Rochester Center for

Economic Research

Assessing the Effects of School Resources on Student Performance: An Update

Hanushek, Eric A.

Working Paper No. 424
August 1996

University of

Rochester

Assessing the Effects of School Resources on Student Performance: An Update

Eric A. Hanushek

Rochester Center for Economic Research
Working Paper No. 424
August 1996

Assessing the Effects of School Resources on Student Performance: An Update

by Eric A. Hanushek

August 1996

Abstract

The relationship between school resources and student achievement has been controversial, in large part because it calls into question a variety of traditional policy approaches. This paper reviews the available educational production literature, updating previous summaries. The close to 400 studies of student achievement demonstrate that there is not a strong or consistent relationship between student performance and school resources, at least after variations in family inputs are taken into account. These results are also reconciled with meta-analytic approaches and with other investigations on how school resources affect labor market outcomes. Simple resource policies hold little hope for improving student outcomes.

*This paper was completed within the context of continuing work with Stephen Rivkin and Lori Taylor. Financial support was provided by the William T. Donner Foundation.

Assessing the Effects of School Resources on Student Performance: An Update

by Eric A. Hanushek

Reflecting its policy significance, an enormous amount of research has focused on the relationship between resources devoted to schools and student performance. Recent interest generated by current policy debates have helped to clarify both the interpretation of this work and the resulting policy implications. This paper updates previous reviews of the literature and adds the perspective of the recent discussions of the results. With over three decades of analysis, new studies have reinforced earlier conclusions: today's schools exhibit continuing inefficiency in their operations as there is no strong or consistent relationship between variations in school resources and student performance. Alternative interpretations of the evidence plus apparently contradictory findings of different strands of this work can be reconciled in a straightforward manner with this conclusion. These results add further impetus for changing the focus of much of current policy development.

Overview of the Analysis of Educational Production Functions

The investigation of the effects of school resources began in earnest with the publication of the "Coleman Report" (Coleman et al., 1966). This Congressionally mandated study by the U.S. Office of Education startled many by suggesting that schools did not exert a very powerful influence on student achievement. Subsequent attention was directed both at understanding the analysis of the Coleman Report[1] and at providing additional evidence about the effects of resources.

The statistical analyses relevant to this work have a common framework that has been well-understood for some time (Hanushek 1979). Student achievement at a point in time is related to the

[1] These analyses suggested serious flaws in the statistical methodology and interpretation of the Coleman Report, but most of those discussions is not relevant for this discussion. See Bowles and Levin (1968), Cain and Watts (1970), Hanushek and Kain(1972).

primary inputs: family influences, peers, and schools. The educational process is also cumulative, so that both historical and contemporaneous inputs influence current performance.

With the exception of the Coleman Report, the subsequent analysis seldom has relied on data collected specifically for the study of the educational process. Instead, it has tended to be opportunistic, employing available data to gain insights into school operations. The focus of much of this work has been the effect of varying resources on student achievement. This focus flows from the underlying perspective of production functions; from its obvious relevance for policy; and from the prevalence of relevant resource data in the administrative records that are frequently used.

Over the past thirty years, a steady stream of analyses has built up a consistent picture of the educational process. This section describes the available studies, while the next considers the results. This summary concentrates on a set of published results available through 1994,[2] updating and extending previous summaries (Hanushek 1981, 1986, 1989). The basic studies meet minimal criteria for analytical design and reporting of results. Specifically, the studies must be published in a book or journal (to ensure a minimal quality standard); must include some measure of family background in addition to at least one measure of resources devoted to schools; and must provide information about statistical reliability of the estimate of how resources affect student performance. The objective was to collect information from all studies meeting these criteria in order to avoid any preselection problems.[3]

The summary relies on all of the separate estimates of the effects of resources on student performance. For tabulation purposes, a "study" is a separate estimate of an educational production

[2]The tabulations do include results in Hanushek, Rivkin, and Taylor (1996), since this updating was conducted as part of that research.

[3]The studies analyzed here include all studies contained in the prior review of 1989 along with a few that had been missed in that review and along with newly published studies. While some studies have undoubtedly been missed in this review, it is virtually impossible that the missed studies would alter the overall conclusions given the numbers of studies reported below.

function found in the literature. Individual published analyses typically contain more than one set of estimates, distinguished by different measures of student performance, by different grade levels, and frequently by entirely different sampling designs. If, however, a publication includes estimates of alternative specifications employing the same sample and performance measures, only one of the alternative estimates is included.[4] Thus, the 90 individual publications that form the basis for this analysis contain 377 separate production function estimates. While a large number of studies were produced as a more or less immediate reaction to the Coleman Report, half of the available studies have been published since 1985.

The studies are drawn from schools across the country, and contain information about a variety of measures of student outcomes. Table 1 provides an overview of the included studies. Three-quarters of the studies measure student performance by standardized tests, while the remainder use a variety of different measures including such things as continuation in school, dropout behavior, and subsequent labor market earnings. Not surprisingly, test score performance measures are more frequently employed for studying education in elementary schools, while a vast majority of the studies of other outcomes relate to secondary schools. Table 1 also displays the level of aggregation of the school input measures—an issue considered in detail below. One-quarter of the studies consider individual classrooms, while 10 percent measure school inputs only at the level of the state. Moreover, fully one-quarter of the studies employing nontest measures rely solely on interstate variations in school inputs.

---

[4]Some judgment is required in selecting from among the alternative specifications. As a general rule, the tabulated results reflect the estimates that are emphasized by the authors of the underlying papers. In some cases, this rule did not lead to a clear choice, at which time the tabulation emphasized statistically significant results among the alternatives preferred by the original author. An alternative approach is followed by Betts (1996). He aggregates all of the separate estimates of a common parameter that are presented in each individual paper.

Table 1. Distribution of outcome measures by schooling level and by aggregation level of school inputs

	Standardized test	Other measure	Total
Schooling level			
Elementary school	162	11	173
Secondary school	120	84	204
Aggregation level of school inputs			
Classroom	89	8	97
School	95	53	148
District	83	8	91
County	2	3	5
State	13	23	36
Total	282	95	377

Source: Author's tabulations.

The Impact of School Resources

Studies of educational performance include a variety of different measures of resources devoted to schools. Commonly employed measures include (1) the real resources of the classroom (teacher education, teacher experience, and teacher-pupil ratios); 2) financial aggregates of resources (expenditure per student and teacher salary); and, 3) measures of other resources in schools (specific teacher characteristics, administrative inputs, and facilities).

The real resource category receives the bulk of attention for several reasons. First, these best summarize variations in resources at the classroom level. Teacher education and teacher experience are the primary determinants of teacher salaries. When combined with teachers per pupil, these variables describe variations in the instructional resources across classrooms. Second, these measures are readily available and well-measured. Third, they relate to the largest changes in schools over the past three decades. Table 2 displays the dramatic increases in these school inputs, with pupil-teacher ratios falling steadily, teacher experience increasing, and the percent of teachers with a masters' degree actually doubling between 1960 and 1990. Fourth, studies of growth in performance at the individual classroom level, commonly thought to be the superior analytical design, frequently have these resource measures available but not the others.

The real resource stand in contrast with the other resource measures. The financial aggregates, particularly expenditure per pupil, are typically not even calculated for the classroom or the school, but instead are only available for the school district or for entire states. Thus, studies employing these are the most aggregated studies. They also tend to have relatively poor measures of family background, and studies focusing on spending are not amenable to value-added specifications (see below). In sum, these studies are noticeably lower quality that the best, and the typical, study

Table 2. **Public School Resources in the United States, 1961-1991**

Resource	1960-61	1965-66	1970-71	1975-76	1980-81	1985-86	1990-91
Pupil-teacher ratio	25.6	24.1	22.3	20.2	18.8	17.7	17.3
% teachers with master's degree	23.1	23.2	27.1	37.1	49.3	50.7	52.6
median years teacher experience	11	8	8	8	12	15	15
current expenditure/ADA (1992-93 $'s)	$1,903	$2,402	$3,269	$3,864	$4,116	$4,919	$5,582

Source: U.S. Department of Education[1994] investigating real classroom resources.[5] The measures of other school resources typically are measured poorly and tend to be available only at the district level. Since these resources tend to be relatively smaller in terms of overall spending, one would not expect these factors to be less important in determining student achievement.

Basic Results. Table 3 presents the overall summary of results. In terms of real classroom resources, only 9 percent of the studies considering the level of teachers education and 15 percent of the studies investigating teacher-pupil ratios find positive and statistically significant effects on student performance.[6] These relatively small numbers of statistically significant positive results are balanced by another set finding statistically significant negative results—reaching 13 percent in the case of teacher-pupil ratios. While a large portion of the studies merely note that the estimated coefficient is statistically insignificant without giving the direction of the estimated effect, those statistically insignificant studies reporting the sign of estimated coefficients are split fairly evenly between positive and negative. A higher proportion of estimated effects of teacher experience are positive and statistically significant: 29 percent. Importantly, however, 71 percent still indicate worsening performance with experience or less confidence in any positive effect. And, because more experienced teachers can frequently choose their school and/or students, a portion of the positive

---

[5]Some studies include expenditure per pupil along with measures of the real classroom resources. In such a case, since variations in classroom instructional expenditure are held constant, expenditure per student is interpreted as spending outside of the classroom. If only some of the classroom resources are included, the interpretation is more ambiguous and depends on the specific specification.

[6]The individual studies tend to measure each of these inputs in different ways. For example, while many studies include an indicator variable for whether or not the teacher has a master's degree, some will include measures of the graduate credits. With teacher-pupil ratio, some measure actual class size, while the majority measure teacher-pupil ratio. A variety of functional forms have been used, ranging from simple linear relationships to different nonlinear forms with thresholds, quadratics, and the like. In all cases, estimated signs are reversed if the measure involves pupil-teacher ratios or class size instead of teacher-pupil ratio. Further, where nonlinearities indicate positive effects over some range but not others, say with ranges of teacher experience, the most favorable for the hypothesis of positive effects is recorded.

Table 3. Percentage Distribution of Estimated Effect of Key Resources on Student Performance, Based on 377 Studies

Resources	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
Real classroom resources						
Teacher-pupil ratio	277	15%	13%	27%	25%	20%
Teacher education	171	9	5	33	27	26
Teacher experience	207	29	5	30	24	12
Financial aggregates						
Teacher salary	119	20%	7%	25%	20%	28%
Expenditure per pupil	163	27	7	34	19	13

Source: Author's tabulations.

effects could actually reflect reverse causation (Greenberg and McCall 1974; Murnane 1981). In sum, the vast number of estimated real resource effects give little confidence that just adding more the resource to schools will lead to a boost in student achievement. Moreover, this statement does not even get into whether or not any effects are 'large'. Given the small confidence in just getting noticeable improvements, it seems somewhat unimportant to investigate the size of any estimated effects.

The financial aggregates provide a similar picture. There is very weak support for the notion that simply providing higher teacher salaries or greater overall spending will lead to improved student performance. Per pupil expenditure has received the most attention, but only 27 percent of the estimated coefficients are positive and statistically significant. In fact, seven percent even suggestion some confidence in the fact that adding resources would harm student achievement. In reality, as discussed below, studies involving per pupil expenditure tend to be the lowest quality studies, and there is substantial reason to believe that even these results overstate the true effect of added expenditure.

Other measures. Outside of the basic resource factors, a vast number of specific measures of teachers and schools have been included at one time or another. Few measures have been repeated frequently enough to permit any sort of tabulation. One set of exceptions involves either administrative inputs or facilities. While these categories include a wide range of specific measures, the results of such investigation as tabulated in Table 4 show little consistent effect on student performance.[7] An additional exception is teacher test score, where teachers have been given some sort of achievement or IQ test and their score on that has been related to their students' performance.

---

[7]Administrative inputs are measured with such things as overall spending, the salaries of administrators, or the qualifications of administrators. Facilities include expenditures and specific measures such as availability of laboratories, the size and presence of a library, and the property of the school. In all cases, results are tabulated such that more of the measured characteristic means greater resources.

Table 4. Percentage Distribution of Other Estimated Influences on Student Performance, Based on 377 Studies

Resources	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
Teacher test scores	41	37%	10%	27%	15%	12%
Administrative inputs	75	12	5	23	28	32
Facilities	91	9	5	23	19	44

Source: Author's tabulations.

Table 4 displays the results of the 41 studies that include teacher test scores. Of all of the explicit measures that lend themselves to tabulation, stronger teacher test scores are most consistently related to higher student achievement, even though only 37 percent provide positive and statistically significant effects.

Aggregation. Studies vary widely in their design, in the character of the underlying samples and data that are available, and in their estimation approach. As displayed in Table 1, one of the most obvious differences relates to the aggregation of the underlying data. While the ideal analysis matches individual students with the school and family resources, this design is frequently precluded by the available data. In a fully-specified linear model, however, aggregation of explanatory variables reduces the precision of any estimates but does not lead to biased estimates. The problem arises there are either nonlinearities, such as interactions of school and family factors, or there are specification problems, such as omitted variables. There is no real expectation about the direction of any effect on estimates that might accompany aggregation of school resource variables.[8] While the next section offers evidence about the interaction of aggregation and specification errors, here we simply describe how the results vary with aggregation of the school resource measures.

Table 5 displays the distribution of studies by level of aggregation of the school resource measures and by the measurement of student outcomes. (This discussion is restricted to teacher-pupil ratios and expenditure per pupil because only five studies consider teacher education measured at the county or state level and only six consider teacher experience at that level). The unmistakable pattern of the results is that resources have a stronger positive influence and more frequently statistically significant as the level of aggregation increases from the school to the district and to the state.

[8] At the same time, aggregation is sometimes helpful. Specifically, when there is measurement error in the explanatory variables, aggregation can improve otherwise biased estimates. In the simplest cases of model misspecification or of errors-in-variables there are predictions about the direction of any biases, but these predictions break down in more complicated situations of multivariate models.

Table 5. Percentage Distribution of Estimated Effect of Teacher-Pupil Ratio and Expenditure per Pupil on Student Performance

Resources	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
A. Teacher-pupil ratio						
Total	277	15%	13%	27%	25%	20%
Classroom	77	12	8	18	26	36
School	128	10	17	26	28	19
District	56	21	16	39	20	4
County	5	0	0	40	40	20
State	11	64	0	27	9	0
B. Expenditure per pupil						
Total	163	27%	7%	34%	19%	13%
Classroom	4	0	0	0	0	100
School	83	17	7	35	23	18
District	43	28	9	37	26	0
County	5	0	0	40	20	40
State	28	64	4	32	0	0

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

Simply put, analyses at higher levels of aggregation are noticeably more likely to conclude that added resources (teacher-pupil ratios or overall spending) improve student performance. The influence of aggregation is especially dramatic when only state-to-state differences in resources are observed, and it is this pattern that leads to serious questions about the interpretation of the results.

State sampling. Overall policies toward schools are made at the individual state level.[9] Individual states through their state constitutions are responsible for providing public schooling and for setting the operating environment for schools. With the exception of Hawaii, all states delegate substantial responsibility for the provision of public schooling to local school districts, but they do so in a very constrained manner. State governments have developed elaborate rules and regulations dictating what local districts can and cannot do in the operations of schools, in the provision of specific programs, in the hiring and firing of teachers, and so forth. The states also govern how funds for schools are raised, including not only the split of responsibility between state and local jurisdictions but also the tax instruments that may be used. States further exert varying influence over the formation and operation of any private schools in the state. Additional variation in the operation of state schooling systems has come from court interpretations of state policies, most notably in the area of school finance. A majority of states have gone through court cases challenging their method of financing local schools based on the varying educational provisions of state constitutions.

Given the variations in policies across states and given the central importance that is frequently attached to modifying state education policies, it would not be surprising to find that state

---

[9]The federal government has always had a rather limited role, directed largely at specific programs and populations. The largest elementary and secondary programs involve money for compensatory programs (such as Title 1), vocational education, and funding for programs for handicapped populations. The federal government probably has a larger impact through laws and regulations (such as the Education for All Handicapped Children Act which determined requirements for special education). The federal judiciary, through its desegregation rulings, has also had enormous impacts on schools. Nonetheless, there is little reason to believe that these elements have had a particularly strong or biasing effect on the educational production process.

policies influence school performance. Unfortunately, little progress has been made at defining or measuring the most important aspects of state policy in terms of their effect on student performance or the efficiency of resource usage. If states that provide a higher level of funding also tend to have more productive policy environments, then a regression analysis that doesn't control for the policy environment will tend to exaggerate the effect of funding on performance.

The magnitude and even direction of any such bias is unknown a priori, because the bias depends on both the importance of variations in state policy and the correlations between state policies and school resources. The existing studies, however, permit some insight into the effects. Specifically, general state policies will have a common effect on each of the districts within a state, so that production function studies employing sample observations from within a given state will not suffer but studies drawing observations across states will.[10] Additionally, the effect of biases is not independent of the modeling strategy. Hanushek, Rivkin, and Taylor (1996) show that, as data are aggregated to level of the omitted variable (e.g., state average data are used when state level factors are left out), any bias must worsen.

Table 6 shows the combined effects of aggregation and of cross-state sampling on the estimated effects of schools. Of the 277 studies of teacher-pupil ratios, 157 come from single-state samples, while 120 are drawn from multiple states. Of the 163 studies of expenditure per pupil, 89 come from single-state samples with the remainder coming from multiple state samples. The multiple state samples are further divided into two groups: those with no intrastate variation in school resources (i.e., where resources are measured at the state level) and those with intrastate variation. Estimation that employs samples crossing states systematically suggest that resources are more important for student performance than those analyzing achievement within individual states. There

[10]The preceding statement assumes linear state effects. To the extent that state policies interact with inputs into the educational process in a nonlinear manner, within state estimates could also suffer biases.

Table 6. Percentage Distribution of Estimated Effect of Teacher-Pupil Ratio and Expenditure per Pupil by State Sampling Scheme and Aggregation

Level of aggregation of recources	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
A. Teacher-Pupil Ratio						
Total	277	15%	13%	27%	25%	20%
Single state samples[a]	157	12	18	31	31	8
Multiple state samples[b]	120	18	8	21	18	35
With within-state variation[c]	109	14	8	20	19	39
Without within-state variation[d]	11	64	0	27	9	0
B. Expenditure per pupil						
Total	163	27%	7%	34%	19%	13%
Single state samples[a]	89	20	11	30	26	12
Multiple state samples[b]	74	35	1	39	11	14
With within-state variation[c]	46	17	0	43	18	22
Without within-state variation[d]	28	64	4	32	0	0

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

a. Estimates from samples drawn within single states.
b. Estimates from samples drawn across multiple states.
c. Resource measures at level of classroom, school, district, or county, allowing for variation within each state.
d. Resource measures aggregated to state level with no variation within each state.

are consistently more positive and more positive and statistically significant estimates from the multiple state samples, while there are noticeably fewer negative and statistically significant estimates. Moreover, the apparent importance of resources increases with aggregation, just what the theory suggests in the case of misspecification at the state level. The fact, however, that positive bias is present in more disaggregated studies that draw multiple state samples provides clear evidence that omission of measures of state policies is important.

Study Quality and Value-added models. One of the concerns about summarizing literatures, particularly in the tabular way done here, is that no weight is given to study quality. Indeed, in selecting studies for tabulation an effort was made to collect the entire universe of studies which met the minimal publication, specification, and reporting criteria. While this approach was taken to minimize any concerns that selection of studies led to the results, it opens the possibility of including low quality studies that might bias the overall results.[11]

One class of studies—those employing a value-added specification—is generally regarded as being conceptually superior and likely to provide the most reliable estimates of education production functions. These studies relate an individual's current performance to the student's performance at same prior time and to the school and family inputs during this intervening time. The superiority of this approach comes from the use of prior achievement to ameliorate any problems arising from missing data about past school and family factors and from differences in innate abilities of students (Hanushek 1979).[12]

---

[11]For an analysis of how study selection affects the summary of studies, see Hanushek 1996a.

[12]A related group of studies employs synthetic cohorts. These studies do not match current and past performance of the same students but instead either adds performance of current students in earlier grades or of students of the same vintage in prior grades. The first approach has none of the features that lead to preferring value-added studies, since past family, past school, and ability effects are not accurately accounted for. The second approach, which would be appropriate if there were no student mobility, leads to substantial errors with in and out movements of students. Moreover, the errors will generally be correlated with socio-economic and school factors, since these are related to

Table 7 provides a summary of value-added results, both for the 96 total separate estimates of resource effects and for the 39 estimates that come from samples in a single state. Clearly, these estimates are very much reduced from the overall set that is available, and thus any conclusions are subject to more uncertainty just due to limited number of underlying investigations. On the other hand, because of the superiority of these analyses, each study deserves more weight than one of the general studies reviewed previously.

These results strongly underscore the lack of effectiveness of general policies to increase teacher-pupil ratios or to hire more teachers with master's degrees or other graduate work. Within the single-state value-added studies, only four percent (i.e., one out of 23 estimates) of the studies of teacher-pupil ratios and none of the 33 studies of teacher education indicate a positive and statistically significant impact on student performance. The reduced sample of studies also lessens the apparent relationship with teacher test scores. The only resource input faring as well in the value-added studies as in the general data base in teacher experience. One would expect that inclusion of prior student achievement would reduce the importance of any reverse causation, so the value-added studies suggest that teacher choice is not driving the relative strength of teacher experience.

The refined analyses included in these higher quality studies strengthens the view that resources are not closely related to student performance. The lack of high quality studies for expenditure per pupil also figures into the preference for considering the results of the real resource models over the aggregate expenditure per pupil results. The expenditure models are almost always aggregated analyses, often lacking very detailed measures of family backgrounds, and estimated in level versus value-added form. This analysis indicates that the results from expenditure studies, weak as they might be, tend to overstate the true effects.

---

mobility behavior. See Hedges, Laine, and Greenwald (1996); Hanushek (1996a).

Table 7. Percentage Distribution of Other Estimated Influences on Student Performance, Based on Value-added Models of Individual Student Performance

Resources	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
A. All studies						
Teacher-pupil ratio	78	12%	8%	21%	26%	35%
Teacher education	40	0	10	35	30	25
Teacher experience	61	36	2	31	20	11
Teacher test score	11	27	9	18	27	18
b. Studies within a single state						
Teacher-pupil ratio	23	4%	13%	30%	39%	13%
Teacher education	33	0	9	33	27	30
Teacher experience	36	39	3	22	17	19
Teacher test score	9	22	11	11	33	22

Source: Author's tabulations.

Interpretation of Results

These results have a simple interpretation: There is no strong or consistent relationship between school resources and student performance. In other words, there is little reason to be confident that simply adding more resources to schools as currently constituted will yield performance gains among students. This finding has a series of obvious policy implications, but, before turning to these, it is useful to clarify precisely what is and is not implied by these data.

Perhaps the most important fact to underscore is that this finding does not imply that all schools and teachers are the same. Quite to the contrary. Substantial evidence suggests that there are large differences among teachers and schools.[13] The simple fact remains that these differences are not closely related to teacher salaries or to resources devoted to programs. The Coleman Report, which found that measured school resources explained a small portion of the variance in student achievement, has been commonly interpreted as implying that 'schools don't make a difference.' This latter interpretation confused the effects of measured differences with the full effects of schools and has been shown to be wrong. There is a significant difference between measured resources (of the kind on which policy frequently focuses) and the true effects of schools. In fact, it is just this difference between true effects and those of standard resources that forms the basis for the policy considerations below.

The previous evidence about the effectiveness of resources is readily interpreted as indicating that there is a distribution of underlying resource parameters. In some circumstances resources are used effectively, but these are balanced by others which indicate ineffective use. The interpretation is

---

[13]The clearest evidence comes from a series of covariance, or fixed-effects, estimates of performance differences across teachers (e.g., Hanushek 1971, 1992; Murnane 1975; Murnane and Phillips 1981; Armor et al. 1976). To give some indication of the order of magnitude, the estimated difference between a "good" and a "bad" teacher in poverty schools of Gary, Indiana, was approximately one grade level per academic year; i.e., a student with a good teacher might progress at 1.5 grade equivalents in a school year, while those with a bad teacher might progress at 0.5 grade equivalents.

easiest to see from the overall distribution of results about parameter estimates in Tables 3-7. If the effect of resources were always zero and a series of valid estimates were obtained across a group of studies, one would expect to find the null hypothesis of no effect rejected five percent of the time (for a 95 percent significance level) with 2 1/2 percent of the studies finding a positive and statistically significant effect and 2 1/2 percent finding a negative and statistically significant effect. In fact, there are uniformly more positive and more negative rejections (except in the high quality studies of Table 7). While there are other explanations, ones that probably contribute some to the results, it seems plausible that some schools and districts find productive uses of added resources and use extra resources to boost the performance of their students.

The concern from a policy viewpoint is that nobody can describe when resources will be used effectively and when not. In the absence of such a description, providing these general resources to a school implies that sometimes resources might be used effectively, other times they may be applied in ways that are actually damaging, and most of the time no measurable student outcome gains should be expected. This heterogeneity of results in the current system guides the policy discussion below.

The other possible explanations of the "fat tails" of the distribution of estimates deserve consideration. The first is publication bias. Hedges' 1990 summary of his prior research and that of others is instructive.

> The published literature is particularly susceptible to the claim that it is unrepresentative of all studies that may have been conducted (the so-called publication bias problem). There is considerable empirical evidence that the published literature contains fewer statistically insignificant results than would be expected from the complete collection of all studies actually conducted. There is also direct evidence that journal editors and reviewers intentionally include statistical significance among their criteria for selecting manuscripts for publication. The tendency of the published literature to overrepresent statistically significant findings leads to biased overestimates of effect magnitudes from published literature, a phenomenon that was confirmed empirically by Smith's study of ten meta-analyses, each of which presented average effect size estimates for both published and unpublished sources. [references omitted] (Hedges (1990, p. 19)

For this discussion, it does not matter whether individual researchers tend to search for 'statistically significant' results or whether journals are biased toward accepting them. In any event, the distribution of results would no longer reflect unbiased statistical tests, and the published results underlying the summaries in Tables 3-7 would overstate the magnitude and significance of each of the resource effects.[14]

The second explanation was alluded to previously. If the estimates are biased, say through misspecification of the underlying relationship, a factor can appear important even though it has no effect on student performance. Its perceived importance and statistical significance will depend on the strength of the omitted factor and on its sample relationship with included resource measures (which will vary from sample to sample). In other words varying specification bias could be driving part of the underlying distribution of estimated effects. This situation corresponds, for example, to the omission of measures of state differences in school regulations and policies which has different effects on the estimates depending upon the aggregation of the resource measures and upon whether samples are drawn across states.

Neither explanation for the observed distributions of resource effects provides more support for the importance of resources.

Controversies about Resource Effects

The preceding interpretations of the general ineffectiveness of school resource policies has been challenged. These challenges are outlined and evaluated here.

Labor Market Outcomes. Taken as a group, the production function studies give little indication that variations of resources have anything to do with present variations in student

[14]It is possible to ignore publication bias in the interpretation here—because publication bias works against the 'no effect' conclusion. The same is not the case when one is working go establish that resource variations are important, as in Hedges, Laine, and Greenwald (1994, 1996).

performance. However, the widely-publicized findings of Card and Krueger (1992a) indicate variations in school resources are related to earnings differences among workers.[15] Several issues could contribute to reconciling these conclusions: differences in levels of resources considered; differences in measurement of student performance; differences in specification; and aggregation bias in the statistical analysis.

The Card and Krueger (1992a) analysis begins with samples of adult workers from the 1970 and 1980 Censuses of Population and fills in information about the schooling circumstances of individuals from information about their year and state of birth. The workers in their sample attended schools between the 1920s and the 1970s, implying variations in the level of resources going far beyond what is found today. This suggests one reconciliation: if added resources have diminishing effects on student achievement, current school operations may be largely "on the flat" of the production function, while Card and Krueger observe ranges from the past where resources had stronger effects.[16] A related possibility might be that the political economy of schools has changed over time. For example, with the rise of teachers unions and the resulting change in bargaining positions, resources might be used in different ways and have different student achievement implications now than in the past (e.g., Hoxby 1996). In other words, it is quite possible that the enormous changes in educational resources did have an effect on outcomes in the first half of this century, but that more recent studies are also correct in finding "no effect" for the sorts of resource changes discussed in current schools.

---

[15]The Card and Krueger (1992a) analysis of school resources and earnings is the most discussed, but it follows a larger line of research. See, for example, Welch (1966), Johnson and Stafford (1973), and Wachtel (1976). An insightful review of past studies that considers underlying characteristics of the studies is Betts (1996).

[16]While not a direct test of this on-the-flat thesis, the lack of significantly stronger resource effects in developing countries introduces some question about this hypothesis; see Hanushek (1995), or, in a growth context, Hanushek and Kim (1996).

A second suggested reconciliation revolves around the measurement of outcomes. The previously compiled production function estimates are heavily weighted toward analyses of standardized test scores, while the Card-Krueger analysis concentrates on labor market earnings.[17] It is possible that schools do not affect test performance of students, but do affect skills and earnings. As Burtless (1996) points out, it seems implausible that schools do not affect what they explicitly are attempting to do (improve test performance) but do affect earnings, something they seldom measure or even consider a direct objective. The previous conclusions from production function estimates, however, hold equally when results are divided between studies of which use test scores as measure of outcomes and other measures of outcomes like college continuation or earnings. This can be seen in Table 8 that presents the available studies for expenditure per student divided by the measure of outcomes. Both the lack of general effects and the biases with aggregation hold regardless of outcome measurement.

One specific issue has received extra attention and is emphasized by Card and Krueger (1996). Do high resource schools encourage students to stay in school longer (which has obvious impacts on earnings)? Answering this question is, perhaps, more difficult than the straight achievement question, because labor market opportunities will affect the school completion decision as will net tuition and parental financial support when contemplating college. That question is a focal point of Hanushek, Rivkin, and Taylor (1996). In that study of school completion, school resources

[17]An important specification issue is that Card and Krueger (1992a) attempts to distinguish between the effects of schooling inputs and the effects of being in different local labor markets by assuming that migration across regions is nonselective. This assumption, however, runs counter to standard economic models, and, as Heckman, Farrar, and Todd (1996a, 1996b) demonstrate, counter to the data. Thus, the data do not support a key identifying condition for the Card-Krueger estimation of school resource effects.

Using a different methodology, however, they do find that school resources appear important in explaining differences in black earnings after the end of segregation (Card and Krueger 1992b). The level of resources are lower and the differences in resources higher in that study than in current situations, again suggesting that resources may matter at very low levels.

Table 8. **Percentage Distribution of Estimated Effect of Expenditure per Pupil on Student Performance by Outcome Measure and Aggregation of Resource Effects** (163 estimates)

Outcome measure	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
A. Test Score Outcomes[a]						
Total	109	25%	9%	28%	21%	17%
Classroom	4	0	0	0	0	100
School	57	19	9	28	21	23
District	38	26	11	37	26	0
County	2	0	0	0	50	50
State	8	75	13	13	0	0
B. Other (Nontest) Outcomes[b]						
Total	54	31%	2%	46%	15%	6%
School	26	12	4	50	27	8
District	5	40	0	40	20	0
County	3	0	0	67	0	33
State	20	60	0	40	0	0

Note: Rows may not add to 100 because of rounding.
a. All studies measure student performance by some form of standardized test score.
b. All studies employ some outcome measure (such as income or school attainment) other than a standardized test score.

have no significant impact on student behavior once individual achievement and school costs are considered.[18] Betts (1996) reviews a number of these studies and does suggest some positive effects of resources. For the studies tabulated here (which differ from those considered by Betts), there tend to be positive effects of expenditure on school attainment, but there are only 25 total studies and only five estimated from within individual states.[19] Thus, the small samples make it difficult to resolve this issue conclusively.

Moreover, there is considerable evidence that test scores are increasingly related to labor market performance (for example, O'Neill, 1990; Bishop, 1991; Grogger and Eide, 1993; Murnane, Willett, and Levy, 1995; Neal and Johnson, 1996). It seems unlikely that school resources affect just the component of earnings that is uncorrelated with cognitive skills. Moreover, school resources are not consistently related to earnings (Betts, 1996). This finding is particularly clear when direct measures of the school resources relevant to individuals are available (Betts, 1995; Grogger, 1996). As an overall summary, the lack of relationship with school resources is more generally true for recent studies of earnings than earlier investigations, while more recent studies have tended to find stronger effects of cognitive skills on earnings.

The final set of reasons that could help to explain the different conclusions involve specification issues. To begin with, many of the direct analyses of earnings include just the level of school resources, but none of the other factors that might influence student achievement and skill development. For example, it is plausible that students attending schools with a high level of

[18]The major focus of that paper is the effect of aggregation of school resource data. At the individual school level, school resources have no significant impact on completion and frequently even have the wrong sign. Aggregation to the state level does boost the apparent significance of school resources, but this is entirely explained by increased bias with model misspecification.

[19]One might expect state effects to be particularly important in determining school continuation, since the availability and expense of public colleges and universities and the opportunity costs implied by different local labor markets vary significantly across states.

resources also have parents who contribute more time, energy, and money to their education. If parental inputs are left out of the calculation, any estimated effects of school resources would tend to overstate the true independent effect of resources. Further, as pointed out above, aggregation of school inputs is also likely to exacerbate any biases due to specification issues (Hanushek, Taylor, and Rivkin 1996). Most of the earnings analyses observe school resources measured only at the aggregate state level. The Card-Krueger estimates come from resource data aggregated to the state level, but no measures of state policy differences are included, so their estimates are subject to this bias.

The end result of this comparison is that the estimates of Card and Krueger (1992a, 1992b) at most suggest that very low levels of resources -- say those found in the poorest states before and during the Great Depression or in segregated school systems -- may have an affect on student outcomes. But there is little reason to believe that this conclusion offers helpful policy advice given the current levels of resources.

Meta-Analysis and the Summary of Results. In some research areas, like considering the health effects of a certain drug therapy, there is frequently an interest in compiling results from a variety of trials. Specialized techniques to combine the results of separate studies and thus assess the magnitude and significance of some relationship have been developed. These approaches go under the general title of "meta-analysis." Quite clearly, the preferred approach to assessing disparate results would involve combining the underlying data of the studies directly to develop statistical inferences and tests of hypotheses across the studies. Unfortunately the original data are seldom available for reanalysis—and even when they are, combining data from different sources can be difficult—which forces a variety of compromises in the aggregation of results. The previous data on studies in Tables 3-8 represent one approach to the aggregation of results, an approach which relies on the minimal set of factors standardly reported. But, instead of simply reporting the distribution of results--which is, sometimes derisively, called vote-counting in the meta-analysis literature—others have attempted to do formal statistical tests.[20]

A well-known version of applying formal statistical tests to education production function data is found in a papers by Hedges, Laine, and Greenwald (1994, 1996). They wish to do formal hypothesis testing using the available data from essentially the same set of published studies employed here. Some of the problems with doing this are immediately evident. Combining testing information is best motivated from thinking about a series of independent laboratories all providing results from a simple, common experiment. But, the education production function estimates are far from a series of independent laboratories producing estimates of a single common parameter. Published estimates pursue a variety of different modeling strategies, so it is hard just to define a common parameter in a way that is susceptible to formal testing. More importantly, published articles frequently do not (and cannot) provide sufficient information. For example, if parameter estimates are correlated across studies, say because they reflect performance in different grades of one school district, estimation of the combined variance of the estimator would require knowledge of the covariances--something that is never provided. To be sure, such problems enter into the distributional tabulations previously presented, but they are clearly less central to the interpretation of the results than in the case of combined significance testing.

The most basic problem with their statistical analysis, however, is that it addresses a completely uninteresting question—one that has little relevance from a policy viewpoint. They suggest that the central hypothesis is whether "money matters." In reality, the question they pose is whether there is any evidence that resources or expenditure differences *ever, under any circumstances*

[20]The primary argument against vote-counting derives from the stylized analysis of combining a series of small experiments employing tests with low power, where more studies can actually lead to false conclusions. These examples have little relevance to the statistical tests developed in a regression framework with the very large samples frequently available.

appear to affect student performance. The formal statement is clear when they test the null hypothesis that all parameters indicating the effect of a specific resource on student performance are simultaneously equal to zero; i.e., H_o: $\beta_1 = \beta_2 = \ldots = \beta_n = 0$, where the β_i are the underlying parameters relating a specific resource to student performance in one of the n available studies. If any single underlying parameter (i.e., one β_i) for the combined sample of studies across varied schooling circumstances is not zero, the null hypothesis is false (that is, someplace there is an effect on student performance). Their statistical procedures are designed in such a case to reject the null hypothesis, leading to acceptance of the alternative that at least one study indicated someplace the resource was related to performance.[21]

The obvious interpretation of the previously reported results, as discussed above, is that there is a distribution of underlying parameters which tends to be centered close to zero. But, even if the distribution were exactly centered on zero and it were very tightly distributed around zero, the methods of Hedges et al. are designed to reject the null hypothesis that all of the underlying parameter values are zero.[22]

Their formal tests lead to rejection of this restricted null hypothesis.[23] These results are sometimes interpreted as a refutation of the conclusion that educational inputs don't affect performance. But in my view, this work both confirms the previous substantive results and points to the same policy conclusions. By thinking of an underlying distribution of resource parameters,

---

[21]In discussing precisely the issue of how to interpret rejection of this null hypothesis, Hedges and Olkin (1985, p. 45) state, "It is doubtful if a researcher would regard such a situation as persuasive evidence of the efficacy of a treatment."

[22]The actual application of the specific tests they employ requires a number of severe restrictions. One crucial is the reliance on very selective samples that are biased toward resource effects. The sampling is discussed in the Appendix.

[23]Note that the precise testing depends importantly on their specific choice of statistical methods and on their selective sampling of available results; see the appendix to this paper.

attention in focused naturally on the need for an appropriate structure of the educational environment to ensure that added resources deliver positive effects. As all of the analysis shows, productive results are possible, even if seldom achieved currently. But, understanding that there is an underlying distribution of effects highlights the inappropriateness of simple resource policies within the context of current schools.[24]

STAR Experiment. In the mid1980s, because of ambiguity about the effects of class size on student performance, the State of Tennessee launched a random-assignment experiment in reducing class sizes. The design was heavily influenced by an early summary of research by Glass and Smith (1979). That study suggested that student achievement was roughly constant across class sizes until the class size got down to approximately 15 to 1. After 15 to 1, reductions in class size yield gains in student performance. Based on this, a group of kindergarten through third graders were randomly assigned to either large classes (22-24 students) or small classes (14-16 students).[25] Students were followed from kindergarten through third grade.

The student testing shows that children in smaller classes did better at the end of kindergarten and that this better performance was maintained through the third grade. The key to interpretation revolves around expectations about student performance over time. One view is that education is a cumulative process, building on past achievement. From this view, if a student learns certain skills in the first grade, they tend to carry over to later grades, albeit possibly with some depreciation.

---

[24]In addition to conducting the combined hypothesis tests, they attempt to provide estimates of the magnitude of any resource effects. They concentrate most of their attention on expenditure per pupil, which is unfortunate because these studies tend to be the weakest of all of the available studies. After considerable manipulation of the sample of studies (see appendix), they do estimate that there is a positive median effect of expenditure on test scores. These estimates are, however, quite inconsistent with aggregate spending and test performance (Hanushek 1996b) and do not change any policy conclusions.

[25]The design was actually more complicated. The large classes were broken into two groups, one with teacher aides and one without aides. To be eligible for participating in the experiment, a school also had to be large enough so as to ensure that there was at least one small and large class.

According to this view, the basic evidence of the STAR study suggests that smaller classes may be important at kindergarten but have no average effect subsequently. Specifically, since the growth in achievement across experimental and control students is the same from first through third grade, the added resources of the smaller classes appear to add nothing to student performance.

The alternative expectation is that students are expected to fall back to a common mean performance each year. This is equivalent to a view that educational performance is not cumulative. Under this set of expectations, maintaining the difference in performance at the end of kindergarten requires continuing application of additional resources. The way to identify the effects of class size in the presence of these alternative interpretations would be to assign some of the experimental children to larger classes after the earliest grades. Unfortunately, this was not done within the experiment. However, follow-ups of these students, after they had returned to regular class settings, showed that they maintained a large portion but not all of the prior differences (Mosteller 1995). This latter finding supports the general cumulative model and indicates that class size reductions after kindergarten have little potential effect on achievement.

The Tennessee experiment does focus attention on earlier grades. The earlier discussion here looked across all grades and could mask differences between earlier and later schooling. To consider this possibility, the previous estimates of the effects of teacher-pupil ratios are divided into elementary and secondary schools. As Table 9 shows, there is little difference between the estimated effects in elementary and in secondary schools, but, if anything, there is less support for increasing teacher-pupil ratios at the elementary level. This evidence does not, however, restrict attention just to the earliest grades as the STAR experiment suggests should be done.

The experimental approach has obvious advantages in situations like this where the treatment--smaller classes--is well-defined and easily implemented. It is unfortunate, given the policy attention

Table 9. Percentage Distribution of Estimated Influence of Teacher-pupil on Student Performance, By Level of Schooling

School level	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
Elementary schools	136	13%	20%	25%	20%	23%
Secondary schools	141	17%	7%	28%	31%	17%
All schools	277	15	13	27	25	20

Source: Author's tabulations.

devoted to class size issues, that there has been no serious follow-up of the STAR experiment with similar experiments.

Policy Implications

These conclusions clearly imply that educational is policy making more difficult than many would like. If resources had a consistent and predictable effect on student performance, policy making would be straightforward. State legislatures could decide how much money to invest in schools and could trust local districts to apply funds in a productive manner. But, the fact that local districts do not use funds effectively complicates this picture. The clearest message of existing research is that uniform resource policies will not work as intended.

An alternative view is that the actual educational approaches can be set centrally. This is consistent with a widely held view that 'what works" is known. For example, Smith, Scoll, and Link (1995) unequivocally states just that. (At the same time, they are totally unsurprised and unconcerned that what works is unrelated to the resources devoted to schools, simply noting that "How money is spent is far more important than how much is spent.") This statement about knowing what works is quite consistent with the myriad of articles and policy prescriptions that promote this or that plan as the panacea. If one believes this perspective, however, it implies that local school administrators are either uncaring or simply don't know what works—implying that just providing better dissemination of information will effectively correct the problems. In reality this is a scathing indictment of today's schools because it implies rather widespread malfeasance.

Similar policy dilemmas face the courts in school finance cases. The courts have entered into education decision making in deciding on suits brought by people who believe that a state legislature are not fulfilling its constitution obligations to provide equitable or sufficient education to identified students in a state. For the most part, both the statement of the issue and the proposed remedies revolve around the level and distribution of resources. If resource availability is not a good measure of educational outcomes or if providing for overall resource levels do not affect performance, the courts face the same dilemma as legislatures. They also are not better off by having people suggest that specific legislated programs would solve any identified problems.

This policy conundrum is precisely what led the Panel on the Economics of Education Reform to concentrate not on the specific resources and policies of schools but on the incentive structure. Its report, *Making Schools Work*, emphasizes the need to radically alter current incentives in schools (Hanushek with others, 1994). The simple premise is that the unresponsiveness of performance to resources is largely a reflection that very little rests on student performance. Good and bad teachers or good and bad administrators can expect about the same career progressions, pay, and other outcomes—making the choice of programs, organization, and behaviors less dependent on student outcomes than on other things that directly affect the actors in schools.

Underlying this view is a more benign opinion of school personnel. Specifically, school personnel are not just ignoring a set of policies that would lead to obvious improvements but are simply following existing incentives. An added part of this argument is that the kinds of policies that will work in given situations with given personnel and students vary and are not easily described and centrally regulated. Given better incentives, school personnel will be interested in searching out what will work in their specific situation. Given current incentives, they may devote their attention and energies elsewhere.

The previous work on educational production has provided substantial evidence that there are vast differences among teachers and schools. It is just that these differences are not easily described by the resources employed or by any simple set of programmatic or behavioral descriptions. The existence of effective teachers and schools, however, implies that one approach to policy is concentrating on ways to reward better performance whenever it is found. In other words, even if the details of what will work are unavailable before the fact (or even after the fact), policy can be described in terms of outcomes, and good outcomes can be rewarded.

Such a description is itself much too simple, because we have limited experience with alternative incentive schemes (Hanushek with others 1994). Nonetheless, current approaches appear to offer little hope of improvement—for we actually have considerable experience with the current organization. And there is considerable experience in settings outside of schools that suggests incentives can have powerful effects.

The existing work does not suggest resources never matter. Nor does it suggest that resources could not matter. It only indicates that the current organization and incentives of schools do little to ensure that any added resources will be used effectively. Faced with this, some simply declare that we should still pursue general resource policies but we should not pursue programs that do not work. This would be fine, if programs that do and do not work could be reliably identified by policy makers. We do know that they have not been accurate in their past identification.

General References

Armor, David, Patricia Conry-Oseguera, Millicent Cox, Niceima King, Lorraine McDonnell, Anthony Pascal, Edward Pauly, and Gail Zellman. *Analysis of the school preferred reading program in selected Los Angeles minority schools*. Santa Monica, CA: Rand Corp., 1976.

Betts, Julian R. "Does school quality matter? Evidence from the National Longitudinal Survey of Youth." *Review of Economics and Statistics* 77, no. 2 (May 1995): 231-47.

---. "Is there a link between school inputs and earnings? Fresh scrutiny of an old literature." In *Does money matter? The effect of school resources on student achievement and adult success*, edited by Gary Burtless, 141-91. Washington, DC: Brookings, 1996.

Bishop, John. "Achievement, test scores, and relative wages." In *Workers and their wages*, edited by Marvin H. Kosters, 146-86. Washington, DC: The AEI Press, 1991.

Bowles, Samuel, and Henry M. Levin. "The determinants of scholastic achievement--an appraisal of some recent evidence." *Journal of Human Resources* 3, no. 1 (Winter 1968): 3-24.

Burtless, Gary. "Introduction and summary." In *Does money matter? The effect of school resources on student achievement and adult success*, edited by Gary Burtless, 1-42. Washington, DC: Brookings, 1996.

Cain, Glen G., and Harold W. Watts. "Problems in making policy inferences from the Coleman Report." *American Sociological Review* 35, no. 2 (April 1970): 328-52.

Card, David, and Alan B. Krueger. "Does school quality matter? Returns to education and the characteristics of public schools in the United States." *Journal of Political Economy* 100, no. 1 (February 1992a): 1-40.

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

---. "Labor market effects of school quality: Theory and evidence." In *Does money matter? The effect of school resources on student achievement and adult success*, edited by Gary Burtless, 97-140. Washington, DC: Brookings, 1996.

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

Glass, Gene V., and Mary Lee Smith. "Meta-analysis of research on class size and achievement." *Educational Evaluation and Policy Analysis* 1, no. 1 (1979): 2-16.

Greenberg, David, and John McCall. "Teacher mobility and allocation." *Journal of Human Resources* 9, no. 4 (Fall 1974): 480-502.

Grogger, Jeffrey T. "Does school quality explain the recent black/white wage trend?" *Journal of Labor Economics* 14, no. 2 (April 1996): 231-53.

Grogger, Jeffrey T., and Eric Eide. "Changes in college skills and the rise in the college wage premium." *Journal of Human Resources* 30, no. 2 (Spring 1993): 280-310.

Hanushek, Eric A. "Teacher characteristics and gains in student achievement: Estimation using micro data." *American Economic Review* 60, no. 2 (May 1971): 280-88.

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

---. "Throwing money at schools." *Journal of Policy Analysis and Management* 1, no. 1 (Fall 1981): 19-41.

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

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

---. "The trade-off between child quantity and quality." *Journal of Political Economy* 100, no. 1 (February 1992): 84-117.

---. "Interpreting recent research on schooling in developing countries." *World Bank Research Observer* 10, no. 2 (August 1995): 227-46.

---. "A more complete picture of school resource policies." *Review of Educational Research* (Fall 1996a).

---. "School resources and student performance." In *Does money matter? The effect of school resources on student achievement and adult success*, edited by Gary Burtless, 43-73. Washington, DC: Brookings, 1996b.

Hanushek, Eric A., and Dongwook Kim. "Schooling, labor force quality, and economic growth". Working Paper 5399, National Bureau of Economic Research, December 1995.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. "Aggregation and the estimated effects of school resources." *Review of Economics and Statistics* 78, no. 4 (November 1996).

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

Heckman, James S., Anne Layne-Farrar, and Petra Todd. "Does measured school quality really matter? An examination of the earnings-quality relationship." In *Does money matter? The effect of school resources on student achievement and adult success*, edited by Gary Burtless, 192-289. Washington, DC: Brookings, 1996a.

Heckman, James, Anne Layne-Farrar, and Petra Todd. "Human capital pricing equations with an application to estimating the effect of schooling quality on earnings." *Review of Economics and Statistics* (November 1996b).

Hedges, Larry V. "Directions for future methodology." In *The future of meta-analysis*, edited by Kenneth W. Wachter and Miron L. Straf, 11-26. New York: Russell Sage, 1990.

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

---. "The effect of school resources on student achievement." *Review of Educational Research* (Fall 1996).

Hedges, Larry V., and Ingram Olkin. *Statistical methods for meta-analysis*. San Diego, CA: Academic Press, 1985.

Hoxby, Caroline Minter. "How teachers' unions affect education production." *Quarterly Journal of Economics* 111, no. 3 (August 1996): 671-718.

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

Mosteller, Frederick. "The Tennessee study of class size in the early school grades." *The Future of Children* 5, no. 2 (Summer/Fall 1995): 113-27.

Murnane, Richard J. *Impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger, 1975.

---. "Teacher mobility revisited." *Journal of Human Resources* 16, no. 1 (Winter 1981): 3-19.

Murnane, Richard J., and Barbara Phillips. "What do effective teachers of inner-city children have in common?" *Social Science Research* 10, no. 1 (March 1981): 83-100.

Murnane, Richard J., John B. Willett, and Frank Levy. "The growing importance of cognitive skills in wage determination." *Review of Economics and Statistics* 77, no. 2 (May 1995): 251-66.

Neal, Derek A., and William R. Johnson. "The role of pre-market factors in black-white differences." *Journal of Political Economy* 104, no. 5 (October 1996): 869-95.

O'Neill, June. "The role of human capital in earnings differences between black and white men." *Journal of Economic Perspectives* 4, no. 4 (Fall 1990): 25-46.

Smith, Marshall S., Brett W. Scoll, and Jeffrey Link. "Research-based school reform: The Clinton administraion's agenda." In *Improving the performance of America's schools*, edited by Technology Board on Science and Economic Policy., 9-27. Washington, DC: National Academy Press, 1995.

Wachtel, Paul. "The effect on earnings of school and college investment expenditures." *Review of Economics and Statistics* 58, no. 3 (August 1976): 326-31.

Welch, Finis. "Measurement of the quality of schooling." *American Economic Review* 56, no. 2 (May 1966): 379-92.

Word, Elizabeth, John Johnston, Helen Pate Bain, B. DeWayne Fulton, Jayne Boyd Zaharies, Martha Nannette Lintz, Charles M. Achilles, John Folger, and Carolyn Breda. *Student/teacher achievement ratio (STAR), Tennessee's K-3 class size study: Final summary report, 1985-1990.* Nashville, TN: Tennessee State Department of Education, 1990.

Sources of Tabulated Results

Akin, John S., and Irwin Garfinkel. "School expenditures and the economic returns to schooling." *Journal of Human Resources* 12, no. 4 (Fall 1977): 460-81.

Anderson, Gary M., William F. Shughart II, and Robert D. Tollison. "Educational achievement and the cost of bureaucracy." *Journal of Economic Behavior and Organization* 15 (1991): 29-45.

Andrews, Donald R., Richaku Fayissa, and Uday S. Tate. "An estimation of the aggregate educational production function for public schools in Louisiana." *Review of Black Political Economy* (Summer 1991): 25-47.

Armor, David, Patricia Conry-Oseguera, Millicent Cox, Niceima King, Lorraine McDonnell, Anthony Pascal, Edward Pauly, and Gail Zellman. *Analysis of the school preferred reading program in selected Los Angeles minority schools*. Santa Monica, CA: Rand Corp., 1976.

Baum, Donald N. "A simultaneous equations model of the demand for and production of local public services: The case of education." *Public Finance Quarterly* 14, no. 2 (April 1986): 157-78.

Behrendt, Amy, Jeffrey A. Eisenach, and William R. Johnson. "Selectivity bias and the determinants of SAT scores." *Economics of Education Review* 5, no. 4 (1986): 363-71.

Beiker, Richard F., and Kurt R. Anschel. "Estimating educational production functions for rural high schools: some findings." *American Journal of Agricultural Economics* 55 (August 1973): 515-19.

Berger, Mark C., and Eugenia F. Toma. "Variation in state education policies and effects on student performance." *Journal of Policy Analysis and Management* 13, no. 3 (Summer 1994): 477-91.

Bingham, Richard D., John S. Heywood, and Sammis B. White. "Evaluating schools and teachers based on student performance." *Evaluation Review* 15, no. 2 (April 1991): 191-218.

Boardman, Anthony, Otto Davis, and Peggy Sanday. "A simultaneous equations model of the educational process." *Journal of Public Economics* 7, no. 1 (April 1977): 23-49.

Borland, Melvin V., and Roy M. Howsen. "Student academic achievement and the degree of market concentration in education." *Economics of Education Review* 11, no. 1 (March 1992): 31-39.

Bosshardt, William, and Michael Watts. "Instructor effects and their determinants in precollege economic education." *Journal of Economic Education* 21, no. 3 (Summer 1990): 265-76.

---. "Instructor effects in economics in elementary and junior high schools." *Journal of Economic Education* 24, no. 3 (Summer 1994): 195-211.

Bowles, Samuel. "Towards an educational production function." In *Education, income and human capital*, edited by W. Lee Hansen, 11-60. New York: National Bureau of Economic Research, 1970.

Brown, Byron W. "Achievement, costs, and the demand for public education." *Western Economic Journal* 10, no. 2 (June 1972): 198-219.

Brown, Byron W., and Daniel H. Saks. "The production and distribution of cognitive skills within schools." *Journal of Political Economy* 83, no. 3 (June 1975): 571-93.

Burkhead, Jesse. *Input-output in large city hIgh schools.* Syracuse, NY: Syracuse University Press, 1967.

Butler, Michael R., and Edward M. McNertney. "Estimating educationalproduction functions: The problem of multicollinearity." *Social Science Journal* 28, no. 4 (1991): 489-99.

Caldas, Stephen J. "Reexamination of input and process factor effects on public school achievement." *Journal of Educational Research* 86, no. 4 (April 1993): 206214.

Card, David, and Alan B. Krueger. "Does school quality matter? Returns to education and the characteristics of public schools in the United States." *Journal of Political Economy* 100, no. 1 (February 1992): 1-40.

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

Cohn, Elchanan. "Economies of scale in Iowa high school operations." *Journal of Human Resources* 3, no. 4 (Fall 1968): 422-34.

Cohn, Elchanan, and Stephen D. Millman. *Input-output analysis in public education.* Cambridge, MA: Ballinger, 1975.

Deller, Steven C., and Edward Rudnicki. "Production efficiency in elementary education: The case of Maine public schools." *Economics of Education Review* 12, no. 1 (1993): 45-57.

Dolan, R. C., and R. M. Schmidt. "Assessing the impact of expenditue on achievement: Some methodological and policy considerations." *Economics of Education Review* 6, no. 3 (1987): 285-99.

Dugan, Dennis J. "Scholastic achievement: its determinants and effects in the education industry." In *Education as an industry*, edited by Joseph T. Froomkin, Dean T. Jamison, and Roy Radner, 53-83. Cambridge, MA: Ballinger, 1976.

Dynarski, Mark. "The scholastic aptitude test: Participation and performance." *Economics of Education Review* 6, no. 3 (1987): 263-74.

Dynarski, Mark, Robert Schwab, and Ernest Zampelli. "Local characteristics and public production:The case of education." *Journal of Urban Economics* 26 (1989): 250-63.

Eberts, Randall W., and Joe Stone. *Unions and public schools: the effect of collective bargaining on American education.* Lexington, MA: Lexington Books, 1984.

Edwards, Linda Nasif. "The economics of schooling decisions: Teenage enrollment rates." *Journal of Human Resources* 10, no. 2 (Spring 1975): 155-73.

Ehrenberg, Ronald G., and Dominic J. Brewer. "Do school and teacher characteristics matter? Evidence from 'High School and Beyond'." *Economics of Education Review* 13, no. 1 (1994): 1-17.

Ehrenberg, Ronald G., Randy A. Ehrenberg, Daniel I. Rees, and Eric L. Ehrenberg. "School district leave policies, teacher absenteeism, and student achievement." *Journal of Human Resources* 26, no. 1 (Winter 1991): 72-105.

Farkas, George, Robert P. Grobe, Daniel Sheehan, and Yuan Shuan. "Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban school district." *American Sociological Review* 55, no. 1 (February 1990): 127-42.

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

Fowler, William J. Jr, and Herbert J. Walberg. "School size,characteristics, and outcomes." *Educational Evaluation and Policy Analysis* 13, no. 2 (Summer 1991): 189-202.

Grimes, Paul W. "Public versus private secondary schools and the production of economic education." *Journal of Economic Education* (Winter 1994): 17-30.

Grimes, Paul W., and Charles A. Register. "Teachers' unions and student achievement in high school economics." *Journal of Economic Education* (Summer 1990): 297-306.

---. "Teacher unions and black students' scores on college entrace exams." *Industrial Relations* 30, no. 3 (Fall 1991): 492-500.

Hanushek, Eric A. "Teacher characteristics and gains in student achievement: Estimation using micro data." *American Economic Review* 60, no. 2 (May 1971): 280-88.

---. *Education and race: an analysis of the educational production process*. Cambridge, MA: Health-Lexington, 1972.

---. "The trade-off between child quantity and quality." *Journal of Political Economy* 100, no. 1 (February 1992): 84-117.

---. "Can equity be separated from efficiency in school finance debates?" In *Essays on the economics of education*, edited by Emily P. Hoffman, 35-73. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1993.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. "Aggregation and the estimated effects of school resources." *Review of Economics and Statistics* 78, no. 4 (November 1996).

Harnisch, Delwyn L. "Characteristics associated with effective public high schools." *Journal of Educational Research* 80, no. 4 (April 1987): 233-41.

Heim, John, and Lewis Perl. *The educational production function: implications for educational manpower policy*. Ithaca, NY: Cornell University, 1974.

Henderson, Vernon, Peter Mieszkowski, and Yvon Sauvageau. *Peer group effects and educational production functions*. Ottawa, Canada: Economic Council of Canada, 1976.

Jencks, Christopher S., and Marsha Brown. "Effects of high schools on their students." *Harvard Educational Review* 45, no. 3 (August 1975): 273-324.

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

Jud, G. Donald, and James L. Walker. "Discrimination by race and class and the impact of school quality." *Social Science Quarterly* 57 (1977): 731-49.

Katzman, Martin. *Political economy of urban schools*. Cambridge, MA: Harvard University Press, 1971.

Keisling, Herbert. "Measuring a local government service: a study of school districts in New York state." *Review of Economics and Statistics* 49 (August 1967): 356-67.

Kenny, Lawrence W. "Economies of scale in schooling." *Economics of Education Review* 2, no. 1 (1982): 1-24.

Kiesling, Herbert J. "Assignment practices and the relationship of instructional time to the reading performance of elementary school children." *Economics of Education Review* 3, no. 4 (1984): 341-50.

Kurth, Michael M. "Teachers' unions and excellence in education: An analysis of the decline in SAT scores." *Journal of Labor Research* 8, no. 4 (Fall 1987): 351-67.

Levin, Henry M. "A new model of school effectiveness." In *Do teachers make a difference?*, edited by U.S. Office of Education. Washington, DC: U.S. Government Printing Office, 1970.

---. "Concepts of economic efficiency and educational production." In *Education as an industry*, edited by Joseph T. Froomkin, Dean T. Jamison, and Roy Radner, 149-90. Cambridge, MA: Ballinger, 1976.

Link, Charles R., and James G. Mulligan. "The merits of a longer school day." *Economics of Education Review* 5, no. 4 (1986): 373-81.

---. "Classmates' effects on black student achievement in public school classrooms." *Economics of Education Review* 10, no. 4 (1991): 297-310.

Link, Charles R., and Edward C. Ratledge. "The influence of the quantity and quality of education on black-white earnings differentials: Some new evidence." *Review of Economics and Statistics* 57, no. 3 (August 1975): 346-50.

---. "Student perceptions, I.Q., and achievement." *Journal of Human Resources* 14, no. 1 (Winter 1979): 98-111.

Lopus, Jane S. "Do additional expenditures increase achievement in the high school economics class?" *Journal of Economic Education* 21, no. 3 (Summer 1990): 277-86.

Maynard, Rebecca, and David Crawford. "School Performance." In *Rural income maintenance experiment: final report*. Madison, WI: University of Wisconsin, 1976.

Michelson, Stephen. "The association of teacher resources with children's characterestics." In *Do teachers make a difference?*, edited by U.S. Office of Education, 120-68. Washington, DC: U.S. Government Printing Office, 1970.

---. "For the plaintiffs--equal school resource allocation." *Journal of Human Resources* 7, no. 3 (Summer 1972): 283-306.

Montmarquette, Claude, and Sophie Mahseredjian. "Does school matter for educational achievement? A two-way nested-error components analysis." *Journal of Applied Econometrics* 4 (1989): 181-93.

Morgenstern, Richard D. "Direct and indirect effects on earnings of schooling and socio-economic background." *Review of Economics and Statistics* 55, no. 2 (May 1973): 225-33.

Murnane, Richard J. *Impact of school resources on the learning of inner city children.* Cambridge, MA: Ballinger, 1975.

Murnane, Richard J., and Barbara Phillips. "What do effective teachers of inner-city children have in common?" *Social Science Research* 10, no. 1 (March 1981): 83-100.

Namboodiri, Krishnan, Ronald G. Corwin, and Linda Eberst Dorsten. "Analyzing distributions in school effects research: An empirical illustration." *Sociology of Education* 66 (October 1966): 278-94.

Pallas, Aaron M., Doris R. Entwisle, Karl L. Alexander, and Doris Cadigan. "Children who do exeptionally well in first grade." *Sociology of Education* 60 (October 1987): 257-71.

Perl, Lewis J. "Family background, secondary school expenditure, and student ability." *Journal of Human Resources* 8, no. 2 (Spring 1973): 156-80.

Raymond, Richard. "Determinants of the quality of primary and secondary public education in West Virginia." *Journal of Human Resources* 3, no. 4 (Fall 1968): 450-70.

Ribich, Thomas I., and James L. Murphy. "The economic returns to increased educational spending." *Journal of Human Resources* 10, no. 1 (Winter 1975): 56-77.

Sander, William. "Expenditure and student achievement in Illinois." *Journal of Public Economics* 52 (1993): 403-16.

Sander, William, and Anthony C. Krautmann. "Local taxes,schooling,and jobs in Illinois." *Economics of Education Review* 10, no. 2 (1991): 111-21.

Sebold, Frederick D., and William Dato. "School funding and student achievement: an empirical analysis." *Public Finance Quarterly* 9, no. 1 (January 1981): 91-105.

Sengupta, Jati K., and Raymond E. Sfeir. "Production frontier estimates of scale in public schools in California." *Economics of Education Review* 5, no. 3 (1986): 297-307.

Smith, Marshall. "Equality of educational opportunity: the basic findings reconsidered." In *On equality of educational opportunity*, edited by Frederick Mosteller and Daniel P. Moynihan, 230-342. New York: Random House, 1972.

Stern, David. "Educational cost factors and student achievement in grades 3 and 6: Some new evidence." *Economics of Education Review* 8, no. 2 (1989): 149-58.

Strauss, Robert P., and Elizabeth A. Sawyer. "Some new evidence on teacher and student competencies." *Economics of Education Review* 5, no. 1 (1986): 41-48.

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

Tuckman, Howard P. "High school inputs and their contribution to school performance." *Journal of Human Resources* 6, no. 4 (Fall 1971): 490-509.

Wachtel, Paul. "The effect of school quality on achievement, attainment levels, and lifetime income." *Explorations in Economic Research* 2 (1975): 502-36. Occasional papers of the National Bureau of Economic Research.

---. "The effect on earnings of school and college investment expenditures." *Review of Economics and Statistics* 58, no. 3 (August 1976): 326-31.

Walberg, Herbert J., and William J. Fowler Jr. "Expenditure and size efficiencies of public school districts." *Educational Researcher* (October 1987): 5-13.

Walberg, Herbert J., and Sue Pinzur Rasher. "Public school effectiveness and equality: New evidence and its implications." *Phi Beta Kappan* (September 1974): 3-9.

Walberg, Herbert J., and Herbert J. Walberg III. "Losing local control." *Educational Researcher* 22, no. 5 (June-July 1994): 19-26.

Watts, Michael. "School district inputs and biased estimation of educational production functions." *Journal of Economic Education* (Fall 1985): 281-86.

Wendling, Wayne, and Judith Cohen. "Education resources and student achievement: Good news for schools." *Journal of Education Finance* 7 (Summer 1981): 44-63.

Winkler, Donald. "Educational achievement and school peer group composition." *Journal of Human Resources* 10, no. 2 (Spring 1975): 189-204.

Appendix
Sampling of Studies Employed by Hedges, Laine, and Greenwald (1996)

The conclusions of the statistical testing of Hedges *et al.* (1996) have received considerable attention, in part because they appear to follow careful statistical procedures. Unfortunately, their testing is dependent upon choosing a very selective sample of the available analytical results (from Table 3). The importance of sample selection is readily understood within the context of available data.

Table A1 shows the sampling percentages, reflecting the proportion of available studies (by results) that are used by Hedges *et al.* (1996). First, for purely technical reasons their methodology requires that they eliminate all studies finding statistically insignificant effects but not reporting the sign (see the last column of Table A1). This action by itself eliminates 13 to 26 percent of the available data. The preliminary elimination of substantial evidence against significant resource effects biases the results towards finding statistically significant results. Second, the sample selection process uniformly retains a higher proportion of the statistically significant positive results than of the overall results. In the cases of teacher education and of expenditure per pupil, the sampling rate for statistically significant positive results is double the overall sampling rate. While they retain just 22 percent of the available estimates of the effects of teacher education, they retain 44 percent of those that show a positive and statistically significant effect. Similarly, for expenditure per pupil, they retain only 17 percent of all studies but 34 percent of those with positive and statistically significant estimated effects. At the same time, with the exception of the teacher education results, Hedges *et al.* (1996) retains a lower proportion of statistically significant negative results than of the overall results. Moreover, among the insignificant results, the sampling tends to retain a relatively higher proportion of the positive estimates than of the negative estimates (with the minor exception of essentially equal sampling rates for expenditure per pupil). The overall sampling in Hedges *et al.* (1996) is dramatically biased toward retaining both statistically significant positive and insignificant but positive results, just the direction that leads to supporting their general conclusions.

Table A1. Sampling Rates for Studies Employed by Hedges, Laine, and Greenwald (1996), Total and by Results[a]
(Percentages)

Outcome measure	number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
Teacher-pupil ratio	23%	31%	19%	43%	17%	0%
Teacher education	22	44	67	27	22	0
Teacher experience	30	30	20	40	30	0
Expenditure per pupil	17	34	9	9	10	0

Source: Author's tabulations.

Note: a. The number of studies by results employed in the statistical analyses of Hedges *et al.* (1996) are compared to the total number of studies available, as found in Table 3.