Review of Educational Research
Fall 1996, Vol. 66, No. 3, pp. 361–396

The Effect of School Resources on Student Achievement

Rob Greenwald
Larry V. Hedges
University of Chicago
Richard D. Laine
University of Chicago
Illinois State Board of Education

A universe of education production function studies was assembled in order to utilize meta-analytic methods to assess the direction and magnitude of the relations between a variety of school inputs and student achievement. The 60 primary research studies aggregated data at the level of school districts or smaller units and either controlled for socioeconomic characteristics or were longitudinal in design. The analysis found that a broad range of resources were positively related to student outcomes, with effect sizes large enough to suggest that moderate increases in spending may be associated with significant increases in achievement. The discussion relates the findings of this study with trends in student achievement from the National Assessment of Educational Progress and changes in social capital over the last two decades.

Over the last two decades the question of how best to improve the achievement of students in our nation's schools has gained increasing prominence. While expenditures have risen dramatically over this period, it has not been apparent that achievement has risen at all, much less commensurately. This potential paradox has led individuals both in the research community and in government to question if further increases in expenditures are warranted and whether such fiscal commitments are likely to result in schools meeting the expectations society holds for them.

Unfortunately, it has proven difficult to determine the relation between school expenditures (the things money may buy) and student achievement. Systematic efforts to do so began 35 years ago with Project Talent (Flanagan et al., 1964) and reached widespread public awareness with Equality of Educational Opportunity (Coleman et al., 1966). Much of the work in the ensuing three decades has employed similar methodology, using regression analysis to estimate the relation between school resources and achievement while controlling for student or family background characteristics. These studies, using the metaphor of the factory, view

Preparation of this article was supported in part by a grant from the Russell Sage Foundation. We would like to thank participants in the Seminar on Meritocracy and Equality and the Workshop on Schools and Schooling at the University of Chicago, and three anonymous reviewers, for their feedback on earlier versions of this work.

schools as producing achievement and thus employ the term *education production function* to describe the relation between school inputs and student outcomes.

The diverse literature presenting the results of education production functions has yielded mixed conclusions about the relation between school resources and student achievement. Coleman et al.'s (1966) original study found that resources had a surprisingly small impact on achievement. Subsequent production function research, including reanalyses of Coleman et al.'s work, produced some results which supported, and other results which challenged, Coleman et al.'s conclusions. Over the past 15 years Hanushek (1981, 1986, 1989, 1991) has published the results of a synthesis of a portion of the education production function literature. His conclusion that the data he assembled did not provide evidence of a strong or consistent relation between resources and student achievement has garnered considerable attention, and acceptance by some individuals, in the academic, legal, and public policy arenas.

Hanushek's synthesis method, vote counting, consists of categorizing, by significance and direction, the relations between school resource inputs and student outcomes (including but not limited to achievement). Unfortunately, vote counting is known to be a rather insensitive procedure for summarizing results (Hedges & Olkin, 1980). It is now rarely used in areas of empirical research where sophisticated syntheses of research are expected. Although Hanushek's work has been challenged since its publication (e.g., Spencer & Wiley, 1981; Baker, 1991), his summaries remain influential in most discussions of production function literature. After reanalysis of Hanushek's evidence, our (Hedges, Laine, & Greenwald, 1994) position was that the data he assessed on the relations between school resource inputs and student outcomes, including achievement, were substantially more consistent and positive than he believed. We found that the typical relation between input and outcome in the data he considered was positive and large enough to have important implications for educational policy. Indeed, for certain variables the typical (median) magnitude of the coefficients actually appeared to be too large to be plausible.

This article reports the results of an analysis of a more comprehensive collection of studies than we examined previously. By exhaustively searching the literature using explicitly specified search criteria, we have obtained a broad and reproducible universe of production function studies. An analysis of a subset of these data (Laine, Greenwald, & Hedges, 1995) and the more thorough analysis presented here lead to the conclusion that a broad range of school inputs are positively related to student outcomes, and that the magnitude of the effects are sufficiently large to suggest that moderate increases in spending may be associated with significant increases in achievement.

Constructing a Universe of Production Function Studies

The universe of production function studies was constructed from articles and books identified in one of four ways:

(1) Studies in the most complete review Hanushek assembled (Hanushek, 1989) were reassessed for possible inclusion.
(2) Electronic databases in economics, education, and psychology were searched.
(3) Literature reviews were employed as a means of identifying articles which

included relevant data.

(4) Citations in sources identified by the first three methods were evaluated.

Studies in the universe Hanushek (1989) constructed were assessed for quality. Of the 38 studies, 9 were discarded due to weaknesses identified in the decision rules for inclusion described below. While the remaining 29 studies were retained, many equations and coefficients failed to satisfy the decision rules we employed. Thus, while more than three quarters of the studies were retained, the number of coefficients from Hanushek's universe was reduced by two thirds.

Three electronic databases were searched extensively: (a) the ERIC database, which indexes journal and technical literature from *Resources in Education* and *Current Index to Journals in Education*; (b) PsycLIT, which corresponds to the American Psychological Association's *Psychological Abstracts*; and (c) EconLit, which is compiled from the American Economic Association's *Journal of Economic Literature* and the *Index of Economic Articles*. These databases were utilized to identify abstracts of potential interest. Each abstract was read to determine whether its article or book was likely to present relevant information. Those documents which (a) contained useful data, (b) presented theories on the utility and construction of education production functions, or (c) reviewed literature were retained for more careful scrutiny.

The ERIC database was searched for the period 1966–1993. *Input-output analysis* was the primary descriptor used in all ERIC searches. Every abstract that included *input-output analysis* and one of the following secondary descriptors was reviewed: *administrator qualifications, class size, cost effectiveness, educational assessment, educational facilities, educational finance, educational resources, expenditure per student, outcomes of education, productivity* (used in lieu of *education production functions*), *resource allocation, school effectiveness, teacher education, teaching experience,* and *teacher salaries.* Collectively these searches resulted in the review of more than 400 abstracts. Articles identified as containing data of interest were then used to determine whether any descriptors had been overlooked.

PsycLIT was searched for journal articles appearing in the period 1974–1993. Book titles were searched only for the more recent period 1987–1993. The level used in all searches was K–12; thus the term *NOT higher education* was used as a primary descriptor in all searches. The outcome measure (secondary descriptor) was identified as one of the following terms (all terms were entered as descriptors in the searches using the *or* function): *academic achievement, educational aspirations, educational objectives, income,* and *school learning.* In addition, the following tertiary descriptors were employed: *classroom environment, school administrators, school counseling, student characteristics, school environment, school facilities, teacher characteristics,* and *teacher education.* This search of PsycLIT yielded more than 1,000 abstracts to review.

EconLit was searched for the period 1969–1993. Unlike the ERIC database and PsycLIT, EconLit does not have a separate descriptor guide. ERIC terms were initially used. This yielded over 500 potentially relevant abstracts. As in the ERIC database, articles previously identified which contained data of interest were then utilized in order to determine if useful descriptors had been overlooked. This led to the combination of descriptors to yield three categories, which were then explored: (a) *analysis of education,* (b) *economics of education & human capital*

& value of human life, and (c) *economics of education & economics of discrimination & economics of minorities.* These combinations of terms were employed because each of the individual descriptors in the second and third categories yielded in excess of 2,000 abstracts independently.

The bibliographies of a number of literature reviews were used to identify other relevant publications (Averch, Carroll, Donaldson, Kiesling, & Pintero, 1972; Bridge, Judd, & Moock, 1979; Glasman & Biniaminov, 1981; Guthrie, Kleindorfer, Levin, & Stout, 1971; MacPhail-Wilcox & King, 1986; Monk, 1989, 1992). In addition, as articles were read for potential inclusion, cited sources which appeared to present relevant data were selected for retrieval.

The initial primary searches in the ERIC database, PsycLIT, and EconLit led to the review of more than 2,000 abstracts. From this pool, over 100 papers were retrieved. The combination of these papers with sources identified through other means yielded over 175 articles and books, which, in addition to the material from our reanalysis (Hedges et al., 1994), were reviewed in order to assemble the universe of studies used in the analysis presented below.

Narrowing the Universe of Studies: Decision Rules for Inclusion

Of the articles and books reviewed for inclusion, only 31, in addition to 29 articles and books drawn from Hanushek's (1989) most complete universe, both met our decision rules and contained the data necessary to perform the meta-analysis we contemplated. The narrowing of the universe assembled through the literature review was completed through the application of the following decision rules:

(1) *The data are presented in a refereed journal or a book.* Two studies retained from Hanushek's (1989) universe—Heim and Perl (1974) and Maynard and Crawford (1976)—were published by research institutes at universities and were retained in the universe.

(2) *The data originate in schools in the United States.* This rule was established due to our desire to apply the results to the United States and the difficulty of attempting to adequately account for cultural and structural differences in educational systems in foreign countries.

(3) *The outcome measure is some form of academic achievement.* While it is true that there are a multitude of objectives for public schools, we have attempted to focus on a specific outcome of students: the results of standardized achievement tests.

(4) *The level of aggregation is at the level of school districts or smaller units.* Moving beyond the level of school districts greatly limits the validity of the relation between inputs and outcomes. Note that one study in the universe (Sander & Krautmann, 1991) utilized counties as the level of aggregation. These data, however, were developed using district-level inputs aggregated to the county level only for the purpose of controlling for variations in socioeconomic status (SES) due to the absence of district-level SES data.

(5) *The model controls for socioeconomic characteristics or is either longitudinal (including a pretest and a posttest) or quasi-longitudinal (including IQ or a measure of earlier achievement as an input).* In order to avoid having student ability and background confound the findings, each production function was required to account for at least one of these factors.

(6) *The data are stochastically independent of other data included in the universe.* The issue of production functions using dependent data appeared in a number of situations. The most frequent occurrence was when a single article presented multiple regression equations with identical models using the same students, but varied the output (e.g., the verbal and quantitative scores on the SAT, or reading and math portions of the Iowa Tests of Basic Skills). Since the results of a single student's scores on various tests will likely be related, including multiple scores from the same student would result in stochastically dependent results. When this occurred, we calculated the median value for the regression coefficient, p-value, and t-ratio (absolute value) before including the data in our analysis. While this greatly reduced the number of coefficients, it increased the validity of our findings by eliminating the bias introduced by the inclusion of related outputs from the same population of students.

A second issue related to dependence of results arose when researchers reported their studies, or close variations, in multiple publications. If the same data and model were employed in multiple publications, all but one of the studies were discarded (Eberts & Stone 1984, 1987, 1988; Grimes & Register, 1990, 1991; Hanushek, 1971, 1972; Levin, 1970, 1976; Register & Grimes, 1991).

After the variables used in the studies constituting the universe were reviewed, school inputs in three general categories were selected for analysis: (a) expenditures (per-pupil expenditure [PPE], teacher salary), (b) teacher background characteristics (teacher ability, teacher education, teacher experience), and (c) size (class size, school size). Details of the individual research reports (articles and books) included in the universe are presented in the Appendix. The following information is provided for each study: (a) author, (b) year of the data, (c) size and description of the sample, (d) grade(s) of the students, (e) background control, (f) input variables, and (g) specifics of the output variable.

Methods

Two meta-analytic methods—combined significance testing and effect magnitude estimation—were employed in the analyses (for greater detail see Cooper & Hedges, 1994; Hedges et al., 1994). A separate analysis using each of these methods was completed for each of the seven input variables examined.

Combined Significance Tests

Combined significance tests provide a means of combining statistical significance values (p-values) from studies which test the same conceptual hypothesis but which may differ in the details of their designs or measurement methods (Hedges & Olkin, 1985). The inverse chi-square (Fisher) method was used to determine if the data are consistent with one or both of the null hypotheses in every study, or if there are effects in a specified direction in at least some of the studies. Two directional hypotheses were tested for each of the resource input variables: (a) the positive case, in which the null hypothesis states that no positive relation exists between the resource input and student outcome for the population coefficients, and (b) the negative case, in which the null hypothesis states that no negative relation exists between the resource input and student outcome for the population coefficients. It is possible to reject the null hypothesis in both the positive and negative cases, which would imply that there is evidence of both some positive and some negative relations. In order to reach the conclusion that "no strong or systematic relationship" (Hanushek, 1989, p. 47) exists between the major educational resource inputs and student outcomes, the data would have to be consistent with the null hypothesis in both the positive and the negative cases.

Effect Magnitude Analyses

Effect magnitude analyses attempt to estimate the strength of the relation between inputs and outputs. Because neither input nor output variables were typically measured on the same scale in all studies, the partial regression coefficients for the resource input variables could not be combined directly. Consequently, the index of effect magnitude used for most inputs was the fully standardized regression coefficient. This coefficient measures the number of standard deviations of change in output which would be associated with a change of one standard deviation in input. The median value of the coefficients for each variable was used as the summary of the effect size for that variable. We created a histogram for each variable in order to examine the distribution of effect sizes.

PPE and teacher salary were initially measured in dollars, and hence the inputs in these categories were directly comparable or could be made so after a correction for inflation (United States Center for Education Statistics, 1994). For these resource inputs, the half-standardized partial regression coefficient was used as the measure of effect magnitude (defined as $\beta_H = b/S_O$, where b is the unstandardized regression coefficient and S_O is the standard deviation of the output variable). The half-standardized regression coefficient measures the number of standard deviations of change in output associated with a one dollar change in input.

Grouping of Coefficients for Analysis

For both combined significance tests and effect magnitude analyses, the data were examined in a number of ways. In the full analysis, all independent p-values or effect magnitudes were analyzed. In order to assess the robustness of the results, the central 90% of the values were analyzed in order to determine the impact of outlier values. Since a number of the studies were based on data collected decades ago by Project Talent (Flanagan et al., 1964) and Equality of Educational Opportunity (Coleman et al., 1966), we analyzed both the full data set and the robustness subsample in two ways: (a) all studies, and (b) the subset of studies based on data collected after 1970 (in one variable, teacher salary, most of the studies were based on data collected after 1980; thus the subset of studies for this variable is of more recent vintage).

The cumulative nature of education presents a problem for education production function modeling, especially given the current high mobility of students in urban school systems. Cross-sectional studies usually attempt to control for background characteristics through the inclusion of a measure of the SES of families. These indexes, though sometimes elaborate constructions, are most frequently devised using parental income and/or parental education. Such controls are not designed to measure the educational background of students directly, but are relied upon due to their consistent relation with student achievement.

Two alternative types of studies designed to better control for student background effects were designated quasi-longitudinal and longitudinal. The former type of study attempts to control for either innate ability (IQ) or past performance (e.g., grade point average [GPA], American College Testing Program [ACT] scores, Scholastic Aptitude Test [SAT] scores). Frequently the measure of innate ability dates from early in a student's academic career, while the output measure is achievement in secondary grades. Whether certain measures of innate ability change appreciably past a certain point in development remains an area of debate within the psychometric community. While most quasi-longitudinal studies also attempt to control for SES, some studies deem the inclusion of past achievement alone as an adequate control for student background. One fifth of the studies in the universe were quasi-longitudinal.

More common in the literature are longitudinal studies, in which a pretest and a posttest are employed to measure student progress over a prescribed time period (usually one academic year). As in the quasi-longitudinal universe, most equations also employ an SES control. While some researchers (Bryk & Raudenbush, 1988) have found larger effects in longitudinal studies than in cross-sectional studies, many researchers involved in education production function modeling are attracted to the specific time course and explicit input-output distinction which longitudinal studies provide. One third of the studies in the universe were longitudinal. We analyzed the data from the subsets of studies which were longitudinal and quasi-longitudinal, and their union.

Measurement Issues

Three resource variables were measured in markedly different ways in different studies. For teacher education, the majority of studies used a dichotomous categorization based on the possession of a master's degree. Other studies utilized a variety of continuous categorizations, using a range of degrees and hours (e.g., MA +15, MA + 30). In some studies there were as many as nine categories of teacher education. We believe that the dichotomous categorization may be more reliable, and we therefore created a *dichotomous* subsample containing only those studies which used possession of a master's degree as the measure employed.

The majority of studies which presented data about teacher experience utilized a continuous measure of years. A few studies utilized discrete intervals (e.g., < 7 or > 7 years). We believe that the continuous categorization was more comparable across studies, and we therefore created a *continuous* subsample which included only those studies measuring teacher experience without specifying discrete intervals.

A few studies providing data about school size utilized the Test of Economic Literacy (TEL) as the output measure. The researchers presenting data using this measure predicted a negative relation between small schools and student outcomes. This was due to the belief that larger school size may allow the recruitment of a teacher specifically trained in economics. We therefore created an *exclude TEL* subsample which removed those studies that used the Test of Economic Literacy as an outcome measure.

Results

While each of the studies included in the universe provided the data required for at least one of the two methodologies (combined significance testing or effect magnitude estimation), not all studies provided the data required for both. While some studies provided data suitable only for use in combined significance testing, other studies provided data suitable only for use in effect magnitude estimation. In order to assess the possibility that the results might be affected by the selective reporting of data from specific studies, we examined the distribution of the production function coefficients for each of the resource variables by its statistical significance (significant or nonsignificant) and its direction (positive or negative). These data are presented in Table 1.

The distribution of data (p-values and t-ratios) for each of the input variables is similar with respect to significance and direction for the full analysis, combined significance testing, and effect magnitude estimation. Thus, the availability of

TABLE 1

Summary of the production function coefficients utilized in the analyses

	Significant		Nonsignificant		
Input variable	Positive	Negative	Positive	Negative	Total
Per-pupil expenditure					
Full analysis	15 (44%)	1 (3%)	10 (29%)	8 (24%)	34
Combined significance	15 (56%)	1 (4%)	5 (19%)	6 (22%)	27
Effect size estimation	11 (41%)	1 (4%)	8 (30%)	7 (26%)	27
Teacher ability					
Full analysis	12 (50%)	1 (4%)	9 (38%)	2 (8%)	24
Combined significance	12 (67%)	1 (6%)	4 (22%)	1 (6%)	18
Effect size estimation	10 (50%)	0 (0%)	8 (40%)	2 (10%)	20
Teacher education					
Full analysis	7 (15%)	6 (13%)	17 (37%)	16 (35%)	46
Combined significance	7 (18%)	6 (16%)	15 (40%)	10 (26%)	38
Effect size estimation	5 (21%)	1 (4%)	8 (33%)	10 (42%)	24
Teacher experience					
Full analysis	20 (29%)	2 (3%)	28 (41%)	18 (27%)	68
Combined significance	18 (30%)	2 (3%)	25 (42%)	15 (25%)	60
Effect size estimation	10 (37%)	1 (4%)	11 (41%)	5 (19%)	27
Teacher salary[a]					
Full analysis	3 (19%)	2 (13%)	9 (56%)	2 (13%)	16
Combined significance	3 (19%)	2 (13%)	9 (56%)	2 (13%)	16
Effect size estimation	3 (23%)	1 (8%)	7 (54%)	2 (15%)	13
Teacher/pupil ratio[b]					
Full analysis	13 (19%)	8 (12%)	34 (50%)	13 (19%)	68
Combined significance	13 (20%)	7 (11%)	32 (50%)	12 (19%)	64
Effect size estimation	8 (21%)	5 (13%)	17 (45%)	8 (21%)	38
School size[c]					
Full analysis	14 (36%)	4 (10%)	12 (31%)	9 (23%)	39
Combined significance	11 (39%)	4 (14%)	8 (29%)	5 (18%)	28
Effect size estimation	11 (36%)	2 (7%)	9 (29%)	9 (29%)	31

Note. Percentages may not sum to 100 due to rounding.

[a]The data for starting salary and average salary are combined due to the small number of coefficients.

[b]$\beta > 0$ means that smaller classes have greater outcomes.

[c]$\beta > 0$ means that smaller schools have greater outcomes.

data for each of the methods of analysis was not thought to substantially alter the results.

Positive coefficients indicate that greater resource inputs are related to higher achievement. Note that the signs of relations have been reversed for the variable school size and for the subset of coefficients presented as pupil/teacher ratio in the category teacher/pupil ratio so that positive coefficients reflect greater achievement in *smaller* schools and *smaller* classes, respectively. For each input variable a majority of the coefficients are positive (as many as 88% in the variable teacher ability). The proportion of the coefficients in the full analysis that are both positive and statistically significant ranges from 15% to 50%, which is from 6 to 20 times the proportion (2.5%) that would be expected to be significant and positive by chance alone, if resource inputs were unrelated to achievement.

The results of the combined significance tests are reported in Tables 2–4. Table 2 provides the data for the full analysis and the post-1970 subset. Table 3 provides the data for the longitudinal and combined longitudinal–quasi-longitudinal subsets. Table 4 provides a summary of results presented in Tables 2 and 3, indicating whether or not there was evidence of a positive or negative effect in each of the variable categories.

Results of Combined Significance Testing: Full Analysis

Examining the results of the tests in the positive direction, shown on the left side of Table 2 (with degrees of freedom and chi-square critical values in the center column), reveals that the null hypothesis for the positive test (that no positive relation exists between resource input and student achievement for the population coefficients) is rejected for every resource input. Thus there is evidence of positive coefficients associated with each of these input variables in the combined significance analyses. This result appears to be quite robust. It holds for both the full analysis of studies and the robustness (trimmed) sample, for both the entire collection of studies and the more recent (post-1970) studies, and for the subsamples created for the variables teacher education, teacher experience, and school size.

Examining the results of the tests in the negative direction, shown on the right side of Table 2, reveals that the null hypothesis for the negative test (that no negative relation exists between resource input and student achievement for the population coefficients) is accepted for the variables PPE, teacher ability, teacher experience, and teacher salary in the combined significance analyses. Thus there is no evidence of negative coefficients associated with any of these input variables in the combined significance analyses. This finding holds for both the full analysis of studies and the robustness (trimmed) sample, and for both the entire collection of studies and for the more recent (post-1970) studies.

The null hypothesis for the negative test is rejected for certain samples in the variables teacher education, teacher/pupil ratio, and school size. Thus there is evidence of coefficients associated with each of these three input variables in the combined significance analyses which are negative. The rejection of the null hypothesis for teacher education appears to be the result of outliers in the full sample (the null hypothesis for the robustness subsample is not rejected). The null hypothesis is rejected for all samples (full analysis and robustness subsample) using the dichotomous subsample.

TABLE 2

Results of combined significance tests: Full analysis

Input variable	Equations (studies)	Positive case (H_0: $\beta \leq 0$)		(df, χ^2)[a]		Negative case (H_0: $\beta \geq 0$)	
		FA	R[b]	FA	R[b]	FA	R[b]
Per-pupil expenditure							
All studies	27 (17)	236.43	190.34	(54, 72.15)	(50, 67.51)	35.91	27.66
Studies post-1970	16 (11)	144.05	121.33	(32, 46.19)	(28, 41.38)	12.47	7.62
Teacher ability							
All studies	18 (9)	428.93	382.87	(36, 51.00)	(32, 46.19)	14.19	2.04
Studies post-1970	9 (4)	239.34		(18, 28.87)		13.9	
Teacher education							
All studies	38 (24)	149.18	115.89	(76, 97.35)	(68, 88.25)	*143.05*	85.95
Studies post-1970	31 (17)	112.64	83.7	(62, 81.38)	(54, 72.15)	*126.90*	69.80
Dichotomous[c]	29 (16)	127.93	112.28	(58, 76.78)	(54, 72.15)	58.60	44.33
Dichotomous[c] post-1970	25 (12)	111.46	95.80	(50, 67.50)	(46, 62.83)	48.63	34.37
Teacher experience							
All studies	60 (29)	397.87	297.65	(120, 146.57)	(108, 126.57)	122.52	66.60
Studies post-1970	47 (19)	299.80	229.00	(94, 117.63)	(86, 108.65)	75.67	58.19
Continuous[d]	30 (24)	229.69	158.62	(60, 79.08)	(52, 69.83)	30.36	17.32
Continuous[d] post-1970	20 (16)	153.22	111.74	(40, 55.76)	(36, 51.00)	21.99	13.84
Teacher salary[e]							
All studies	16 (8)	95.12	95.10	(32, 46.19)	(28, 41.34)	27.49	17.71
Studies post-1980	13 (5)	77.89	55.03	(26, 38.89)	(22, 33.92)	28.83	17.04
Teacher/pupil ratio[f]							
All studies	64 (32)	327.18	243.99	(128, 155.41)	(116, 142.14)	147.60	89.22
Studies post-1970	52 (23)	268.54	185.35	(104, 128.80)	(92, 115.39)	*136.35*	77.96
School size[g]							
All studies	28 (18)	273.59	227.54	(56, 74.47)	(52, 69.83)	*113.69*	67.63
Studies post-1970	26 (16)	269.32	223.27	(52, 69.83)	(48, 65.17)	*111.99*	*65.94*
Exclude TEL[h]	24 (14)	272.57	226.52	(48, 65.17)	(44, 60.48)	33.25	17.03
Exclude TEL, post-1970	22 (12)	268.31	222.26	(44, 60.48)	(40, 55.76)	31.56	15.34

FA = full analysis, R = robust subsample.

Italics indicates failure to reject the null hypothesis in the (+) case, rejection of the null hypothesis in the (–) case.

[a] *df* = degrees of freedom. The chi-square (χ^2) values provided are at the $\alpha = 0.05$ level. The degrees of freedom and chi-square values are identical for the positive and negative cases.

[b] The robustness subsamples indicated are the middle 90%, trimming 5% from each side of the distribution.

[c] Dichotomous subsample includes only equations indicating possession of a master's degree.

[d] Continuous subsample includes only equations utilizing years of experience, without specifying discrete intervals.

[e] The results are similar if teacher salary is divided into starting salary and average salary and separate analyses are performed. The data is combined due to the small number of coefficients. Note that a post-1980 subsample is used.

[f] $\beta > 0$ means that smaller classes have greater outcomes.

[g] $\beta > 0$ means that smaller schools have greater outcomes.

[h] *Exclude TEL* subsample excludes equations with Test of Economic Literacy as the dependent variable.

A missing value indicates that due to the limited number of equations, a robustness sample was not created for this variable.

TABLE 3
Results of combined significance tests: Longitudinal subsample

Input variable	Equations (studies)	Positive case (H_0: $\beta \leq 0$) FA	Positive case (H_0: $\beta \leq 0$) R[b]	(df, χ^2)[a] FA	(df, χ^2)[a] R[b]	Negative case (H_0: $\beta \geq 0$) FA	Negative case (H_0: $\beta \geq 0$) R[b]
Per-pupil expenditure							
Longitudinal	7 (3)	*20.93*		(14, 23.69)		20.16	
Longitudinal/Quasi	20 (10)	174.43	128.35	(40, 55.76)	(36, 51.00)	26.10	17.85
Teacher ability[c]							
Longitudinal	4 (3)	30.53		(8, 15.51)		12.23	
Teacher education							
Longitudinal	19 (12)	*44.06*	*34.92*	(38, 53.38)	(34, 48.60)	*111.40*	*82.59*
Longitudinal/Quasi	20 (13)	61.71	*44.06*	(40, 55.79)	(36, 51.00)	*111.40*	*82.61*
Long.-Dichotomous[de]	13 (7)	43.43	34.30	(26, 38.89)	(22, 33.92)	27.63	13.34
Teacher experience							
Longitudinal	30 (14)	177.50	128.26	(60, 79.08)	(52, 69.83)	55.27	37.78
Longitudinal/Quasi	33 (17)	203.37	154.13	(66, 85.97)	(58, 76.78)	55.53	38.04
Long.-Continuous[f]	15 (12)	87.47	67.55	(30, 43.77)	(26, 38.89)	15.80	7.65
Long./Quasi-Cont.[f]	17 (14)	99.08	79.16	(34, 48.60)	(30, 43.77)	16.06	7.91
Teacher salary[g]							
Longitudinal/Quasi	8 (3)	61.27		(16, 26.30)		14.85	
Teacher/pupil ratio[h]							
Longitudinal	26 (10)	90.81	73.70	(52, 69.83)	(48, 65.17)	*73.19*	47.92
Longitudinal/Quasi	38 (18)	177.28	117.32	(76, 97.35)	(68, 88.25)	85.61	51.25
School size[i]							
Longitudinal	6 (5)	25.67		(12, 21.03)		*54.65*	
Longitudinal/Quasi	10 (9)	*29.35*	*20.12*	(20, 31.41)	(16, 26.30)	*84.91*	*38.84*
Long.-Exclude TEL[j]	4 (3)	25.63		(8, 15.51)		0.74	
Long./QL-Exclude TEL[j]	6 (5)	28.33		(12, 21.03)		4.48	

FA = full analysis, R = robust subsample.
Italics indicates failure to reject the null hypothesis in the (+) case, rejection of the null hypothesis in the (–) case.
[a]*df* = degrees of freedom. The chi-square (χ^2) values provided are at the $\alpha = 0.05$ level. The degrees of freedom and chi-square values are identical for the positive and negative cases.
[b]The robustness subsamples indicated are the middle 90%, trimming 5% from each side of the distribution.
[c]No studies in the variable *teacher ability* are quasi-longitudinal.
[d]Dichotomous subsample includes only equations indicating the possession of a master's degree.
[e]No studies in the variable *teacher education* in the dichotomous subsample are quasi-longitudinal.
[f]Continuous subsample includes only equations utilizing years of experience, without specifying discrete intervals.
[g]No studies in the variable *teacher salary* are longitudinal. The results are similar if teacher salary is divided into starting salary and average salary and separate analyses are performed.
[h]$\beta > 0$ means that smaller classes have greater outcomes.
[i]$\beta > 0$ means that smaller schools have greater outcomes.
[j]*Exclude TEL* subsample excludes equations with Test of Economic Literacy as the dependent variable.
A missing value indicates that due to the limited number of equations, a robustness sample was not created for this variable.

TABLE 4

Summary of results of combined significance tests: Full analysis and longitudinal subsample

Input variable	Equations (studies)	Evidence of positive effects?		Evidence of negative effects?	
		FA	R[a]	FA	R[a]
Per-pupil expenditure					
All studies	27 (17)	Yes	Yes	No	No
Studies post-1970	16 (11)	Yes	Yes	No	No
Longitudinal	7 (3)	No		No	
Longitudinal/Quasi	20 (10)	Yes		No	
Teacher ability					
All studies	18 (9)	Yes	Yes	No	No
Studies post-1970	9 (4)	Yes		No	
Longitudinal[b]	4 (3)	Yes		No	
Teacher education					
All studies	38 (24)	Yes	Yes	Yes	No
Studies post-1970	31 (17)	Yes	Yes	Yes	No
Dichotomous variables[c]	29 (16)	Yes	Yes	No	No
Dichotomous[c] post-1970	25 (12)	Yes	Yes	No	No
Longitudinal	19 (12)	No	No	Yes	Yes
Longitudinal/Quasi	20 (13)	Yes	No	Yes	Yes
Long. dichotomous[c]	13 (7)	Yes	Yes	No	No
Long./Quasi-dichotomous[c]	13 (7)	Yes	Yes	No	No
Teacher experience					
All studies	60 (29)	Yes	Yes	No	No
Studies post-1970	47 (19)	Yes	Yes	No	No
Continuous variables[d]	30 (24)	Yes	Yes	No	No
Continuous[d] post-1970	20 (16)	Yes	Yes	No	No
Longitudinal	30 (14)	Yes	Yes	No	No
Longitudinal/Quasi	33 (17)	Yes	Yes	No	No
Long. continuous[d]	15 (12)	Yes	Yes	No	No
Long./Quasi-continuous[d]	17 (14)	Yes	Yes	No	No
Teacher salary[e]					
All studies	16 (8)	Yes	Yes	No	No
Studies post-1980	13 (5)	Yes	Yes	No	No
Longitudinal/Quasi	8 (3)	Yes		No	
Teacher/pupil ratio[f]					
All studies	64 (32)	Yes	Yes	No	No
Studies post-1970	52 (23)	Yes	Yes	Yes	No
Longitudinal	26 (10)	Yes	Yes	Yes	No
Longitudinal/Quasi	38 (18)	Yes	Yes	No	No

TABLE 4 *(continued)*

Input variable	Equations (studies)	Evidence of positive effects? FA	Evidence of positive effects? R[a]	Evidence of negative effects? FA	Evidence of negative effects? R[a]
School size[g]					
All studies	28 (18)	Yes	Yes	Yes	No
Studies post-1970	26 (16)	Yes	Yes	Yes	Yes
Exclude TEL[h]	24 (14)	Yes	Yes	No	No
Exclude TEL[h] post-1970	22 (12)	Yes	Yes	No	No
Longitudinal	6 (5)	Yes		Yes	
Longitudinal/Quasi	10 (9)	No	No	Yes	Yes
Longitudinal-Exclude TEL[h]	4 (3)	Yes		No	
Long/Quasi-Exclude TEL[h]	6 (5)	Yes		No	

Note. FA = full analysis, R = robust subsample. Evidence of positive effects indicates that the null hypothesis was rejected at $\alpha = 0.05$ level. Evidence of negative effects indicates that the null hypothesis was rejected at $\alpha = 0.05$ level.

[a]The robustness samples indicated are middle 90%, trimming 5% from each side of the distribution.

[b]No studies in the variable *teacher ability* are quasi-longitudinal.

[c]Dichotomous subsample includes only equations indicating the possession of a master's degree.

[d]Continuous subsample includes only equations utilizing years of experience, without specifying discrete intervals.

[e]No studies in the variable *teacher salary* are longitudinal. The results are similar if teacher salary is divided into starting salary and average salary and separate analyses are performed. A post-1980 subsample is used.

[f]$\beta > 0$ means that smaller schools have greater outcomes.

[g]$\beta > 0$ means that smaller schools have greater outcomes.

[h]*Exclude TEL* subsample excludes equations with Test of Economic Literacy as the dependent variable.

A missing entry indicates that due to the limited number of equations, a robustness sample was not created for this variable.

In teacher/pupil ratio, the null hypothesis is rejected for the full analysis in the post-1970 subsample. This rejection appears to be due to the presence of outliers, as the null hypothesis is accepted for the robustness sample of post-1970 studies. For the variable school size, the rejection of the null hypothesis for the full analysis appears to be the result of outliers, as the null hypothesis is accepted for the robustness subsample. However, for the subset of studies which contain post-1970 data, there is evidence of negative effects in both the full analysis and the robustness subsample. These findings are explained by the hypothesis about studies utilizing the TEL as the output measure explained above. The exclusion of studies using the TEL as an outcome results in the acceptance of the null hypothesis for the negative case in school size. Thus there is no evidence of negative coefficients associated with this input variable in the combined significance analyses when the outcome measure is not the TEL.

Results of Combined Significance Testing: Longitudinal Subsample

Table 3 provides the data for the combined significance tests for the longitudinal and the combined longitudinal–quasi-longitudinal subsets. Since only one third of the studies in the universe were longitudinal, the quantity of data is much more limited in these subsamples. For example, though 9 studies (18 coefficients) presented data on teacher ability in the full analysis, only 3 of these studies were longitudinal (4 coefficients). We are reluctant to make policy recommendations based on data assembled from a very limited number of studies. In order to increase the number of studies under consideration, the longitudinal and quasi-longitudinal subsets have been combined. In two cases—PPE and teacher/pupil ratio—the quantity of data was greatly enhanced due to the addition of quasi-longitudinal studies. In PPE, twice as many studies were quasi-longitudinal than were longitudinal. In this case, as in a number of others, the longitudinal–quasi-longitudinal subsets represent a majority (in this case more than two thirds) of the data presented in the full analysis, and this may be sufficient to justify their use as a basis for extrapolation.

Examining the results of the tests in the positive direction, shown on the left side of Table 3, reveals that the null hypothesis for the positive test (that no positive relation exists between a resource input and student achievement for the population coefficients) is rejected for every resource input in the longitudinal subset with the exception of PPE and teacher education. Due to the limited number of coefficients in PPE, it is impossible to determine if this finding is due to an outlier. When the larger data set incorporating both longitudinal and quasi-longitudinal studies which include PPE are considered, the null hypothesis is rejected. In teacher education, the null hypothesis in the positive direction is rejected when the dichotomous subsample is considered. The results in the combined longitudinal–quasi-longitudinal subset for the variable school size appear to be due to the studies which utilize the TEL as the outcome measure. Excluding these studies leads to the rejection of the null hypothesis for the positive case.

Examining the results of the tests in the negative direction, shown on the right side of Table 3, reveals that the null hypothesis for the negative test (that no negative relation exists between a resource input and student achievement for the population coefficients) is accepted for the variables PPE, teacher ability, teacher experience, and teacher salary in the longitudinal subset. In teacher/pupil ratio, the rejection of the null hypothesis in the negative direction appears to be due to the presence of outliers, as the null hypothesis is accepted for the robustness subsample. As in the positive case, the findings for teacher education appear to be due to the studies which utilized nondichotomous measures of teacher education. The null hypothesis is accepted for those studies using the dichotomous subsample in teacher education. Similarly, in school size, the rejection of the null hypothesis appears to be due to those studies which used the TEL as the outcome measure. The exclusion of studies using the TEL as an outcome measure results in the acceptance of the null hypothesis for the negative case in school size.

Table 4 provides the same information as Tables 2 and 3, but in a more simplified form. Table 4 summarizes the evidence, indicating whether either positive or negative effects are present. It is easy to see that the data from the full analyses support conclusions similar to those drawn from the longitudinal and combined longitudinal–quasi-longitudinal subsamples.

Results of Effect Magnitude Estimation: Full Analysis

The results of the effect size analyses are presented in Table 5. The median effect magnitudes are, in general, consistent with the results of the combined significance tests. The median effect (standardized or half-standardized regression coefficient) is positive for all of the resource variables. As in the combined significance analysis, the signs of relations have been reversed for the variable school size and for the subset of coefficients presented as pupil/teacher ratio in the category teacher/pupil ratio so that positive coefficients reflect greater achievement in *smaller* schools and *smaller* classes, respectively.

The pattern of effect sizes for the newer (post-1970) studies is the same as that for the entire collection of studies. The median of the coefficients for the variables teacher ability, teacher education, teacher experience, teacher/pupil ratio, and school size (for the *exclude TEL* subset) appear to be somewhat more positive for recent studies (although the teacher ability finding is based on a single publication, Ferguson, 1991). These findings are important because newer studies are certainly more relevant to conditions in contemporary schools, and they tend to utilize stronger methodology. While the median of the coefficients for the variables PPE and teacher salary are smaller among the more recent studies, they are still large enough to have important implications for policy. Figure 1 presents histograms of the effect sizes for the full analysis of each of the variables, and for the longitudinal–quasi-longitudinal subsample for the variable PPE. Some variation in estimates would be expected due to sampling error. In a collection of studies, some effect estimates would be expected to be negative, even if all parameters were small but positive. These histograms demonstrate that the vast majority of effect sizes are positive for each variable.

Results of Effect Magnitude Estimation: Longitudinal Subsample

The results from the effect size analyses of the longitudinal and combined longitudinal–quasi-longitudinal subsamples are presented in Table 6. The results of the full analysis for each variable are presented for comparison. For a number of variables, the quantity of information in the longitudinal and longitudinal–quasi-longitudinal analyses is too limited to draw conclusions. The number of coefficients in teacher ability, teacher education, and school size (*exclude TEL* subsample) is three or smaller. The number of studies in the longitudinal–quasi-longitudinal subsample for teacher salary is only two. Thus the only variables with sufficient data to warrant interpretation are PPE, teacher experience, and teacher/pupil ratio.

In PPE and teacher/pupil ratio, as in the combined significance analysis, far more data are available if the longitudinal and quasi-longitudinal subsamples are combined. In the case of both PPE and teacher/pupil ratio, the median regression coefficients in the longitudinal subsample are negative, indicating negative relations between expenditures and class size, respectively, and student achievement. In the case of PPE the data are drawn from only one study, and in teacher/pupil ratio the data are based on four coefficients. Including the quasi-longitudinal studies increases the number of studies substantially for both PPE and teacher/pupil ratio to data sets sufficiently large to draw conclusions.

The median effect size for PPE in the combined longitudinal–quasi-longitudi-

TABLE 5
Median regression coefficients

Input variable	Equations (studies)	Sample: Full analysis	Sample: Publication bias robustness[a]
Per-pupil expenditure[b]			
All studies	27 (14)	.0003	.0003
Studies post-1970	11 (8)	.0001	.0001
Teacher ability			
All studies	20 (6)	.0724	.0500
Studies post-1970	6 (1)	.2230	.2110
Teacher education			
All studies	24 (12)	.0003	–.0050
Studies post-1970	15 (8)	.0430	.0406
Dichotomous[c]	15 (8)	.0430	.0406
Dichotomous[c] post-1970	13 (6)	.0476	.0430
Teacher experience			
All studies	27 (15)	.0482	.0461
Studies post-1970	11 (8)	.0984	.0461
Continuous[d]	22 (12)	.0431	.0400
Continuous[d] post-1970	11 (8)	.0984	.0461
Teacher salary[e]			
All studies	13 (5)	.0263	.0135
Studies post-1980	11 (3)	.0007	.0005
Teacher/pupil ratio[f]			
All studies	38 (21)	.0295	.0266
Studies post-1970	29 (15)	.0470	.0470
School size[g]			
All studies	31 (15)	.0299	.0183
Studies post-1970	21 (12)	.0299	.0183
Exclude TEL[h]	28 (12)	.0376	.0183
Exclude TEL[h] post-1970	18 (9)	.0499	.0190

Note. These results are summaries of fully standardized regression coefficients unless otherwise indicated.

[a]Publication bias robustness sample created by counting the effect size of each nonsignificant coefficient twice.

[b]Half-standardized regression coefficients. Units are dollars, adjusted to 1993–94.

[c]Dichotomous subsample includes only equations indicating the possession of a master's degree.

[d]Continuous subsample includes only equations utilizing years of experience, without specifying discrete intervals.

[e]Half-standardized regression coefficients. Units are thousands of dollars, adjusted to 1993–94.

[f]$\beta > 0$ means that smaller classes have greater outcomes. Regression coefficients for pupil/teacher ratio were converted to coefficients for teacher/pupil ratio by expanding a Taylor series about the mean.

[g]$\beta > 0$ means that smaller schools have greater outcomes.

[h]*Exclude TEL* subsample excludes equations with Test of Economic Literacy as the dependent variable.

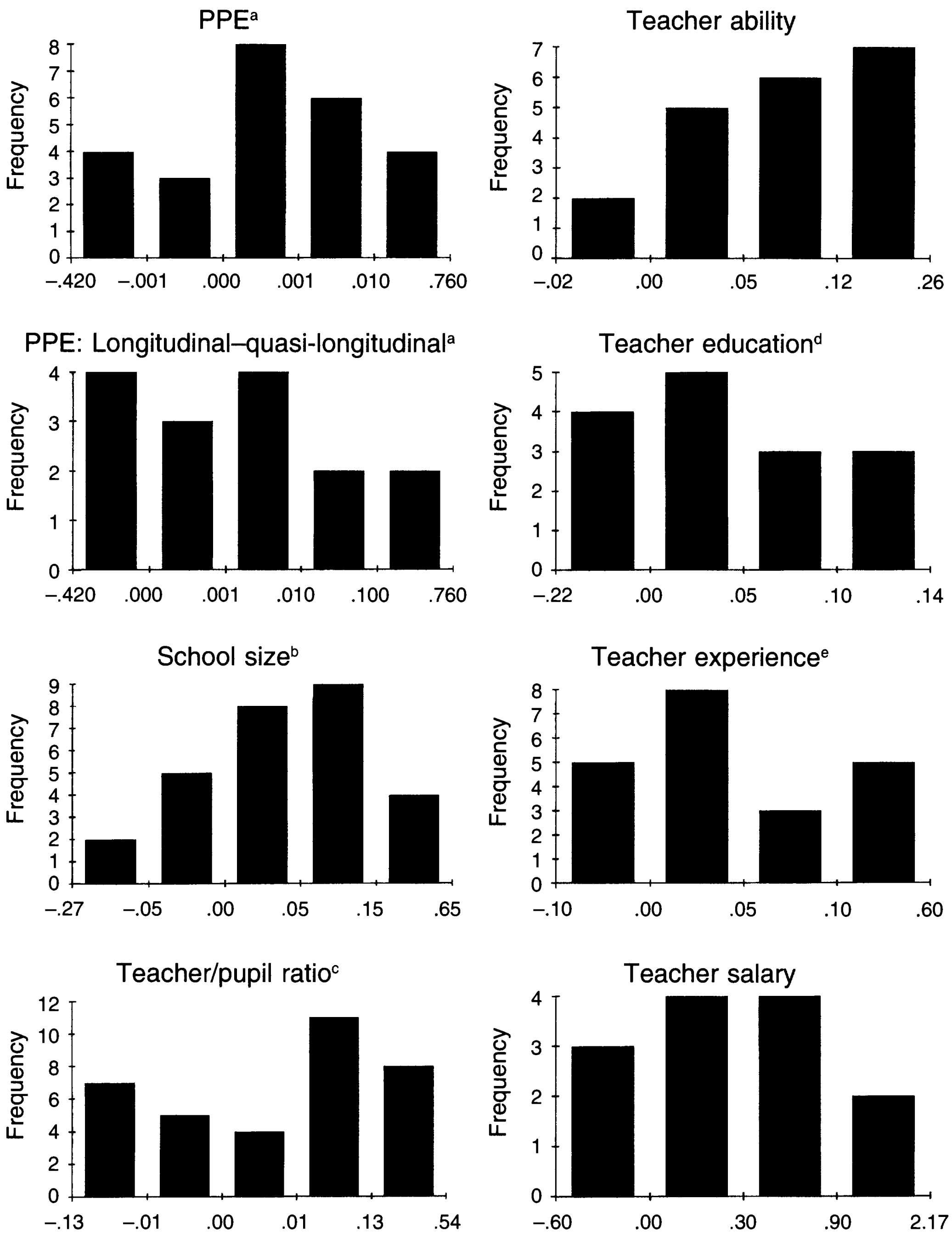

FIGURE 1. *Histograms of effect sizes*

Note. The width of the effect size intervals depicted are nonuniform in order to provide the most information. The effect sizes are fully standardized regression coefficients, except PPE and teacher salary, which are half-standardized. PPE = per-pupil expenditure.

[a]In PPE and PPE: Longitudinal–quasi-longitudinal, one outlier ($\beta = 4.05$) omitted.

[b]$\beta > 0$ means that smaller schools have greater outcomes. Equations using the Test of Economic Literacy excluded.

[c]$\beta > 0$ means that smaller classes have greater outcomes. Three outliers omitted ($\beta = -0.82$, 1.7, 7.5).

[d]Dichotomous subsample limited to equations that utilized absence/possession of a master's degree as the education level.

[e]Continuous subsample limited to equations that utilized years, without specifying intervals. One outlier omitted ($\beta = 1.92$).

TABLE 6
Comparison of median regression coefficients: Full analysis and longitudinal subsample

	Full analysis		Longitudinal (*quasi-long.)	
Input variable	Equations (studies)	Coefficient	Equations (studies)	Coefficient
Per-pupil expenditure[a]	27 (14)	.0003	3 (1)	–.3394
			16 (8)	.0026*
Teacher ability[b]	20 (6)	.0724	2 (1)	.0671
Teacher education[c]	24 (12)	.0003	2 (2)	–.0099
Dichotomous[cd]	15 (8)	.0430	1 (1)	.0000*
Teacher experience	27 (15)	.0482	5 (3)	.0726
			7 (5)	.0726*
Continuous[e]	22 (12)	.0431	5 (3)	.0726
			6 (4)	.0604*
Teacher salary[f]	13 (5)	.0263	7 (2)	.5590*
Teacher/pupil ratio[g]	38 (21)	.0295	4 (3)	–.0058
			14 (9)	.0295*
School size[h]	31 (15)	.0299	2 (2)	–.0342
			6 (6)	.0726
Exclude TEL[i]	28 (12)	.0376	1 (1)	.0400
			3 (3)	.0351*

Note. These results are summaries of fully standardized regression coefficients unless otherwise indicated.
[a]Half-standardized regression coefficients. Units are dollars, adjusted to 1993–94.
[b]No studies in the variable *teacher ability* are quasi-longitudinal.
[c]No studies in the variable *teacher education* are quasi-longitudinal.
[d]Dichotomous subsample includes only equations indicating the possession of a master's degree.
[e]Continuous subsample includes only equations utilizing years of experience, without specifying discrete intervals.
[f]Half-standardized regression coefficients. Units are thousands of dollars, adjusted to 1993–94. No studies in the variable *teacher salary* are longitudinal.
[g]$\beta > 0$ means that smaller classes have greater outcomes.
[h]$\beta > 0$ means that smaller schools have greater outcomes.
[i]*Exclude TEL* subsample excludes equations with Test of Economic Literacy as the dependent variable.

nal subsample is an order of magnitude larger than that found in the full analysis. This result is quite surprising, and we believe it provides more conclusive evidence about direction, rather than size, of effect. We hope that additional longitudinal studies are completed in order to enlarge the data available for this analysis. The median effect size for teacher experience in both the longitudinal and longitudinal–quasi-longitudinal subsamples are greater than, but on the same order of magnitude as, the effect size for the full analysis. The median effect size for teacher/pupil ratio in the longitudinal–quasi-longitudinal subsample is equal to that found in the full analysis. In general, the effect size analyses of the longitudinal and longitudinal–quasi-longitudinal subsamples support the conclusions drawn from the full analysis.

What the Effect Sizes Mean

In most of the variables the effect size of the median regression coefficient is substantial. For two variables, PPE and teacher salary, it is relatively simple to interpret the effect sizes. For each of these variables, the half-standardized regression coefficient measures the number of standard deviations of change in output associated with a one dollar change in input. For the other variables, a fully standardized regression coefficient measures the number of standard deviations of change in output which would be associated with a change of one standard deviation in input. For these variables, a number of assumptions must be made in order to translate standardized regression coefficients into information that is easily interpretable. Table 7 presents a summary of the effects of $500/student on achievement. The quantity $500 was selected for two reasons. It is approximately 10% of the national average of PPE (United States Center for Education Statistics,

TABLE 7
The effect of $500[a] per student on achievement[b]

	Sample	
Input variable	Full analysis	Publication bias robustness[c]
Per-pupil expenditure[d]	0.15	0.15
Teacher education[e]	0.22	0.20
Teacher experience[f]	0.18	0.17
Teacher salary[g]	0.16	0.08
Teacher/pupil ratio[h]	0.04	0.04

[a]1993–94 dollars

[b]All achievement outcomes are in standard deviation units.

[c]Publication bias robustness sample created by counting the effect size of each nonsignificant coefficient twice.

[d]Extrapolation from effect size estimate = 0.0003 SD in achievement/$1 of PPE.

[e]Dichotomous subsample includes only equations indicating the possession of a master's degree. Assumptions: class size = 25; teacher salary represents 50% of PPE; average teacher salary = $35,000/year; possession of an MA = $3,500 additional teacher salary/year; 50% of teachers possess a BA; 1 SD of teacher education = 10%. Extrapolation from effect size estimate = 0.0430 SD in achievement/SD of teacher education (0.0406 in robust).

[f]Continuous subsample includes only equations utilizing years of experience, without specifying discrete intervals. Assumptions: class size = 25; teacher salary represents 50% of PPE; 1 SD of teacher experience = 3.5 years; 3.5 years of experience = $1,500 in teacher salary. Extrapolation from effect size estimate = 0.0431 SD in achievement/SD of teacher experience (0.0400 in robust).

[g]Assumptions: class size = 25; teacher salary represents 50% of PPE. Extrapolation from effect size estimate = 0.0263 SD in achievement/$1,000 of teacher salary (0.0135 in robust).

[h]$\beta > 0$ means that smaller classes have greater outcomes. Assumptions: class size = 25; teacher salary represents 50% of PPE; average teacher salary = $35,000/year; 1 SD of class size = 3 students. Extrapolation from effect size estimate = 0.0295 SD in achievement/SD of class size (.0266 in robust).

Assumptions concerning class size, average teacher salary, and teacher salary as a percentage of PPE are from *Digest of Education Statistics,* 1994.

1994) and represents a spending increase of the order of magnitude that legislative bodies may be willing to consider when determining fiscal allocations for schools. Note that we were unable to determine a reliable method to estimate the effect on achievement of a specific dollar amount devoted to decreasing school size or recruiting teachers with greater verbal ability. Thus these variables are not presented in Table 7.

For PPE, the median half-standardized coefficient for the full analysis is .0003, which suggests that an increase in PPE of $500 would be associated with an increase in achievement of nearly one sixth of one standard deviation. This is a somewhat smaller effect than was estimated in our reanalysis (Hedges et al., 1994), but it is derived from a more reliable, higher-quality, and more recent data set. The magnitude of effects for the global resource measure (PPE) and the largest single component of school costs (teacher salaries) are comparable. The median half-standardized coefficient for the variable teacher salary is .0263, which suggests that an increase of $500/student (which converts to $12,500 in teacher salary, assuming that class size is 25 and that teacher salaries account for 50% of PPE) is associated with an increase of one sixth of one standard deviation in student achievement.

The details of the evaluation of the magnitude of the effects of teacher ability, teacher education, teacher experience, and teacher/pupil ratio are presented in the notes to Table 7. The magnitudes for teacher education and teacher experience are higher than, but of the same order of magnitude as, PPE. That is, one would expect comparable and substantial increases in achievement if resources were targeted to selecting (or retaining) more educated or more experienced teachers.

The data in Table 7 suggest that the smallest increase in student achievement may be expected from utilizing increased expenditures to reduce class size. However it is important to recognize that these analyses are based on the variable teacher/pupil ratio, which is an imperfect measure of class size. The reasons for such imperfection include the fact that class sizes vary within a school, with certain classes being significantly smaller than others (e.g., special education), and the fact that schools vary in the proportion of certified personnel who are not functioning as classroom teachers. The variance in the correspondence between teacher/pupil ratio and class size constitutes measurement error, which may affect the size of the median effect for this variable.

Many educators believe that class size is a critical variable in student learning. The data provide some evidence to support this belief, and the positive coefficient is consistent with the findings of a number of high-quality studies on class size which do not present their data in the same framework and thus are difficult to combine with production function studies. The results on class size presented here are consistent with the extensive experimental literature, which suggests that smaller class sizes produce greater achievement (Glass & Smith, 1979; Hedges & Stock, 1983). (Note that these 2 studies and 2 additional class size studies [Finn & Achilles, 1990; McGiverin, Gilman, & Tillitski, 1989] were reviewed but were not included in our data set because they did not report data in a form that could be used in these analyses. The data from the former study provide evidence of effects on the same order of magnitude as we have found.)

When the results of the combined significance tests and the effect magnitude analyses are examined together, the findings suggest a substantially positive relation between educational resource inputs and academic achievement. These results are similar to those obtained in our earlier reanalysis (Hedges et al., 1994). However, the present results seem to suggest even stronger and more consistent relations between educational resources and student outcomes.

Why Have Previous Reviews Failed to Detect Positive Effects?

In previous work, we criticized Hanushek for summarizing the estimates of production function coefficients by computing the proportion of coefficients with significant *p*-values (vote counting). When individual studies have relatively low statistical power, only a small proportion of studies would be expected to obtain statistical significance, even if each study were estimating the same (nonzero) effect. However, we have not seen any power calculations for education production function models. If statistical power were relatively high for plausible values of resource effects, the conclusion from vote count analyses might be less suspect.

Precise calculations of statistical power in regression models require the covariance matrix of the variables involved, which is rarely reported. We elected instead to compute approximate power by assuming that the standard error of the coefficient in each study remained the same, but the expected value of the regression coefficient was set so that a 10% change in PPE corresponded to a change in output of a specific number of standard deviations. We argue that a large, but plausible, value might be that a 10% change in PPE could result in a change in output of 1 standard deviation over 12 years of schooling, or .083 standard deviations per year.

Nineteen of the 27 equations in the PPE analysis reported the information necessary to make this approximate power computation (namely, the standard error of the coefficient and the standard deviation of the output). Three of the 19 would have had power in excess of 90%, 2 would have had power just under 50%, and the remaining 14 would have had power less than 20%. The average power to detect this large, but plausible, effect would have been no greater than about 33% among these studies. Hence a large proportion of significant results would not be expected (even if all studies were estimating this effect), and vote counting would be expected to miss effects.

Publication Bias

The analyses reported in this article did not include data from unpublished papers or from analyses in published papers that were reported so incompletely that neither values of coefficients nor their levels of statistical significance (*p*-values) were given. It might be argued that because of publication bias the omitted data might have been systematically different from that available, and that if they could have been included, these omitted data might have changed the results of our analyses. There is no doubt that publication bias exists in many fields (see Begg, 1994) and that it can compromise the interpretation of empirical research studies.

The most frequently studied model of publication bias is the case in which each research study tests only one hypothesis and the likelihood of publication depends only on the two-tailed *p*-value for that single hypothesis test (e.g., Hedges, 1984; Hedges & Olkin, 1985). This model may describe research literatures fairly well, such as those consisting of experiments that are reasonably focused on a single hypothesis. It does not, however, describe education production function studies, in which the required statistical method is inherently multivariate and every study tests many hypotheses, some of which are virtually guaranteed to attain statistical significance. In cases such as this, the relation between a single one of the effects in a study and the likelihood of publication is surely weaker, and the effects of publication bias are therefore also weaker. Even if the simplified model did apply, the effect of publication bias would be to reduce the magnitude of the positive effects somewhat, not to change their direction.

We completed a sensitivity analysis to see what effects publication bias might have had on our results. The best studies of publication bias involve the long-term follow-up of experiments (clinical trials) in medicine. Easterbrook, Berlin, Gopalan, and Matthews (1991), Dickersin, Min, and Meinert (1992), and Dickersin and Min (1993) found the relative rates of publication of studies with nonsignificant results to be 61%, 80%, and 86%, respectively. This suggests that between 14% and 39% of statistically nonsignificant results are not published. Although we believe that the effects of publication bias should be less severe in this literature for the reasons given above, we used a value exceeding the highest estimate from medicine, and estimated that only 50% of the statistically nonsignificant results were observed. If we assume that the statistically nonsignificant results observed are not different from those which were eliminated by publication selection, we can simulate the effect of publication bias by giving twice as much weight to each observed insignificant result. We actually accomplished the weighting by counting each statistically nonsignificant result twice. This procedure is analogous to that used in estimation from stratified sample surveys; if a stratum is sampled with half the probability of occurrence in the population, it is weighted twice as much to obtain a valid population estimate (see Hedges, 1992).

Giving twice as much weight to the statistically nonsignificant results did not change the results of the combined significance tests. The summary table for the full analysis in each of the seven resource variables would be identical for the full analysis when all nonsignificant coefficients are counted twice. The publication bias robustness column in Table 5 provides evidence of the changes to the median effect sizes when a similar weighting procedure is employed. While the median value of the standardized regression coefficients remains the same in PPE, the effect sizes in teacher education (dichotomous subsample), teacher experience (continuous subsample), and teacher/pupil ratio are reduced by less than 10%. Larger reductions in median effect size were present in teacher ability (30%) and teacher salary and school size (each close to 50%).

The publication bias robustness column in Table 7 demonstrates that even when each nonsignificant effect is counted twice, the magnitude of the effects are still large enough to be important to educators. In four of the five categories for which such estimates are calculated, the difference between the full analysis and the robustness analysis is small. In teacher salary, the reduction is significant, as would be the changes in teacher ability and school size if data were available to provide a similar estimation of the effect of $500/student in these categories. Taken together, the robustness analyses in the combined significance testing and effect size estimate provide additional evidence of the positive and substantial relation between resource inputs and student achievement, even if nonsignificant results are disproportionately underrepresented in the research literature.

Connecting Education Production Functions With Societal Changes

Some have argued that since resources over the past few decades have increased and achievement appears to have declined, the two variables cannot be positively related. This argument necessarily assumes that everything else relevant to the cost of education and the production of student achievement has remained constant. This assumption is incorrect, as there have been important changes in the social capital available in families which substitute for school resources.

Are These Findings Inconsistent With Recent Achievement Trends?

Perhaps the best evidence on national achievement trends is provided by the National Assessment of Educational Progress (NAEP) trend data (Mullis et al., 1994). These are the only data that have been collected from nationally representative samples using the same design and measurement instruments during each wave of data collection since the early 1970s. The overall average achievement scores in reading and mathematics have increased slightly since 1971 and 1973, respectively, when NAEP trend data collection began. However, the overall means obscure important changes in scores. While the national average achievement of White students has remained fairly stable, the national average reading and math achievement of Blacks and Hispanics has increased by about one half of a standard deviation from the earliest trend assessment to the latest (in 1992). Overall achievement in science has declined somewhat since 1970, when the NAEP trend data collection began. The national average achievement for Whites in science is down a bit less than the overall average, and the average achievement of Blacks in science is essentially unchanged. However, the average achievement of Hispanics in science has actually increased somewhat.

Thus, overall achievement in the core academic subjects of reading and mathematics has actually increased in the last two decades, and the achievement of Blacks and Hispanics in these subjects has increased substantially. Achievement in science, which comprises a much smaller part of the elementary and secondary school curriculum, has decreased overall; but even in science, Hispanics have shown increases. It would appear that schools have produced modest increases in achievement for all students in the core academic areas of mathematics and reading, and produced rather substantial increases in the achievement of Blacks and Hispanics.

Social Capital

There is considerable research in psychology and education to support the hypothesis that home environment has very strong effects on student achievement, stronger in fact than social class effects (e.g., *The Home Environment*, 1993). The most important home environment variables involve a parent (or parents) expending time participating in or facilitating activities with children which enhance learning (reading with the child, playing games with educational content, helping with homework, etc.). These home environment variables have been characterized as *social capital* (see Coleman, 1987, 1988), since they reflect a consistent allocation of parental time and expertise to the child for the purposes of fostering greater achievement (investment of time on the part of parents to produce a return).

While there is little national data on the most important home environment or social capital variables, there are some indicators which are likely to be correlated with them. These indicators have shown marked declines in the last quarter century. One aspect of social capital is the amount of time mothers have to devote to their children, presumably some of which is devoted to informal educational activities. Maternal work competes for time with educational aspects of child rearing, and, consequently, social capital is decreased when mothers work. The percentage of children with mothers in the workforce has increased steadily over the last few decades, from 16% in 1950 to 59% in 1990 (Hernandez, 1994).

Presumably, the presence of a father in the home increases the total amount of adult time that can be deployed for all family activities, including educative activities in the home. Thus social capital would be smaller if no father were present. The proportion of children living with only their mothers has more than tripled since 1950, increasing from 6.4% in 1950 to 20.0% in 1990 (Hernandez, 1994). Finally, it might be expected that children born to unwed mothers would generally have fewer adult resources in the home and hence less social capital than children born to married couples. The proportion of births to unmarried women has more than tripled since 1960, increasing from about 5% to about 22% in 1986 (Hobbs & Lippman, 1990).

Each of these indicators provides evidence of a substantial decline in certain facets of social capital in the last few decades. There may be some evidence of changes in families, including family size, which may serve to counterbalance a portion of this decline (Grissmer, Kirby, Berends, & Williamson, 1994). Nonetheless, we believe that the overall effect may still remain negative. If social capital has declined, and if the declines are at least as pronounced for disadvantaged groups such as Blacks and Hispanics as for the population as a whole, then the increases in achievement of the latter groups are even more impressive. The fact that achievement has not declined substantially (and has increased substantially for some subgroups) is evidence for the positive effects of increasing school expenditures. We conclude that as family structures have changed and social capital has eroded, increases in school expenditures substitute for the informal educational resources we characterized as social capital investments. (We are not alone in drawing this conclusion; Flyer and Rosen,1994, have made precisely this argument.)

Conclusions

We believe that this represents a more significant contribution than our earlier work (Hedges et al., 1994; Laine et al., 1995) on the question of whether school resources and student achievement are related. The general conclusion of the meta-analysis presented in this article is that school resources are systematically related to student achievement and that these relations are large enough to be educationally important. Global resource variables such as PPE show strong and consistent relations with achievement. Smaller schools and smaller classes are also positively related to student achievement. In addition, resource variables that attempt to describe the quality of teachers (teacher ability, teacher education, and teacher experience) show very strong relations with student achievement.

While we believe these findings contradict the long-held beliefs of certain individuals, we also recognize that they may be considered by some as merely confirming the obvious. We ask that the reader, before adopting this perspective, consider an observation made by Bernd Heinrich (1989) as he described his attempt to understand corvid feeding behavior:

> Right from the start, I felt that after the solution was found, it would—as is usually the case—seem almost self-evident and then quickly be taken for granted. If it makes very *good* common sense, we say it is self-evident. After it fits into a theory (which is, after all, only formalized common sense), we feel that it could have been predicted. (p. 13)

While many would have hoped that increasing resources would be positively related to achievement, we did not expect that the synthesis of data from a wide variety of studies over a three decade period would yield conclusions so uniform in direction and comparable in magnitude.

While the findings of our research should provide a clear direction for policymakers—that money is positively related to student achievement—the results are not intended to specify the allocation of existing and new dollars in schools. The normative level of aggregation in the universe of studies we assembled measured inputs at the school level and outcomes at the student level. This is logical, as few school districts have developed financial data collection and reporting systems which track resources to the classroom, the level of greatest interest to those who study student learning. Thus the data available to address questions about optimal resource allocation remain far from ideal. A quarter of a century ago, Dreeben (1968) stated that it was necessary to understand classroom organization and instructional practice before addressing the question of what is learned in school. Though studies following Dreeben's criteria remain rare, we hope that a decade from now it will be possible to complete a research synthesis of primary studies measuring resource inputs at the classroom and student levels, and employing a diversity of outcomes.

We do not argue that money is everything. How we spend the money and the incentives we create for both children and teachers are equally important. We hope that our work is viewed as only one stage in assessing how schools function, as the universe of studies we assembled allowed us to address only a portion of the criteria Murnane (1981) identified as important to providing an accurate "snapshot of a school system at work": (a) using progress, rather than level of achievement, as the measure of school effectiveness (longitudinal studies address this); (b) including teacher and peer characteristics in the resource category; (c) measuring the resources reaching specific children, rather than school or district averages; (d) using the individual as the unit of observation; and (e) monitoring instructional techniques and time devoted to specific tasks. One consequence of employing relatively undifferentiated resource variables is that much of the policy debate has tended to be centered around the question, Does money matter?, rather than the more prescriptive question, How does money matter? It is only by addressing the latter question that local educators may meet the needs of their students by the most educationally and economically efficient means possible. The gap between the psychology literature, which frequently employs outcome measures that schools themselves do not typically utilize, and the economics literature, which often emphasizes outcomes with which schools (and often many businesses) may not be concerned, must be closed in order for educational research to have a greater impact on children. At the same time, we must recognize how schools continue to change, and how often the disjunction between experiences inside and outside of schools may be related to the socioeconomic status of students. Instead of reform without the possibility of enhanced resources, policymakers should advocate reform which incorporates high standards, continuing assessment, and adequate resources.

APPENDIX

Descriptions of education production function studies

Author(s)	YOD	Sample	Grade	Control	Inputs	Outputs
Baum	'70–71	46 SD	Sec	SES	PPE	Reading
Bieker & Anschel	'67–69	226 St	Sec	SES, Long	PPE	Composite
Boardman, Davis, & Sanday	'64–65 (EEO)	16,456 St	Sec	SES	TAb TExp T/P	Composite
Borland & Howsen	'89–90	170 SD	Elem	QL	TSal T/P	Composite
Bosshardt & Watts	'87	>3,000 St	Sec	Long	TExp SSize	TEL
Bowles	'64–65 (EEO)	1,000 St	Sec	SES	TAb	Verbal
Brown & Saks	'70–71	104,790 St	Elem	SES	TEd TExp T/P	Created comp (reading/math/ English)
Burkhead, w/Fox & Holland	'61–62 '60–61 '60–61	39 Sc 22 Sc 177 Sc	Sec Sec Sec	SES	TEd TExp T/P SSize	Reading Verbal Reading
Caldas	'89–90	737 Sc 468 Sc 96 Sc	Elem Sec Dual	SES	T/P SSize	Composite
Cohn	'62–63	377 Sc	Sec	QL	TSal T/P SSize	Composite
Cohn, Millman, & Chew	'71–72	53 Sc	Sec	SES	TEd TExp TSal T/P SSize	Created comp (verbal/math)
Deller & Rudnicki	'85–89	139 Sc	Elem	SES	PPE SSize	Composite
Dolan & Schmidt	'80–84	128 SD	Elem Sec	SES, QL	TSal T/P	Created comp (reading/math)
Dugan	'69–70	47 Sc	Sec	SES	PPE TEd TExp T/P	Created comp (reading/math)

APPENDIX *(continued)*

Author(s)	YOD	Sample	Grade	Control	Inputs	Outputs
Eberts & Stone (1987)	’78	9,468 (U) St 5,411 (NU) St	Elem	SES, Long	TEd TExp T/P	Math
Eberts & Stone (1988)	’78	14,959 St	Elem	SES, Long	TEd TExp SSize	Math
Ehrenberg & Brewer	’81–82	3,128 (W) St 1,055 (AA) St 549 (L) St	Sec	SES, Long	PPE TEd TExp T/P	Created comp (math/reading/ vocab)
Ferguson	’85–86	857–890 SD (2.4 million St)	Elem Sec	SES	TAb TEd T/P SSize	Reading
Fowler & Walberg	’84–85	293 Sc	Sec	SES	T/P SSize	Math
Grimes	’86–87	1,224 St	Sec	SES, QL	PPE T/P SSize	TEL
Grimes & Register	’86–87	1,626 St	Sec	SES, QL	PPE TExp SSize	TEL
Gyimah-Brempong & Gyapong	’86–87	152 SD	Sec	SES, QL	PPE T/P	Created comp (math/English)
Hallinan & Sorensen	Unk	451 St	Elem	SES, Long	T/P	Composite
Hanushek (1971)	’68–69	838 St	Elem	SES, Long	TAb TExp	Composite
Hanushek (1972)	’64–65 (EEO)	471 (W) Sc 242 (AA) Sc	Elem	SES	TAb TExp	Created comp (math/verbal)
Hanushek (1992)	’71–75	441 St	Elem	Long	TAb TEd TExp T/P	Reading
Harnisch	’82 (HSB)	18,684 St	Sec	SES, Long	TEd SSize	Composite
Heim & Perl	’67–68	63 SD	Elem	SES, QL	TEd TExp T/P	Reading and created comp (reading/math)

APPENDIX (continued)

Author(s)	YOD	Sample	Grade	Control	Inputs	Outputs
Jencks	’64–65 (EEO)	1,030 Sc	Elem	SES	T/P	Verbal
Katzman	’64–65	56 Sc	Elem	SES, QL	TEd TExp	Math or reading
Kenny	’59–60 (PT)	4,270 St	Sec	SES	TEd TExp T/P	Created comp (math/read/sci/ soc sci/verbal)
Kiesling (1967)	’57–60	79-121 SD	Elem Sec	SES, QL	PPE	Composite
Kiesling (1984)	’74	3,374 St	Elem	SES, Long	TEd TExp T/P	Reading
Levin	’64–65 (EEO)	597 (W) St	Elem	SES	TAb TExp	Verbal
Lewis & Ouellette	Unk	383 St	Elem	Long	TEd TExp	Social studies
Link & Mulligan (1986)	’76–77	103 (L) St 1,986 (W) St	Elem	SES, Long	TExp	Created comp (reading/math)
Link & Mulligan (1991)	’76–77	1,022 (L) St 10,871 (W) St	Elem	Long	T/P	Created comp (reading/math)
Link & Ratledge	’69–70	500 St	Elem	SES, Long	TEd TExp T/P	Reading
Lopus	’86–87	528 St (PPE) 655 St (other var.)	Sec	SES	PPE TEd TExp T/P	TEL
Maynard & Crawford	’71–72	18 Sc, 16 Sc	Elem Sec	SES, Long	PPE TEd TExp T/P SSize	Composite
Michelson (1970)	’64–65 (EEO)	458 (AA) St	Elem	SES	TEd	Verbal or math
Michelson (1972)	’70–71	110 Sc	Elem	SES	TEd TExp SSize	Reading
Monk	’89–90	1,492 St	Sec	Long	TEd TExp	Created comp (reading/math)

APPENDIX *(continued)*

Author(s)	YOD	Sample	Grade	Control	Inputs	Outputs
Murnane	'70–71 '71–72	410–440 (AA) St	Elem	SES, Long	TEd TExp T/P	Reading or created comp (reading/math)
Murnane & Phillips	'72–75	199–277 St	Elem	SES, Long	TEd TExp	Vocabulary
Perl	'59–60 (PT)	3,265 St	Sec	SES	PPE TEd TExp TSal T/P SSize	Created comp (verbal/abstract reasoning)
Register & Grimes	'86–87	1,570 (U) St 790 (NU) St	Sec	SES, QL	PPE T/P SSize	Composite
Ribich & Murphy	'59–60 (PT)	8,249-8,466 St	Sec	SES	PPE	Composite
Ritzen & Winkler	'64–65	217 St (AA) 194 St (W)	Elem	SES, QL	PPE	Reading
Sander	'89–90	154 Sc	Sec	SES	TSal T/P SSize	Composite
Sander & Krautmann	'86–87	102 counties	Sec	SES	PPE TEd TExp TSal T/P SSize	Composite
Schneider	'77–78	493 St	Elem	Long	TEd	Created comp (reading/math)
Sebold & Dato	'75–76	100 SD	Elem Sec	SES, QL	PPE	Reading or created comp (reading/math?/ writing)
Smith	'64–65 (EEO)	Unk	Elem Sec	SES	PPE TAb TEd TExp SSize	Verbal
Stern	'83–85	2,452-3,652 Sc	Elem	SES	TSal T/P SSize	Created comp (math/reading/ writing)

APPENDIX *(continued)*

Author(s)	YOD	Sample	Grade	Control	Inputs	Outputs
Strauss & Sawyer	'77–78	105 SD	Sec	SES	TAb T/P	Created comp (math/reading)
Summers & Wolfe	'70–71	627 St	Elem	SES, Long	TAb TExp T/P SSize	Composite
Walberg & Fowler	'83–84	261 SD	Sec	SES	PPE	Created comp (reading/math/ writing)
Walstad & Soper	'85–86	2,483 St	Sec	SES, Long	SSize	TEL
Winkler	'64–65	388 (AA) St 385 (W) St	Elem	SES, QL	T/P	Reading

Note. YOD = year of data (EEO = Equality of Educational Opportunity data, PT = Project Talent data, HSB = High School & Beyond data, Unk = Unknown).

Sample = sample size (Sc = schools, SD = school districts, St = students, AA = African Americans, L = Latinos, W= Whites, NU = non-unionized schools, U = unionized schools).

Elem = kindergarten through Grade 8, Sec = Grades 9–12, Dual = kindergarten through Grade 12.

SES = socioeconomic status, Long = longitudinal (pretest and posttest), QL = quasi-longitudinal (control for background but not pretest).

PPE = per-pupil expenditure, TAb = teacher ability, TEd = teacher education, TExp = teacher experience, TSal = teacher salary, T/P = teacher/pupil ratio, SSize = school size.

Created comp = In order to eliminate stochastic dependence, for production functions that utilized the same model and the same population of students but different outcomes, we calculated the median. TEL = Test of Economic Literacy.

References

References marked with asterisks indicate studies which provided data utilized in the meta-analysis.

Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J., & Pintero, J. (1972). *How effective is schooling? A critical review and synthesis of research findings* (Rep. No. R–956–PCSF–RC). Santa Monica, CA: RAND Corporation.

Baker, K. (1991). Yes, throw money at schools. *Phi Delta Kappan, 72,* 628–631.

*Baum, D. N. (1986). A simultaneous equations model of the demand for and production of local public services: The case of education. *Public Finance Quarterly, 14,* 157–178.

Begg, C. B. (1994). Publication bias. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York: Russell Sage Foundation.

*Bieker, R. F., & Anschel, K. R. (1973). Estimating educational production functions for rural high schools: Some findings. *American Journal of Agricultural Economics, 55,* 515–519.

*Boardman, A., Davis, O., & Sanday, P. (1977). A simultaneous equations model of

the educational process. *Journal of Public Economics, 7,* 23–49.
*Borland, M. V., & Howsen, R. M. (1992). Student academic achievement and the degree of market concentration in education. *Economics of Education Review, 11,* 31–39.
*Bosshardt, W., & Watts, M. (1990). Instructor effects and their determinants in precollege economic education. *Journal of Economic Education, 21,* 265–276.
*Bowles, S. (1970). Towards an educational production function. In W. L. Hansen (Ed.), *Education, income, and human capital* (pp. 11–61). New York: National Bureau of Economic Research.
Bridge, R. G., Judd, C. M., & Moock, P. R. (1979). *The determinants of educational outcomes: The impact of families, peers, teachers, and schools.* Cambridge, MA: Ballinger.
*Brown, B. W., & Saks, D. H. (1975). The production and distribution of cognitive skills within schools. *Journal of Political Economy, 83,* 571–593.
Bryk, A. S., & Raudenbush, S. W. (1988). Toward more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education, 97,* 65–108.
*Burkhead, J., Fox, T. G., & Holland, J. W. (1967). *Input and output in large-city high schools.* Syracuse, NY: Syracuse University Press.
*Caldas, S. J. (1993). Reexamination of input and process factor effects on public school achievement. *Journal of Economic Research, 86,* 206–214.
*Cohn, E. (1968). Economies of scale in Iowa high school operations. *Journal of Human Resources, 3,* 422–434.
*Cohn, E., Millman, S. D., & Chew, I. K. (1975). *Input-output analysis in public education.* Cambridge, MA: Ballinger.
Coleman, J. (1987). Families and schools. *Educational Researcher, 16*(6), 32–38.
Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94,* 95–120.
Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.
Cooper, H. M., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.
*Deller, S. C., & Rudnicki, E. (1993). Production efficiency in elementary education: The case of Maine Public Schools. *Economics of Education Review, 12,* 45–57.
Dickersin, K., & Min, Y–I. (1993, April 28). NIH clinical trials and publication bias [53 paragraphs]. *Online Journal of Current Clinical Trials* [On-line serial]. Available: Doc. No. 50
Dickersin, K., Min, Y–I, & Meinert, C. L. (1992). Factors influencing the publication of research results: Followup of applications submitted to two institutional review boards. *Journal of the American Medical Association, 267,* 374–378.
*Dolan, R. C., & Schmidt, R. M. (1987). Assessing the impact of expenditure on achievement: Some methodological and policy considerations. *Economics of Education Review, 6,* 285–299.
Dreeben, R. (1968). *On what is learned in school.* Reading, MA: Addison-Wesley.
*Dugan, D. J. (1976). Scholastic achievement: Its determinants and effects in the education industry. In J. T. Froomkin, D. T. Jamison, & R. Radner (Eds.), *Education as an industry* (pp. 53–83). Cambridge, MA: National Bureau of Economic Research.
Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *Lancet, 337,* 867–872.
Eberts, R. W., & Stone, J. A. (1984). *Unions and public schools: The effect of collective bargaining on American education.* Lexington, MA: D.C. Heath.

*Eberts, R. W., & Stone, J. A. (1987). Teacher unions and the productivity of public schools. *Industrial and Labor Relations Review, 40,* 354–363.

*Eberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference? *Economics of Education Review, 7,* 291–299.

*Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter?: Evidence from High School and Beyond. *Economics of Education Review, 13,* 1–17.

*Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28,* 465–498.

Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27,* 557–577.

Flanagan, J. C., Davis, F. B., Dailey, J. T., Shaycoft, M. F., Orr, D. B., Goldberg, I., & Neyman, C. A., Jr. (1964). *Project Talent: The identification, development, and utilization of human talents* (Final Report for Cooperative Research Project No. 635, U.S. Office of Education). Pittsburgh, PA: University of Pittsburgh.

Flyer, F., & Rosen, S. (1994). *The new economics of teachers and education* (Working Paper No. 4828). Cambridge, MA: National Bureau of Economic Research.

*Fowler, W. J., & Walberg, H. J. (1991). School size, characteristics, and outcomes. *Educational Evaluation and Policy Analysis, 13,* 189–202.

Glasman, N. S., & Biniaminov, I. (1981). Input-output analyses in schools. *Review of Educational Research, 51,* 509–539.

Glass, G. V., & Smith, M. E. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1,* 2–16.

*Grimes, P. W. (1994). Public versus private secondary schools and the production of economic education. *Journal of Economic Education, 25,* 17–30.

*Grimes, P. W., & Register, C. A. (1990). Teachers' unions and student achievement in high school economics. *Journal of Economic Education, 21,* 297–306.

Grimes, P. W., & Register, C. A. (1991). Teacher unions and Black students' scores on college entrance exams. *Industrial Relations, 30,* 492–500.

Grissmer, D.W., Kirby, S. N., Berends, M., & Williamson, S. (1994). *Student achievement and the changing American family.* Santa Monica, CA: RAND Corporation.

Guthrie, J. W., Kleindorfer, G. B., Levin, H. M., & Stout, R. T. (1971). *Schools and inequality.* Cambridge, MA: MIT Press.

*Gyimah-Brempong, K., & Gyapong, A. O. (1991). Characteristics of education production functions: An application of canonical regression analysis. *Economics of Education Review, 10,* 7–17.

*Hallinan, M. T., & Sorensen, A. B. (1985). Class size, ability group size, and student achievement. *American Journal of Education, 94,* 71–89.

*Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro-data. *American Economic Review, 61,* 280–288.

*Hanushek, E. A. (1972). *Education and race: An analysis of the educational production process.* Lexington, MA: D.C. Heath.

Hanushek, E. A. (1981). Throwing money at schools. *Journal of Policy Analysis and Management, 1,* 19–41.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24,* 1141–1177.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45–65.

Hanushek, E. A. (1991). When school finance "reform" may not be good policy. *Harvard Journal on Legislation, 28,* 423–456.

*Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100,* 84–117.

*Harnisch, D. L. (1987). Characteristics associated with effective public high schools. *Journal of Educational Research, 80,* 233–241.

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics, 9,* 61–85.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science, 7,* 246–255.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5–14.

Hedges, L. V., & Olkin, I. (1980). Vote counting methods in research synthesis. *Psychological Bulletin, 88,* 359–369.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Hedges, L. V., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal, 20,* 63–85.

*Heim, J., & Perl, L. (1974). *The educational production function: Implications or educational manpower policy* (Monograph No. 4). Ithaca, NY: Cornell University, Institute of Public Employment.

Heinrich, B. (1989). *Ravens in winter.* New York: Summit Books.

Hernandez, D. J. (1994). Children's changing access to resources : A historical perspective. *Social Policy Report, VIII,* 1.

Hobbs, F., & Lippman, L. (1990). *Children's well-being: An international comparison* (U.S. Bureau of the Census, International Population Reports, Series P-85, No. 8). Washington, DC: U.S. Government Printing Office.

The home environment and school learning: Promoting parental involvement in the education of children. (1993). San Francisco: Jossey-Bass.

*Jencks, C. S. (1972). The quality of data collected by "The equality of educational opportunity survey." In F. Mosteller & D. P. Moynihan (Eds.), *On equality of educational opportunity* (pp. 437–512). New York: Random House.

*Katzman, M. (1971). *The political economy of urban schools.* Cambridge, MA: Harvard University Press.

*Kenny, L. W. (1982). Economies of scale in schooling. *Economics of Education Review, 2,* 1–24.

*Kiesling, H. J. (1967). Measuring a local school government: A study of school districts in New York State. *Review of Economics and Statistics, 49,* 356–367.

*Kiesling, H. J. (1984). Assignment practices and the relationship of instructional time to the reading performance of elementary school children. *Economics of Education Review, 3,* 341–350.

Laine, R. D., Greenwald, R., & Hedges, L. V. (1995). Money does matter: A research synthesis of a new universe of education production function studies. In L. O. Picus (Ed.), *Where does the money go?: Resource allocation in elementary and secondary schools* (pp. 44–70). Newbury Park, CA: Corwin Press.

Levin, H. M. (1970). A new model of school effectiveness. In *Do teachers make a difference?: A report on recent research on pupil achievement* (pp. 55–78). Washington, DC: U.S. Government Printing Office.

*Levin, H. M. (1976). Concepts of economic efficiency and educational production. In T. Joseph, J. T. Froomkin, D. Jamison, & R. Radner (Eds.), *Education as an industry* (pp. 149–191). Cambridge, MA: National Bureau of Economic Research.

*Lewis, J., & Ouellette, H. (1979). Teacher training, teacher experience and pupil achievement. *Journal of Instructional Psychology, 6,* 3–5.

*Link, C. R., & Mulligan, J. G. (1986). The merits of a longer school day. *Economics*

of Education Review, 5, 373–381.
*Link, C. R., & Mulligan, J. G. (1991). Classmates' effects on Black student achievement in public school classrooms. *Economics of Education Review, 10,* 297–310.
*Link, C. R., & Ratledge, E. C. (1979). Student perceptions, I.Q. and achievement. *Journal of Human Resources, 14,* 98–111.
*Lopus, J. S. (1990). Do additional expenditures increase achievement in the high school economics class? *Journal of Economic Education, 21,* 277–286.
MacPhail-Wilcox, B., & King, R. A. (1986). Production functions revisited in the context of educational reform. *Journal of Education Finance, 12,* 191–222.
*Maynard, R., & Crawford, D. (1976). School performance. In D. L. Bawden & W. S. Harrar (Eds.), *Rural income maintenance experiment: Final report* (Vol. VI, Part II, pp. 1–104). Madison: University of Wisconsin, Institute for Research on Poverty.
McGiverin, J., Gilman, D., Tillitski, C. (1989). A meta-analysis of the relation between class size and achievement. *The Elementary School Journal, 90,* 47–56.
*Michelson, S. (1970). The association of teacher resourceness with children's characteristics. In *Do teachers make a difference?: A report on recent research on pupil achievement* (pp. 120–168). Washington, DC: U.S. Government Printing Office.
*Michelson, S. (1972). For the plaintiffs—equal school resource allocation. *Journal of Human Resources, 7,* 283–306.
Monk, D. H. (1989). The education production function: Its evolving role in policy analysis. *Educational Evaluation and Policy Analysis, 11,* 31–45.
Monk, D. H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Educational Evaluation and Policy Analysis, 14,* 307–332.
*Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13,* 125–145.
Mullis, I. V. S., Dossey, J. A., Campbell, J. R., Gentile, C. A., O'Sullivan, C., & Latham, A. S. (1994). *NAEP 1992 trends in academic progress.* Washington, DC: National Center for Education Statistics.
Murnane, R. (1981). Interpreting the evidence on school effectiveness. *Teachers College Record, 83* (1), 19–35.
*Murnane, R. J. (1975). *The impact of school resources on the learning of inner city children.* Cambridge, MA: Ballinger.
*Murnane, R. J., & Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research, 10,* 83–100.
*Perl, L. J. (1973). Family background, secondary school expenditure, and student ability. *Journal of Human Resources, 8,* 156–180.
*Register, C. A., & Grimes, P. W. (1991). Collective bargaining, teachers, and student achievement. *Journal of Labor Research, 12,* 99–110.
*Ribich, T. I., & Murphy, J. L. (1975). The economic returns to increased educational spending. *Journal of Human Resources, 10,* 56–77.
*Ritzen, J. M., & Winkler, D. R. (1977). The revealed preferences of a local government: Black/White disparities in scholastic achievement. *Journal of Urban Economics, 4,* 310–323.
*Sander, W. (1993). Expenditures and student achievement in Illinois: New evidence. *Journal of Public Economics, 52,* 403–416.
*Sander, W., & Krautmann, A. C. (1991). Local taxes, schooling, and jobs in Illinois. *Economics of Education Review, 10,* 111–121.
*Schneider, B. L. (1985). Further evidence of school effects. *Journal of Educational Research, 78,* 351–356.
*Sebold, F. D., & Dato, W. (1981). School funding and student achievement: An

empirical analysis. *Public Finance Quarterly, 9,* 91–105.
*Smith, M. (1972). Equality of educational opportunity: The basic findings reconsidered. In F. Mosteller & D. P. Moynihan (Eds.), *On equality of educational opportunity* (pp. 230–342). New York: Random House.
Spencer, B. D., & Wiley, D. E. (1981). The sense and the nonsense of school effectiveness. *Journal of Policy Analysis and Management, 1,* 43–52.
*Stern, R. (1989). Educational cost factors and student achievement in Grades 3 and 6: Some new evidence. *Economics of Education Review, 8,* 149–158.
*Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5,* 41–48.
*Summers, A., & Wolfe, B. (1977). Do schools make a difference? *American Economic Review, 67,* 639–652.
United States Center for Education Statistics. (1994). *Digest of education statistics, 1994.* Washington, DC: U.S. Government Printing Office.
*Walberg, H. J., & Fowler, W. J. (1987). Expenditure and size efficiencies of public school districts. *Educational Researcher, 16*(7), 5–13.
*Walstad, W. B., & Soper, J. C. (1988). A report card on the economic literacy of U.S. high school students. *American Economic Review, 78,* 251–256.
*Winkler, D. (1975). Educational achievement and school peer group composition. *Journal of Human Resources, 10,* 189–204.

Authors

ROB GREENWALD is Searle Fellow, Department of Education, University of Chicago, 5835 S. Kimbark Ave., Chicago, IL 60637; green@cicero.spc.uchicago.edu. He specializes in finance and litigation.

LARRY V. HEDGES is Stella M. Rowley Professor of Education and Social Science, Department of Education, University of Chicago, 5835 S. Kimbark Ave., Chicago, IL 60637; hedge@cicero.spc.uchicago.edu. He specializes in statistics and research synthesis.

RICHARD D. LAINE is Associate Superintendent for Policy, Planning and Resource Management, Illinois State Board of Education, 100 W. Randolph St., Suite 14-300, Chicago, IL 60601; rlaine@spr5.isbe.state.il.us. He is also a doctoral candidate in the Department of Education at the University of Chicago. He specializes in education policy, education finance, and academic standards.

Received November 21, 1995
Revision received April 4, 1996
Accepted June 20, 1996