Educational Considerations

Volume 33 | Number 2 | Article 9

4-1-2006

How Much Do Schools and Districts Matter? A Production Function Approach to School Accountability

James L. Phelps

Michael F. Addonizio
Wayne State University

Recommended Citation

Phelps, James L. and Addonizio, Michael F. (2006) "How Much Do Schools and Districts Matter? A Production Function Approach to School Accountability," *Educational Considerations*: Vol. 33: No. 2. https://doi.org/10.4148/0146-9282.1227

Commentary

How Much Do Schools and Districts Matter? A Production Function Approach to School Accountability[1]

James L. Phelps and Michael F. Addonizio

In 1989, President George H.W. Bush convened a first-ever education summit in Charlottesville, Virginia, with the governors of the states and territories. At this unprecedented summit, political leaders at the federal and state levels agreed to establish national education goals for America's elementary and secondary schools. This national focus on education goals culminated in the 1994 passage by the U.S. Congress of legislation declaring that "all students can learn and achieve to high standards and must realize their potential if the United States is to prosper."[2]

The 1994 reauthorization of the Elementary and Secondary Education Act of 1965 established Adequate Yearly Progress (AYP) as the accountability measure for Title I schools and districts. Each state was required to develop its own formula based on state assessments in at least reading and mathematics. States varied considerably in their approaches to AYP, with the result that Title I schools and districts were held to different standards across the states. The 2001 reauthorization of Title I, the No Child Left Behind (NCLB) Act, sought to bring more uniformity to the states' AYP requirements. This legislation also substantially changed how AYP results are used, focusing on low-performing Title I schools and offering educational alternatives to their students.

Under NCLB, schools and districts that fail to make AYP for two consecutive years are required to undergo a set of reforms and sanctions designed to improve student achievement. The scope of these reforms and sanctions widens as a school continues to fall short of AYP requirements to include the offer of transfer to children who wish to leave the school, the provision of supplementary educational services outside of the normal school day, the replacement of school staff, and the conversion of the school to charter status. New requirements and sanctions are also imposed on school districts that fail to make AYP, including the withholding of funds by the state, replacement of district staff, and the abolition or reorganization of the district.

James L. Phelps is an educational consultant and former Deputy Superintendent of Public Instruction in Michigan. Michael F. Addonizio is Professor of Education Economics and Policy in the Department of Educational Administration at Wayne State University.

In response to these federal mandates, the states have adopted or refined outcome goals for schools and students and placed new emphasis on school accountability for student achievement. States are now devoting considerable attention to the task of distinguishing between effective and ineffective schools. Much less attention, however, has been given to the task of identifying effective school districts despite the considerable emphasis placed on district as well as school performance by NCLB. This joint focus on school and district performance raises the question: How do district policies, leadership, and support services influence the quality of teaching and learning in public schools? This article uses a rich longitudinal school-level database to estimate a model of student achievement and analyzes the residuals in the model to obtain estimates of the contributions of unobserved school and school district characteristics to student performance. The second section of the paper reviews recent research on current approaches to determining school quality on the basis of student achievement test scores. Section three presents an alternative approach to assessing school and district effectiveness using an educational production function. A production function model is specified in the fourth section, and empirical results are presented in section five. A summary of findings in presented in the final section, along with implications for state and federal policy and programs regarding school accountability.

The School Accountability Movement in the United States

By 2000, 48 states had implemented standardized testing, including tests in mathematics and English/language arts or reading, as an integral part of statewide school accountability programs.[3] The other two states– Iowa and Nebraska– require their districts to test students in specified grades or grade spans. Other elements of this educational reform movement include standards for student and school performance, teacher competency testing, and school accreditation programs which often include recognition and rewards for high performance and assistance and, in many states, sanctions for poor performance. These elements of performance-based school reform were emphasized in the 1994 reauthorization of the federal Title I program as well as many state reform initiatives.[4] Thirty-three states have set performance goals for schools or districts and use the results of state assessments to hold these units accountable for meeting these outcome goals.

These performance-based reforms respond to school critics who have cited the lack of sufficient incentives for schools to improve the quality of teaching and learning;[5] that is, these reforms seek to provide such incentives by developing measures of student achievement and school quality and tying financial and other rewards to those measures. Such rewards may take the form of school-level financial bonuses or statewide public recognition for excellence. Of course, such recognition may also translate directly into school district financial rewards in the context of inter-district school choice programs, where high performing schools attract residents of neighboring districts.

The creation of effective incentives, however, depends crucially on the valid and reliable measurement of school performance. Increasingly, policymakers agree that school performance should be measured in terms of the academic performance of the students in the school. The most prevalent measure of school performance is average test score levels among students in a particular grade. Test score levels are often reported in terms of the percentage of students at a school scoring in particular ranges, such as the proportion failing, proficient or advanced. It is also understood, however, that any measure of school performance that is based on student performance should account for differences in student characteristics (particularly socioeconomic) and school resources. School level scores fail to do this.

A number of states base school building accountability systems on comparisons of student achievement test scores from one year to the next; that is, states compare the achievement of students at selected grades in a given year with the achievement of students from the previous year at the same grade in the same school. Such change scores are clearly superior to single-year level scores as an indicator of school quality because they provide a control for the different endowments and starting points of the students. However, as Linn and Haug observed, such comparisons of student performance at a grade level across years assume that student characteristics that affect achievement levels are relatively stable from year to year for students attending a given school. This assumption, while reasonable for most schools, is invalid for schools in neighborhoods undergoing rapid demographic and economic change.

Other important determinants of student performance may change as well, confounding the interpretation of change scores as indicators of school quality. Kane and Staiger have shown that a substantial portion of the variability in schools' change scores is due to non-persistent factors such as an extended leave of a teacher, a teacher strike, or changes in rules regarding test taking, that affect test scores in one year but not the next. Examining fourth-grade math scores from North Carolina, Kane and Staiger estimated that only about one fourth of the variance in school change scores was attributable to persistent factors associated with the school.[6] Linn and Haug, using data from Colorado's fourth grade reading assessments, computed two change scores (change in percent of students proficient or advanced) for each of 734 schools, one from 1997 to 1999 and one from 1998 to 2000, and found a correlation of -.03 between them.[7] The authors concluded that "there is a complete lack of stability in the two-year change scores. Knowing the magnitude of the gain or loss in percent proficient or advanced from 1997 to 1999 tells you essentially nothing about the change from 1998 to 2000".[8]

School change scores, then, are flawed indicators of real change in school quality. They are influenced not only by measurement error, but also by changes in the student population and in the teaching staff, making their interpretation as indicators of effective or failing schools problematic. A third approach to inferring school performance on the basis of student test scores uses the average gain in test performance between the end of one grade and the end of the next grade. This cohort gain or "value-added" approach, which compares the performance of this year's fourth-grade students with their own performance in third grade, requires states to invest in data systems that link test scores of individual students across years. This approach is used in a handful of states including Arizona, North Carolina and Tennessee. Test score changes and gains are generally viewed as less biased than level scores as a means of comparing schools serving different student populations. They are, however, more difficult to measure reliably.[9] Moreover, school gain scores have been found to be positively correlated with the proportion of white and nonpoor students, thus confounding their interpretation as measures of school effectiveness.[10]

Further, the assessment of *district* quality, also required by NCLB, is similarly confounded. Indeed, even assuming away these problems in interpreting school change scores, what is to be inferred about the performance of a school district in which most but not all schools show improvement over a change cycle? The quality of district leadership, policies, communications, and school supports is difficult to discern through the use of school change scores. A more valid and reliable assessment of school and district effectiveness requires more information. Such an assessment is outlined in the next section.

Assessing School Performance: A Production Function Approach

To accurately estimate the "quality" of a school, that is, the school's contribution to student learning, one must account for the relative contributions of children's families, communities, peers, and school resource levels to student learning. Put another way, one should not confound school quality with other fundamental determinants of student performance, particularly when assessments of school quality trigger school rewards and sanctions.

One approach to developing school performance measures relies upon the concept of production efficiency and techniques for measuring such efficiency. This approach utilizes the economist's notion of a production function.[11] Production models have three parts: the outcomes sought, the necessary ingredients or inputs, and the process that transforms inputs into outcomes. These three parts are linked together by a production function. This production function reveals the maximum amount of outcome possible for various combinations of inputs. If the supply levels of the various inputs are known and the production function is also known, the maximum level of outcome (i.e., production) can be determined. Anything short of maximum attainable output indicates technical inefficiency.

A second dimension to production efficiency involves input costs. Consider, for example, two alternative educational programs that utilize different input combinations to produce the same outcome, say, the ability to do mathematics at a specified level. While both programs involve teachers' time, textbooks, worksheets, and the like, one may emphasize student-teacher contact while the other relies heavily on computer-assisted self-instruction. Assuming that each program makes the best possible use of each set of inputs—that is, each program is technically efficient—the less-costly input combination is preferred on allocative efficiency grounds. Put another way, production efficiency requires both technical and allocative efficiency.

Analysis of educational production is notoriously difficult. First of all, education is characterized by multiple outcomes. Schools are charged with developing cognitive skills in a number of subject areas, as well as affective traits, promoting democratic values and furthering other social goals. Some outcomes are jointly produced, (e.g., cognitive skills and self-esteem), while others may be mutually exclusive (e.g., higher academic standards and higher graduation rates). Second, even if it were possible to separate outcomes, there is no obvious way to assign *a priori* weights to reflect the relative value of each. Consequently, there is no unambiguous way to sum the various production activities into a single outcome measure.

Researchers have responded to the problem of joint production of educational outcomes by focusing on one relatively easy to measure outcome and assuming the other outcomes are produced as by-products. This approach emphasizes student learning and the testing of cognitive skills in key subjects such as reading and mathematics and simplifies the analysis of schools' production efficiency considerably. This approach also enjoys a wide political consensus across states and school districts and provides the basis of school accountability systems in virtually every state.

At the same time, there is growing recognition that any measure of school performance (i.e., production efficiency) must account for inputs that are beyond the control of those in the school, particularly student and community characteristics and school resource levels. The production function approach allows us to estimate the marginal educational contributions of identified educational inputs, both "controllable" and "uncontrollable," and to identify those controllable inputs with positive marginal products. These estimated products can then be compared with corresponding input costs to improve allocative efficiency. The production function approach can also be used to identify school districts and schools that consistently produce levels of student achievement that exceed (or fall short of) levels predicted by the identified inputs. These consistently higher or lower than predicted performance levels can be attributed to practices or characteristics of the schools and districts that are not identified in the production model. Levin contends that these unmeasured and often unobserved practices and characteristics can be very important to school performance.[12] Levin builds upon Leibenstein's seminal article on x-efficiency in which incentives and other generally unmeasured organizational attributes of the firm are viewed as making a greater contribution to firm efficiency than the marginal reallocation of inputs.[13]

The Production Function Model

Hanushek proposed a framework for an educational production function that distinguishes among family background, peer, and school inputs.[14] A simplified version of this production function is of the following form:

$$A = f(B, P, S)$$

where A represents all outcomes, B represents all family background inputs, P represents all peer inputs, S represents all school inputs and $f(.)$ is the function or production process that transforms the inputs into outcomes. Citing the absence of a well-developed theory of learning to guide the estimation of this model, Monk observed that researchers generally choose input measures on intuitive grounds because they are important for policymaking, or because the data are readily available.[15] All three factors have influenced our selection of input variables and outcome measures. Following Hanushek's framework, we estimated the following model:

$$A = b_0 + b_1 SES + b_2 RLADMIN + b_3 RLSUPPORT + b_4 RLINSTRUCT + b_5 RNLINSTRUCT + b_6 Tch\_yrs + b_7 Tch\_sal + b_8 Tch\_age + b_9 PCT\_mas + b_{10} Tot\_adm + b_{11} TotalPP + \in$$

where A is measured student achievement in reading and mathematics for grades three and five (READ3, READ5, MATH3, and MATH5);[16]

SES is an index of family and peer inputs;
RLADMIN is licensed administrators per 1,000 students;
RLSUPPORT is licensed support staff per 1,000 students;
RLINSTRUCT is licensed instructional staff per 1,000 students;
RNLINSTRUCT is non-licensed instructional staff per 1,000 students;
Tch_yrs is teachers' average years of teaching experience;
Tch_sal is average teacher salary;
Tch_age is average teacher age;
Pct_mas is percent of teachers with a masters degree;
Tot_adm is total average daily attendance; and
Total PP is total operating expenditures per pupil.
$\in$ is an error term

A pooled time series of school-level data was obtained from the Minnesota Department of Children, Families and Learning for all elementary schools in Minnesota for four years, 1998 through 2001. All schools reporting data to the state were included in the study. Reporting of school-level data was optional in 1998, and 506 schools participated that year. Participation rose to 671 schools in 1999, 690 in 2000, and 694 in 2001, thereby including all elementary schools in the state. Data for all variables were reported by participating schools, with the exception of teachers' average years of teaching experience for 1998. For that variable, schools' 1999 data were also used in the 1998 data base. Achievement data consisted of building average scores on statewide assessments of reading and mathematics in grades three and five for each of the four years.[17] The SES index is a weighted average of five component variables: (1) percent of children in the school who are eligible for free or reduced price lunch; (2) percent of children who are minority; (3) percent of children who are in special education; (4) reported disciplinary incidents as a percent of building enrollment; and (5) intra-district mobility rate.[18]

Results

Our model was estimated by weighted least squares (WLS), with each observation (school) weighted by the square root of the school's average daily membership.[19] Separate stepwise regressions were run for each of the outcome measures (READ3, READ5, MATH3, and MATH5) for each of the four years. Descriptive statistics are presented in Table 1, and regression results are given in Table 2. The F-value tolerances for entry and removal of independent variables in the stepwise regression routine were set at .20 and .25, respectively.[20]

The cross-section regressions reveal the importance of the SES index in explaining variation in student test scores. SES was statistically significant at the .01 level in each equation, with an R^2 ranging from .487 to .740. Thus, the index explained anywhere from about half to three-quarters of the variation in test scores. The SES effect was more pronounced in reading, but was also substantially greater with grade 5 math results than with grade 3 math. Clearly, such powerful SES effects would render school level scores meaningless as indicators of school quality.

The most influential school variables were teacher characteristics. Teacher salary was statistically significant at the .01 level in five equations and at the .05 level in two others. All coefficients were positive. The effect was greatest for grade 5 math performance, with significance in every year. Coefficients on teacher age were positive and statistically significant in four equations, all for reading (third and fifth grades for both 2000 and 2001). Finally, and somewhat surprisingly, the teacher experience coefficient was negative and statistically significant in four equations – 1998 READ5, 1999 MATH5, 2000 MATH3, and 2000 MATH5. Taken together, these findings suggest that higher salary schedules have succeeded in recruiting and retaining more skilled teachers, all else equal. Beyond that, the inconsistent findings regarding teacher age and experience are open to varying interpretations and remain ambiguous.

Table 1
Descriptive Statistics

Variable	1998		1999		2000		2001	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
READ3	1401.34	74.64	1419.19	79.23	1451.76	83.78	1479.28	79.50
READ5	1407.73	84.51	1442.69	85.79	1483.16	93.76	1536.52	93.97
MATH3	1389.79	93.16	1451.66	97.09	1470.15	95.39	1489.12	91.80
MATH5	1384.64	84.67	1409.77	86.30	1461.05	91.08	1484.99	87.85
SES	365.05	258.43	367.60	270.79	346.6	231.14	344.42	232.39
RLADMIN	2.65	1.78	2.78	1.97	2.81	1.75	2.58	1.78
RLSUPPORT	3.47	2.38	3.35	2.43	3.73	2.13	3.24	2.33
RLINSTRUCT	63.49	13.36	65.26	13.73	67.43	14.62	68.05	13.46
RNLINSTRUCT	18.46	11.14	19.63	11.52	21.07	12.24	21.95	20.60
Tch_yrs	14.91	3.53	14.91	3.53	15.00	3.49	14.35	3.39
Tch_sal	41374.07	5223.84	40569.14	5326.03	40413.01	5322.70	42876.11	5430.58
Tch_age	42.29	3.45	42.22	3.55	41.74	3.30	41.77	3.38
PCT_mas	40.29	20.05	37.35	20.63	37.60	19.98	37.74	20.63
Tot_adm	493	210.86	461.97	213.64	458.66	211.42	452.84	212.40
Total PP	4859.78	3922.34	4818.02	3812.75	5188.81	2522.82	5213.26	2527.40
N	506		671		690		694	

Licensed instructional staff per 1,000 students was also found to positively influence student performance. The coefficient on RLINSTRUCT was positive and statistically significant in six equations. Interestingly, five of the six involved math achievement. No other resource measures were found to be statistically significant in more than two of the sixteen equations. In summary, teaching staff seemed important to student achievement, with investment in well-paid teachers and higher teacher-pupil ratios yielding a positive marginal product.

These regression models, as one might expect, are subject to considerable collinearity among the independent variables. This multicollinearity increases the variance of our coefficient estimates (while our large sample sizes decrease this variance), and the stepwise regression procedure may overestimate the influence of SES on student achievement; that is, the estimated marginal effect of an independent variable on student achievement will depend, in part, on the order in which it is entered into the estimated model. Consequently, the models were re-estimated with the order of entry of the independent variables controlled by the researchers.[21]

Specifically, in view of the substantial evidence confirming the importance of family and peer effects on student achievement and the mixed findings regarding school effects, each outcome variable was regressed against the SES index only and then regressed over the measures of school resources. Finally, each outcome variable was regressed over both the SES and school resource variables. The differences in the coefficient of determination, or R^2, were interpreted as upper and lower bounds for the estimated influence of each set of independent variables on student achievement given the multicollinearity among the variables. These changes in the coefficients of determination are presented in Table 3.

The R^2 change associated with the SES index was subtracted from the model's total R^2 to derive a lower bound for the effect of the school resource measures on student achievement. The SES index was found to explain between 45 and 71 percent of the variance in student achievement. When the SES index entered the regression first and the school resource measures second (collectively designated SCHOOL 2 in the table), the SES index is attributed with virtually all the power to explain variance in student achievement and negligible influence is attributed to school resources. In contrast, when the school resource measures are entered first (SCHOOL 1), their assigned explanatory power is about one half the explanatory power of SES, indicating considerable collinearity among the SES and school resource measures. This collinearity makes it difficult to disentangle and estimate their separate influences on student achievement. Moreover, any unobserved school and district effects, as opposed to the observed effects of the school resource variables, are concealed in the error terms of the regressions. The magnitude of the unexplained variance is (1-TOTAL), labeled E/U for "error/unexplained" variance.

Analysis and Discussion of Residuals

The residuals in these regressions consist of school and district fixed effects, both unobserved, along with random error. In order to estimate the magnitude of these unobserved but nonrandom effects, the residuals were examined for each observation (school) to

https://newprairiepress.org/edconsiderations/vol33/iss2/9
DOI: 10.4148/0146-9282.1227

Table 2
Weighted Least Squares (WLS) Regression Results, 1998-2001

1998	**Dep = MATH5**					
Step	***Predictor Entered***	***Beta***	***R-sq Change***	***Cumulative R-sq***	***Adj. R-sq***	***SEE***
1	SES	-0.845**	0.621	0.621	0.621	257.4244
2	TCH_SAL	0.089**	0.004	0.625	0.624	256.3583
3	RLINSTRUCT	0.108**	0.005	0.63	0.629	254.7226
4	RLADMIN	-0.037	0.001	0.631	0.629	254.585
	Dep = MATH3					
1	SES	-0.737**	0.516	0.516	0.515	300.831
2	RNLINST	0.083*	0.002	0.518	0.516	300.361
3	TCH_SAL	0.093	0.004	0.521	0.519	299.482
4	RLADMIN	-0.049	0.002	0.523	0.52	299.203
5	PCT_MAS	-0.053	0.002	0.524	0.521	298.908
	Dep = READ5					
1	SES	-0.88**	0.715	0.715	0.714	212.774
2	TCH_SAL	0.123*	0.009	0.724	0.723	209.555
3	TCH_YRS	-0.086*	0.004	0.728	0.726	208.293
4	TOT_ADM	-0.055*	0.002	0.73	0.728	207.788
5	TOTAL PP	-0.043	0.002	0.732	0.729	207.252
6	PCT_MAS	0.041	0.001	0.733	0.73	207.06
7	RNLINST	0.033	0.001	0.734	0.73	206.925
	Dep = READ3					
1	SES	-0.909**	0.74	0.74	0.739	179.518
2	TCH_SAL	0.103**	0.002	0.742	0.741	178.865
3	RLINSTRUCT	0.06	0.003	0.745	0.743	178.143
4	TOTAL PP	-0.042*	0.001	0.746	0.744	177.877
5	TCH_YRS	-0.052	0.001	0.747	0.745	177.66
6	TOT_ADM	-0.04	0.001	0.748	0.745	177.387
1999	**Dep = MATH5**					
1	SES	-0.812**	0.626	0.626	0.626	424.792
2	TCH_SAL	0.126**	0.008	0.634	0.633	240.519
3	TCH_YRS	-0.081*	0.004	0.638	0.636	239.428
4	RLSUPPORT	-0.042	0.001	0.639	0.637	239.172
	Dep = MATH3					
1	SES	-0.779**	0.537	0.537	0.536	299.629
2	TCH_SAL	0.077	0.002	0.54	0.538	299.064
3	RLINSTRUCT	0.072*	0.003	0.543	0.541	298.269

Table 2 Continued
Weighted Least Squares (WLS) Regression Results, 1998-2001

1999	Dep = READ5					
Step	Predictor Entered	Beta	R-sq Change	Cumulative R-sq	Adj. R-sq	SEE
1	SES	-0.788**	0.687	0.687	0.687	221.732
2	PCT_MAS	0.068**	0.006	0.693	0.692	219.848
3	TOTAL PP	-0.057	0.003	0.695	0.694	219.102
4	RNLINST	-0.039	0.001	0.697	0.695	218.779
	Dep = READ3					
1	SES	-0.835**	0.623	0.623	0.623	222.226
2	TCH_SAL	0.135**	0.006	0.629	0.628	220.559
3	RLINSTRUCT	0.057	0.002	0.632	0.63	220.069
4	TOTAL PP	-0.048	0.002	0.633	0.631	219.742
5	TOTAL ADM	-0.046	0.001	0.634	0.631	219.633
6	TCH YRS	-0.044	0.001	0.636	0.632	219.376
2000	Dep = MATH5					
1	SES	-0.903**	0.703	0.703	0.702	217.937
2	PCT_MAS	0.089	0.008	0.711	0.709	215.316
3	TCH_YRS	-0.11**	0.003	0.713	0.711	214.554
4	TCH_SAL	0.143**	0.005	0.718	0.716	212.831
5	(PCT_MAS deleted)	---	-0.001	0.718	0.716	212.841
6	RNLINST	0.057*	0.003	0.72	0.718	212.054
	Dep = MATH3					
1	SES	-0.854**	0.641	0.641	0.641	262.585
2	TCH_YRS	-0.073*	0.002	0.643	0.642	262.172
3	TCH_SAL	0.097	0.004	0.647	0.645	261.017
4	RLIINSTRUCT	0.082*	0.003	0.65	0.647	260.087
5	TOTAL PP	-0.06	0.003	0.653	0.65	259.212
	Dep = READ5					
1	SES	-.829**	0.668	0.668	0.667	248.37
2	TCH_AGE	0.053*	0.003	0.67	0.669	247.57
3	RLSUPPORT	0.033	0.001	0.671	0.67	247.418
	Dep = READ3					
1	SES	-0.858**	0.677	0.677	0.677	218.507
2	TCH_AGE	0.102*	0.005	0.682	0.681	217.126
3	TCH_YRS	-0.079	0.002	0.684	0.683	216.556
4	TCH_SAL	0.069	0.001	0.685	0.683	216.37
5	PCT_MAS	-0.045	0.001	0.686	0.684	216.165

https://newprairiepress.org/edconsiderations/vol33/iss2/9
DOI: 10.4148/0146-9282.1227

Table 2 Continued
Weighted Least Squares (WLS) Regression Results, 1998-2001

2001	**Dep = MATH5**					
Step	**Predictor Entered**	**Beta**	**R-sq Change**	**Cumulative R-sq**	**Adj. R-sq**	**SEE**
1	SES	-0.847**	0.604	0.604	0.603	254.654
2	TCH_SAL	0.123*	0.004	0.608	0.607	253.507
3	RLINSTRUCT	0.119**	0.004	0.612	0.61	252.272
4	RLSUPPORT	-0.063*	0.002	0.614	0.611	251.926
5	TCH_YRS	-0.054	0.002	0.615	0.613	251.562
6	TOT_ADM	-0.041	0.001	0.616	0.613	251.381
7	RNLINST	-0.035	0.001	0.617	0.614	251.236
	Dep = MATH3					
1	SES	-0.722**	0.487	0.487	0.486	296.404
2	TCH_YRS	0.049	0.005	0.492	0.491	295.053
3	RLINSTRUCT	0.105*	0.003	0.495	0.493	294.452
4	RLSUPPORT	-0.079*	0.003	0.498	0.495	293.703
5	TCH_SAL	0.048	0.002	0.5	0.496	293.428
	Dep = READ5					
1	SES	-0.883**	0.681	0.681	0.68	244.95
2	TCH_AGE	0.137**	0.007	0.688	0.687	242.34
3	RLINSTRUCT	0.068*	0.003	0.691	0.689	241.47
4	TCH_YRS	-0.074	0.001	0.692	0.69	241.07
5	TOT ADM	-0.041	0.001	0.693	0.691	240.8
6	RNLINST	-0.034	0.001	0.694	0.692	240.58
	Dep = READ3					
1	SES	-0.791**	0.616	0.616	0.615	225.163
2	TCH_AGE	0.098**	0.011	0.627	0.625	222.107
3	TOT_ADM	-0.038	0.001	0.628	0.626	221.874

** denotes $p < .01$
* denotes $.01 < p < .05$

identify schools and districts that consistently over- or under-performed as compared with outcome levels predicted by the SES and school resource measures. For example, a school that consistently exceeded its test performance as predicted by its students' characteristics (SES) and resource levels is assumed to benefit from positive but unobserved school and district attributes, attributes sometimes referred to as X-efficiency. For each outcome (i.e., grade level and subject), the residual was averaged by school building over the four years. Data for all four years were available for 476 schools. If the residuals were random, they would necessarily have a mean of zero.[22] They are not random, however, if they include the effects of unobserved variables that influence student achievement. Specifically, the average building residual reflected the joint effect on achievement made by the school and district. To decompose this effect into school and district effects, the residuals were averaged by school district, and the district average was subtracted from the total residual. The district average was interpreted as the upper bound for the district effect, and the difference between the total building residual and the district average was interpreted as the upper bound for the school effect.

To estimate the magnitude of these unobserved building and district effects on student achievement, the achievement measures were then regressed over these average residuals and the SES and school resource measures. The R^2 changes resulting from these stepwise regressions are presented in Table 4. As the results reported in Table 4 indicate, the district accounted for between 6 and 12 percent of the variance in measured achievement across all estimated models, averaging about 11 percent for mathematics and 8

Table 3
Upper and Lower Bounds for Estimates of R^2 Changes

Summary for MATH3

	Year				
	98	99	00	01	Avg.
SES	0.6122	0.5034	0.4830	0.4492	0.5120
SCHOOL 1	0.3120	0.2226	0.2517	0.2545	0.2602
SCHOOL 2	0.0129	0.0081	0.0107	0.0200	0.0129
SES+SCH2	0.6251	0.5115	0.4937	0.4692	0.5249
E/U	0.3749	0.4885	0.5063	0.5308	0.4751
N	506	671	691	695	

Summary for MATH5

	Year				
	98	99	00	01	Avg.
SES	0.6867	0.6033	0.5863	0.5704	0.6117
SCHOOL 1	0.3605	0.2905	0.3007	0.2656	0.3043
SCHOOL 2	0.0202	0.0107	0.0107	0.0074	0.0123
SES+SCH2	0.7069	0.6140	0.5970	0.5778	0.6239
E/U	0.2931	0.3860	0.4030	0.4222	0.3761
N	506	671	691	695	

Summary for READING3

	Year				
	98	99	00	01	Avg.
SES	0.7105	0.5849	0.6460	0.5753	0.6292
SCHOOL 1	0.3543	0.2642	0.3528	0.3257	0.3243
SCHOOL 2	0.0074	0.0141	0.0129	0.0195	0.0135
SES+SCH2	0.7179	0.5990	0.6589	0.5948	0.6427
E/U	0.2821	0.4010	0.3411	0.4052	0.3574
N	506	671	691	695	

Summary for READING5

	Year				
	98	99	00	01	Avg.
SES	0.7002	0.6655	0.6396	0.6483	0.6634
SCHOOL 1	0.3711	0.3535	0.3543	0.3308	0.3524
SCHOOL 2	0.0180	0.0120	0.0088	0.0129	0.0129
SES+SCH2	0.7182	0.6775	0.6484	0.6612	0.6763
E/U	0.2818	0.3225	0.3516	0.3388	0.3237
N	506	671	691	695	

percent for reading. The building accounted for between 11 and 18 percent of the variance in measured achievement, averaging about 16 percent for mathematics and 14 percent for reading. When the district is omitted from the regression, and the entire effect is attributed to the building, the building effect rises to an average of 22 percent for reading and 27 percent for mathematics. These effects, which reflect unobserved qualities of school administrators, faculty, support staff, and the climate they create, along with other unobserved variables, are substantial. The R^2 changes associated with building and district effects were then added to the R^2 changes associated with SES and school resource effects to obtain an estimate of the total explained variance in student achievement (R^2_{total}). The unexplained variance is estimated as $(1\text{-}R^2_{total})$ and is attributable to random error.

One may expect that these unobserved school and district effects would be roughly consistent across grades and subjects; that is, a good elementary school is good in all grades and subjects. To further examine the consistency of these effects across subjects and grades, the simple correlations across subjects and grades were examined. These correlation coefficients are presented in Table 5. The correlations are relatively high, confirming that the fixed effects or levels of x-efficiency taking place within a school building and school district tended to be consistent across subjects and grades over the four-year period examined. This conforms to intuition. The effects of such unobserved school and district variables as climate, communications, shared vision and goals, leadership, and incentives should be reflected throughout the school and not restricted to particular grades and subjects.

More generally, this consistent pattern of fixed effects or x-efficiency among the district and building residuals provides a measure of school and district influence on the quality of teaching and learning in the classroom. Not surprisingly, effective schools are found in effective districts. This finding was consistent across subjects and grade levels. Such a pattern of residuals reflects the effects of activities, climate, policies, incentives, instructional practices, and other inputs that are consistently present in the schools and districts but are not captured by the SES or school resource variables.[23]

Summary and Policy Conclusions

In keeping with a vast research literature on educational productivity, this analysis revealed that the socioeconomic characteristics of students remain the most influential factor in predicting achievement outcomes. A high SES school building (three standard deviations above the mean) can be expected to add about 30 percentile points to the average achievement level, raising a student from the 50th percentile to about the 80th, while a correspondingly low SES building would fall 30 percentile points below the mean. This relationship is depicted in standardized units in Figure 1.

SES exerted a much larger influence on academic achievement than did the various measures of school resources. Further, our estimates of school district and building fixed effects were considerably larger than the estimated effects of the school resource variables. This finding is consistent with Leibenstein, who observed in his seminal article on

https://newprairiepress.org/edconsiderations/vol33/iss2/9
DOI: 10.4148/0146-9282.1227

Table 4
Analysis of Residuals: Building and District Fixed Effects

Summary for MATH3

	Year				
	98	99	00	01	Avg.
Four Year Avg. N (E/U for the same N)					476
TOTAL	0.8389	0.8049	0.7922	0.7861	0.8055
BUILDING	0.1431	0.1696	0.1336	0.1978	0.1610
DISTRICT	0.0707	0.1238	0.1649	0.1191	0.1196
B AND D	0.2138	0.2934	0.2985	0.3169	0.2807
SES+SCH2	0.6251	0.5115	0.4937	0.4692	0.5249
ERROR	0.1611	0.1951	0.2078	0.2139	0.1945

Summary for MATH5

	Year				
	98	99	00	01	Avg.
Four Year Avg. N (E/U for the same N)					476
TOTAL	.09353	0.8830	0.8955	0.8584	0.8931
BUILDING	0.1281	0.1544	0.1805	0.1704	0.1584
DISTRICT	0.1003	0.1146	0.1180	0.1102	0.1108
B AND D	0.2284	0.2690	0.2985	0.2806	0.2691
SES+SCH2	0.7069	0.6140	0.5970	0.5778	0.6239
ERROR	0.0647	0.1170	0.1045	0.1416	0.1070

Summary for READING3

	Year				
	98	99	00	01	Avg.
Four Year Avg. N (E/U for the same N)					476
TOTAL	0.9503	0.8212	0.9152	0.8163	0.8758
BUILDING	0.1111	0.1474	0.1274	0.1632	0.1373
DISTRICT	0.0640	0.0850	0.0948	0.0931	0.0842
B AND D	0.1751	0.2324	0.2222	0.2563	0.2215
SES+SCH2	0.7179	0.5990	0.6589	0.5948	0.6427
ERROR	0.0497	0.1788	0.0848	0.1837	0.1243

Summary for READING5

	Year				
	98	99	00	01	Avg.
Four Year Avg. N (E/U for the same N)					476
TOTAL	0.9383	0.9100	0.8723	0.8832	0.9009
BUILDING	0.1253	0.1338	0.1503	0.1487	0.1395
DISTRICT	0.0860	0.0863	0.0822	0.0752	0.0824
B AND D	0.2113	0.2201	0.2325	0.2239	0.2220
SES+SCH2	0.7182	0.6775	0.6484	0.6612	0.6763
ERROR	0.0617	0.0900	0.1277	0.1169	0.0991

Figure 1
Influence of SES on Student Achievement

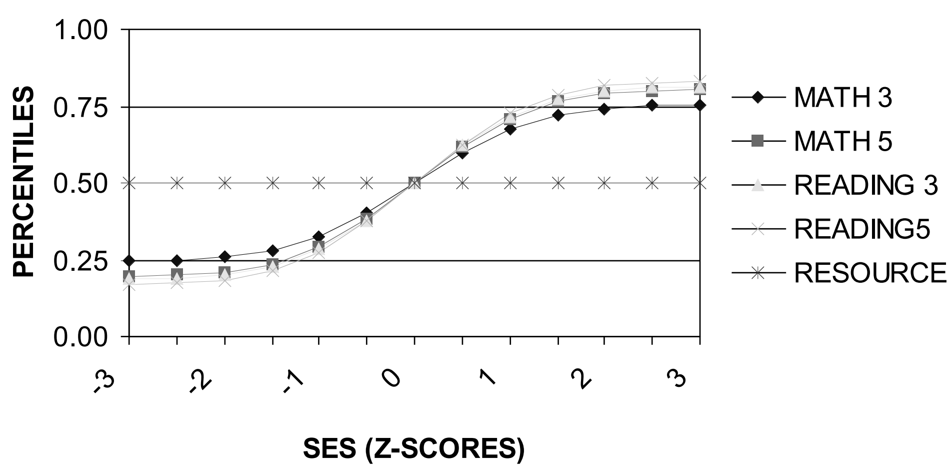


Table 5
Consistency of Building and District Effects: Correlations Among Estimates Across Grades and Subjects

Correlations				
		MATH 5		READ 5
	MATH 3		READ 3	
B-D		0.57		0.61
DAVE		0.75		0.52
BAVE		0.67		0.56
		READ 5		READ 5
	MATH 3		MATH 5	
B-D		0.61		0.88
DAVE		0.52		0.89
BAVE		0.56		0.87

X-efficiency in organizations that incentives, motivation, culture, and other organizational characteristics have far greater implications for efficiency than the allocation of inputs at the margins.

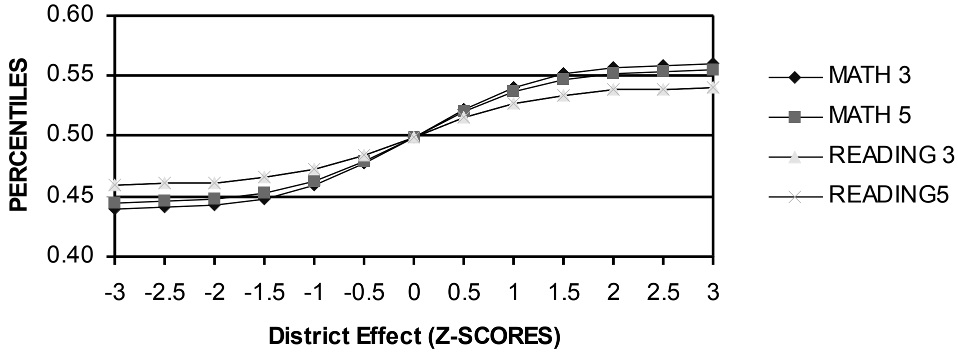
By our estimates, unobserved district characteristics exerted an influence on achievement outcomes, adding about five points at the high end (i.e., three standard deviations above the mean) and subtracting about five points at the low (i.e., three standard deviations below the mean). These effects are depicted in standardized units in Figure 2.

Unobserved building characteristics also exerted an influence on achievement outcomes, adding about seven points at the high end and subtracting about seven points at the low. These estimated effects are depicted in standardized units in Figure 3.

These findings hold several important implications for school accountability policies. First, holding schools accountable for average levels of measured achievement outcomes is tantamount to holding them accountable for the SES of the community. Level scores of student achievement say little about school quality. To ascribe high quality to schools in which children attain high scores on achievement tests is to confuse school quality with student attributes. Second, when SES and school resource variables are taken into consideration, high-performing and low-performing schools are found in all SES strata. Holding schools accountable for achievement outcomes when SES and school resources are taken into consideration may be appropriate. This could be accomplished by means of "value-added" analysis of the results of annual testing of every student in a school.

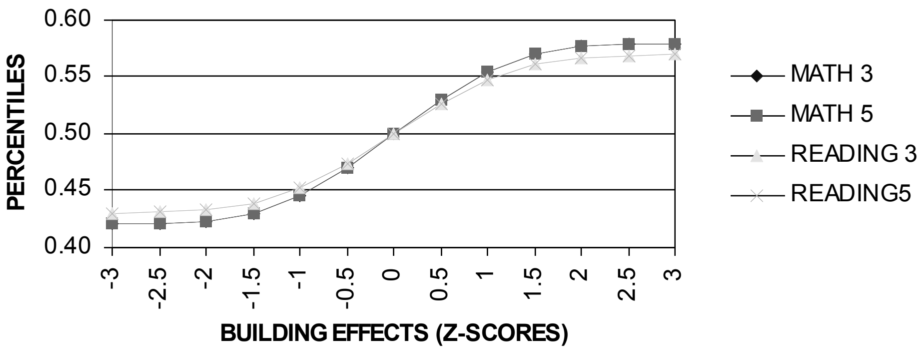
Further, a production function model of student achievement could be used to identify school districts and buildings that consistently exceed predicted performance levels. These school and districts should be the subject of case studies to identify the sources of their x-efficiency. Insights gained into school and, particularly, district climate, policies, operations, and incentives could be invaluable as states look for ways to improve teaching and learning in their public

Figure 2
Estimated District Fixed Effects



https://newprairiepress.org/edconsiderations/vol33/iss2/9
DOI: 10.4148/0146-9282.1227

Figure 3
Estimated Building Fixed Effects



schools in an economic environment that promises little in the way of increased resources in the near future. Case studies of this sort are not unusual in education research but are generally not conducted as part of an ongoing and systematic state-level effort to improve teaching and learning in our public schools. Currently, state departments of education and regional educational service agencies generally do not gather information regarding the behavior, activities, policies, or leadership at the school district or building levels that could explain differences in achievement outcomes across schools. Such qualitative data could be of enormous value to the schools. As the saying goes, "Not everything that counts can be counted," but leadership and sound practice can be observed and replicated.

Endnotes

[1] An earlier version of this paper was presented at the 2004 Annual Conference of the American Education Finance Association in Salt Lake City, Utah. Dr. Addonizio's work was supported by a grant from the State Policy Center of Wayne State University. Their support is gratefully acknowledged.

[2] *Goals 2000: Educate America Act*, P.L. 103-227, Section 301(1).

[3] Margaret E. Goertz and Mark C. Duffy, *Assessment and Accountability in the 50 States: 1999-2000*, CPRE Research Report Series RR-046 (Philadelphia, PA: Consortium for Policy Research in Education, 2001).

[4] For a discussion of this outcomes-based approach to school accountability, including the measurement of student and school performance and the design of associated school ratings, rewards, and sanctions, see Richard Elmore, C. Abelmann, and Susan Fuhrman "The New Accountability in State Education Reform: From Process to Performance," in *Holding Schools Accountable: Performance-based Reform in Education*, Helen F. Ladd, ed. (Washington, DC: Brookings Institution, 1996), 65-98. See also Richard King and J. Mathers, "Improving Schools Through Performance-based Accountability and Financial Rewards," *Journal of Education Finance* 23 (Fall 1997): 147-176.

[5] See, for example, Eric Hanushek, *Making Schools Work: Improving Performance and Controlling Costs* (Washington, DC: The Brookings Institution, 1994); see also, Henry M. Levin, "Raising School Productivity: An X-efficiency Approach," *Economics of Education Review* 16 (June 1997): 303-311.

[6] Kane and Staiger also found that changes in fourth-grade math gains have a correlation of 0.45 with changes in the next year, suggesting that 90% of the variance in the change in change scores is transitory. The authors report even less persistence in reading scores. The authors conclude that "...if one were to look for signs of improvement by closely tracking changes in school-level scores from one year to the next, most of what one observed would be temporary – either due to sampling variation or some other nonpersistent cause." Thomas J. Kane and Douglas O. Staiger, "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives* 23 (Fall 2002): 97.

[7] They also found a correlation of -.05 for a weighted index of all performance levels in a school. Robert L. Linn and C. Huag, "Stability of School-Building Accountability Scores and Gains," *Education Evaluation and Policy Analysis* 24 (Spring 2002): 29-36.

[8] Ibid., 33.

[9] Kane and Staiger, "The Promise and Pitfalls of Using Imprecise School Accountability Measures."

[10] Helen F. Ladd and R. Walsh, "Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right," *Economics of Education Review* 21 (February 2002): 1-17; Charles Clotfelter and Helen F. Ladd, "Recognizing and Rewarding Success in Public Schools," in *Ladd*, 23-63.

[11] Considerable controversy exists as to whether educational phenomena can be adequately represented in a strict production function framework. For an overview of the debate about the existence of an educational production function, see David Monk, *Educational Finance: An Economic Approach* (New York: McGraw-Hill, 1990), especially chapter 11. This brief summary draws upon Monk's overview.

[12] Levin, "Raising School Productivity."

[13] Harvey Leibenstein, "Allocative Efficiency and X-Efficiency," *The American Economic Review* 56 (March 1966): 392-425.

[14] Eric A. Hanushek, "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources* 14 (Summer 1979): 351-88.

[15] Monk, *Educational Finance*.

[16] In using reading and mathematics test scores as performance indicators, we assume these tests are valid for this use; that is, we

assume that these tests measure skills demanded by society (outcome validity) and that they accurately measure performance with respect to these skills (noncorruptability). For a discussion of school performance indicators, see Robert H. Meyer, "Value-Added Indicators of School Performance: A Primer," *Economics of Education Review*, 16 (June 1997): 283-301.

[17] Individual student scores on Minnesota's reading and mathematics assessments are based on a scale ranging from a minimum of approximately 50 to a maximum of approximately 2,500. The minimum and maximum scale scores vary slightly from year to year according to the performance of students at the extremes of the achievement range.

[18] Each of these component variables was found to be statistically significant in regressions of student achievement for each of the four years. Each component variable was then assigned a weight inversely proportional to its variance averaged over the four years. With this weighting method, each component variable contributes approximately the same amount of variance to the total variance of the composite SES variable. The SES index is an inverse measure of socioeconomic status; that is, a higher index score reflects lower socioeconomic status. For a complete discussion of the construction of composite measures, see J. P. Guilford, *Fundamental Statistics in Psychology and Education* (New York: McGraw-Hill, 1965), 416-426.

[19] Weighted least squares is an appropriate estimation technique when one suspects that the error terms are not of equal variance for each observation (heteroskedasticity). The most common instance of heteroskedasticity is with aggregate data, such as the school-level data examined here, where the dependent variable is a mean value for the individuals in the observational unit. The accuracy of the dependent variable will be a function of the number of individuals in the aggregate; that is, observations for the more populous units (e.g., schools) are presumably more accurate and should exhibit less variation about the true value than data drawn from smaller schools. This leads to different values of the error term variance for each observation--the heteroskedastic problem. For discussion see, for example, Eric Hanushek and John Jackson, *Statistical Methods for Social Scientists* (San Diego, CA: Academic Press, 1977) 142-153.

[20] For a discussion of the stepwise regression routine, see, for example, John Neter and William Wasserman, *Applied Linear Statistical Models* (Homewood, IL: Richard D. Irwin, Inc., 1974), 383.

[21] The following regressions are unweighted. These unweighted regressions yielded slightly lower coefficients of determination in 14 of 16 equations as compared with the weighted regressions. The average difference was approximately .028.

[22] The assumption that the error term has a mean value of zero is, of course, a part of the classical linear regression model. See, for example, Domar Gujarati, *Essentials of Econometrics* (New York: McGraw-Hill, Inc., 1992), 186-187.

[23] For a discussion of the importance of such generally unobserved school and district characteristics, see Levin, "Raising School Productivity."

https://newprairiepress.org/edconsiderations/vol33/iss2/9
DOI: 10.4148/0146-9282.1227