and the characteristics of public schools in the United States." *Journal of Political Economy* 100 (February 1992): 1-40.

---. "School quality and black-white relative earnings: A direct assessment." *Quarterly Journal of Economics* 107, no. 1 (February 1992): 151-200.

---. "The Economic Return to School Quality: A Partial Survey". Working Paper #334, Industrial Relations Section, Princeton University, October 1994.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966.

Ferguson, Ronald. "Paying for public education: New evidence on how and why money matters." *Harvard Journal on Legislation* 28, no. 2 (Summer 1991): 465-98.

Hanushek, Eric A. "Efficient estimators for regressing regression coefficients." *The American Statistician* 28, no. 2 (May 1974): 66-67.

---. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources* 14, no. 3 (Summer 1979): 351-88.

---. "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature* 24, no. 3 (September 1986): 1141-77.

---. "The impact of differential expenditures on school performance." *Educational Researcher* 18, no. 4 (May 1989): 45-51.

---. "Money might matter somewhere: A response to Hedges, Laine, and Greenwald." *Educational*

Researcher 23, no. 4 (May 1994): 5-8.

Hanushek, Eric A., and Steven G. Rivkin. "Understanding the 20th century explosion in U.S. school costs". Rochester Center for Economic Research, Working Paper 388, August 1994.

Hanushek, Eric A., and Lori Taylor. "Alternative assessments of the performance of schools: Measurement of state variations in achievement." *Journal of Human Resources* 25, no. 2 (Spring 1990): 179-201.

Hanushek, Eric A., and with others. *Making schools work: Improving performance and controlling costs.* Washington, DC: Brookings Institution, 1994.

Heckman, James J. "Aggregation Bias and Schooling Quality". University of Chicago, December 1994 (mimeo).

Heckman, James J., Anne S. Layne-Farrar, and Petra E. Todd. "Does Measured School Quality Really Matter?". University of Chicago (mimeo), September 1994.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald. "Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes." *Educational Researcher* 23, no. 3 (April 1994): 5-14.

Johnson, George E., and Frank P. Stafford. "Social returns to quantity and quality of schooling." *Journal of Human Resources* 8, no. 2 (Spring 1973): 139-55.

Maddala, G. S. *Econometrics.* New York: McGraw-Hill, 1977.

Rivkin, Steven G. "Schooling and employment in the 1980s: Who succeeds?" Ph.D. Dissertation, University of California, Los Angeles, 1991.

Speakman, Robert, and Finis Welch. "Does school quality matter? --A Reassessment". Texas A&M University (mimeo), January 1995.

Summers, Anita, and Barbara Wolfe. "Do schools make a difference?" *American Economic Review* 67, no. 4 (September 1977).

Swamy, P. A. V. B. "Efficient inference in a random coefficient regression model." *Econometrica* 38, no. 2 (March 1970): 311-23.

Theil, Henri. *Principles of Econometrics.* New York: John Wiley & Sons, 1971.

Tiebout, Charles M. "A pure theory of local expenditures." *Journal of Political Economy* 64 (October 1956): 416-24.

White, Halbert. "Insturmental variables regression with independent observations." *Econometrica* 50, no. 2 (March 1982): 483-99.

Appendix A

Table A1. Within Group Parameter Estimates for Test Score Production and School Attainment [t-statistics in parentheses]

Variable	12th Grade Test Score		Years of Schooling Completed	
	School Fixed Effects	State Fixed Effects	School Fixed Effects	State Fixed Effects
10th Grade Test Score	.86 (134.9)	.86 (139.7)	.17 (19.8)	.16 (20.8)
Father's Education	.10 (7.10)	.11 (8.58)	.10 (6.16)	.11 (7.06)
Father's Education Unknown	.91 (4.34)	1.08 (5.30)	1.14 (4.50)	1.18 (4.94)
Mother's Education	.08 (5.31)	.09 (5.83)	.11 (5.40)	.11 (6.12)
Mother's Education Unknown	.56 (2.17)	.72 (2.85)	1.16 (2.87)	1.08 (2.84)
Fam.Income between $12,000 and $20,000			-.08 (-0.88)	-.12 (-1.40)
Fam.Income over $20,000			.28 (2.82)	.27 (2.88)
Female	-.33 (-5.89)	-.33 (-6.06)	.16 (2.42)	.16 (2.69)
Black	-.58 (-5.11)	-.59 (-6.65)	.39 (3.46)	.47 (5.57)
observations	11,386	11,386	2,309	2,309