Educational Evaluation and Policy Analysis
Summer 1997, Vol. 19, No. 2, pp. 141–164

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

Eric A. Hanushek
University of Rochester

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 article 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.

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 has helped clarify both the interpretation of this work and the resulting policy implications. This article 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 that has resource policies at its heart. Added resources within the current organization and incentives of schools are neither necessary nor sufficient for improving student achievement. Instead, incentive structures that encourage better performance and recognize differences of students, teachers, and schools offer much greater likelihood of success than the centralized decision-making approaches currently prevalent.

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 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 30 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 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 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.[5]

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% measure school inputs only at the level of the state. Moreover, fully one quarter of the studies employing non-test measures rely solely on interstate variations in school inputs.

The Impact of School Resources

The overall approach here is to summarize the combined evidence about the effects of various school resources. As will be apparent, given the

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 level	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

Note: Source—Author's tabulations.

large number of studies it is quite possible to find individual studies supporting one or another positions—such as supporting the efficacy of providing some specific programs or resources. Because there are widely divergent results from individual studies, this analysis concentrates on systematic effects that hold across the available studies.

Studies of educational performance include a variety of different measures of resources devoted to schools. Commonly employed measures include

- The real resources of the classroom (teacher education, teacher experience, and teacher-pupil ratios);
- Financial aggregates of resources (expenditure per student and teacher salary); and
- 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 percentage of teachers with master's degrees 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 measures stand in contrast with the other 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 of noticeably lower quality than the best—and the typical—study investigating real classroom resources.[6] The measures of other school resources also are frequently measured poorly and tend to be available only at the district level. At the same time, because 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% of the studies considering the level of teachers' education and 15% of the studies investigating teacher-pupil ratios find positive and statistically significant effects on student performance.[7] These relatively small numbers of statistically significant positive results are balanced by another set finding statistically significant negative results—reaching 13% 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 re-

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 degrees	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

Note: Source—U.S. Department of Education, 1994.

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

Note: Source—Author's tabulations.

porting 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%. Importantly, however, 71% 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 effects could actually reflect reverse causation (Greenberg & McCall, 1974; Murnane, 1981). In sum, the vast number of estimated real resource effects gives little confidence that just adding more of any of the specific resources 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% of the estimated coefficients are positive and statistically significant. In fact, from the statistically significant negative estimates, we see that 7% even suggest some confidence 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 the reported 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.[8] An additional exception is teacher test score, where teachers have been given some sort of achievement or IQ test and their score on those has been related to their students' performance. 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% 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. Problems arise when there are either nonlinearities, such as

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

		Statistically significant		Statistically insignificant		
Resources	Number of estimates	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

Note: Source—Author's tabulations.

interactions of school and family factors, or specification problems, such as omitted variables. Even with these problems, however, there is no real expectation about the direction of any effect on estimates that might accompany aggregation of school resource variables.[9] 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 for teacher-pupil ratio and expenditures. (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 appear to have a stronger positive influence and to be more frequently statistically significant as the level of aggregation increases from the school to the district to the state. For example, for teacher-pupil ratios, the percentage of positive and statistically significant estimates goes from 12% to 21% and 64% as the estimates go from the classroom level to aggregation at the district and state level, respectively. 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.[10] Individual states,

TABLE 5
Percentage Distribution of Estimated Effect of Teacher-Pupil Ratio and Expenditure Per Pupil on Student Performance

		Statistically significant		Statistically insignificant		
Resources	Number of estimates	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.

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 methods 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 policies influence school performance. Unfortunately, little progress has been made at identifying, defining, or measuring the most important aspects of state policies in terms of their effect on student performance or the efficiency of resource usage. Whether well measured or not, such state factors can have a significant impact on the results of common statistical analyses, such as those summarized here. For example, 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 specification 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 from these specification biases, but studies drawing observations across states will.[11] Additionally, the effect of biases is not independent of the modeling strategy. Hanushek, Rivkin, and Taylor (1996) show that as data are aggregated to the 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 suggests that resources are more important for student performance than those analyzing achievement within individual states. Looking, for example, at the teacher-pupil results, there are consistently more positive and statistically significant estimates from the multiple-state samples (18%) compared to single state samples (12%). There are also noticeably fewer negative and statistically significant estimates (8% for multiple-state samples versus 18% for single-state samples). Similar results hold for expenditure per pupil. Moreover, the apparent importance of resources increases with aggregation, just what the theory suggests in the case of misspecification at the state level. At the state level of estimation, almost two thirds of the estimates for both teacher-pupil ratios and expenduture per pupil are positive and statistically significant. The fact 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,

TABLE 6

Percentage Distribution of Estimated Effect of Teacher-Pupil Ratio and Expenditure Per Pupil by State Sampling Scheme and Aggregation

		Statistically significant		Statistically insignificant		
Level of aggregation resources	Number of estimates	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

Note: 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.

in selecting studies for tabulation, an effort was made to collect the entire universe of studies that 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.[12]

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 some 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).[13]

Table 7 provides a summary of value-added results, both for all 96 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 because of a 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 4% (i.e., 1 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 database is 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

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

Note: Source—Author's tabulations.

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.

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 the contrary. Substantial evidence suggests that there are large differences among teachers and schools.[14] The simple fact remains that these differences are not closely related to teacher salaries or to other measured 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. It ignores the 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 that indicate ineffective use. The interpretation is 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 5% of the time (for a 95% significance level), with 2.5% of the studies finding a positive and statistically significant effect and 2.5% 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 they will 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.[15]

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 on the aggregation of the resource measures and on whether samples are drawn across states. Again, the underlying biases would push the results toward finding more statistically significant estimates than would be the case when there are not systematic resource effects.

Neither explanation for the observed distributions of resource effects provides more support for the importance of resources. Both point to the conclusion that the weak results previously displayed are actually overstating the strength of any resource relationships.

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 performance. However, the widely publicized findings of Card and Krueger (1992a) indicate that variations in school resources are related to earnings differences among workers.[16] 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.[17] 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., Borland & Howsen, 1992; Hoxby, 1996; Peltzman, 1993). 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.

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.[18] 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 that use test scores as a measure of outcomes and those that use other measures of outcomes like college continuation or earnings. This can be seen in Table 8, which 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 impact on earnings)? Answering this question is, perhaps, more difficult than answering 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 have no significant impact on student behavior once individual achievement and school costs are considered.[19] Betts (1996) reviews a number of these studies of educational attainment and suggests some positive effects of resources. For the studies tabulated here (which

TABLE 8

Percentage Distribution of Estimated Effect of Expenditure Per Pupil on Student Performance by Outcome Measure and Aggregation of Resource Effects (163 estimates)

		Statistically significant		Statistically insignificant		
Outcome measure	Number of estimates	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.

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 5 estimated from within individual states.[20] 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, Bishop, 1991; Grogger & Eide, 1993; Murnane, Willett, & Levy, 1995; Neal & Johnson, 1996; O'Neill, 1990). 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 explain the different conclusions involves 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 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 will 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, Rivkin, & Taylor, 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.[21]

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 affect 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, such as in 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 that 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.[22]

A well-known version of applying formal statistical tests to education production-function data is found in Hedges, Laine, and Greenwald (1994) and Greenwald, Hedges, and Laine (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 to define a common parameter in a way that is susceptible to formal testing. More important, 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. To deal with the lack of independence of results, Hedges et al. preselect a very specific sample of available evidence. This procedure—forced by their methodology—not only throws away considerable information about resource effects but also leads to badly biased samples. As described in the appendix, their sample, by itself, would be sufficient to produce their conclusions.

The most basic problem with this statistical analysis, however, is that it addresses a completely uninteresting question—one that has little relevance from a policy viewpoint. Hedges et al. 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* 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., $H_0 : \beta_1 = \beta_2 = \ldots = \beta_n = 0$, 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 a systematic effect on student performance). The 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 the resource was someplace related to performance.[23]

The obvious interpretation of the previously reported results, as discussed above, is that there is a distribution of underlying parameters that 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.[24]

Their formal tests lead to rejection of this restricted null hypothesis.[25] 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, attention is 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.[26]

STAR Experiment

In the mid-1980s, because of ambiguity about the effects of class size on student performance, the state of Tennessee launched a random-assignment experiment in reducing class size (Word et al., 1990). 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 appeared to yield gains in student performance. Based on this, a group of kindergarten through third-graders in Tennessee were randomly assigned to either large classes (22–24 students) or small classes (14–16 students).[27] Students were followed over time as they progressed 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.[28] 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. 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, because 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.

An 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.

Yet a third alternative would be that the lowered class size did not affect learning but instead influenced the socialization of students into schools and learning settings. Such an effect would be consistent with a one-time shift in the level of student performance. It would also suggest that any resources devoted to lowering class sizes should be concentrated just on the earliest grades.

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 either the general cumulative model or the socialization 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 in this article 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 devoted to class-size issues, that there has been no serious follow-up of the STAR experiment with similar experiments. As discussed in Hanushek with others (1994), improved experiments can potentially save considerable money by pinpointing when and where resources might productively be applied instead of moving directly to major public funding of full-scale programs.[29]

Policy Implications

The interpretation of these results depends fundamentally on how the policy- and decision-making process is conceived. At one level, these conclusions clearly imply that educational policy-making is 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.

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 state legislatures are not fulfilling their constitutional obligations to provide equitable or adequate education to identified students in each state. While frequently motivated by concerns about student achievement, in reality both the judicial statement of the issue and the proposed remedies revolve around the level and distribution of resources. If resource availability is not a good index of educational outcomes or if providing for overall resource levels does not ensure a desired level of performance, the courts face the same dilemma as legislatures. Simply providing more funding or a different distribution of funding is unlikely to improve student achievement (even though it may affect the tax burdens of school financing across the citizens of a state).

A variation of this general theme is to argue that, while resources alone may not be sufficient to guarantee achievement, adequate resources are surely necessary. Undoubtedly, this is an ac-

TABLE 9
Percentage Distribution of Estimated Influence of Teacher-Pupil on Student Performance, by Level of Schooling

School level	Number of estimates	Statistically significant: Positive	Statistically significant: Negative	Statistically insignificant: Positive	Statistically insignificant: Negative	Statistically insignificant: 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

Note: Source—Author's tabulations.

curate statement at some level because a school with no funds would not be expected to add anything to student achievement. Nonetheless, as shown in Table 2, real spending per student increased by more than 70% between 1970 and 1990, even though student performance appears to have remained essentially unchanged.[30] Further, nothing in the previous analytical results about the effects of resources suggests that there is a level below which resources have clear and powerful effects on achievement—which would be a demonstration that some schools are below the threshold of "necessity." Just asserting that there is some level of necessary expenditure does not make the case for pure resource policies in today's schooling environment. While it is not possible to define scientifically how much is "necessary," it seems clear that the dramatically larger spending of today has taken virtually every school system in the country beyond some minimal level.

A related issue—one highlighted in some recent school-finance court cases—centers on whether funding for schools is "adequate." Such concepts may have popular appeal, but they have no policy superiority to traditional district equity arguments when translated into resource requirements. First, what is adequate is a purely political and economic issue that it likely to change both with the demands of the economy and with political views on appropriate levels of government support of programs. Second, and more important, the previous analyses of the lack of a relationship between resources and student performance hold no matter what goals are placed for student achievement or how they are arrived at. Thus, there is no objective method of indicating what resources are required for an "adequate" level of student performance.

If the object of policy is student achievement, simply changing the resources available to schools while retaining the existing decision-making in schools is unlikely to have the desired effects. Its main impact will be to increase the costs of schools.

The considerations of overall spending levels, either in legislatures or the courts, largely rest on the premise that local districts are best situated and motivated to use funds wisely and productively. There is ample evidence, however, that policymakers do not fully believe that. The extensive bodies of rules and regulations at the federal and state levels are mainly designed to ensure that local districts do not do undesirable things in operating their schools and indicate a considerable distrust of the motivations and/or abilities of local districts. To set regulations appropriately, one would need to know how resources or process considerations affect student performance—which we do not know in any way sufficient for designing most regulatory approaches to good schooling. An extension of this that pervades much of the thinking and decision-making about schools is the view that educational approaches can be effectively set centrally. This is consistent with a widely held view that "what works" is known. For example, Smith, Scoll, and Link (1996) unequivocally assert 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" (p. 23).) 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 because otherwise they would use available resources more effectively.[31] It also suggests 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.

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 alter current incentives in schools radically (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. This then makes the choice of programs, organization, and behaviors less dependent on student outcomes than on other things that more directly affect the actors in schools.

Underlying the incentive perspective is also 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 instead 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 that these policies are not easily described and centrally regulated. The assumption is that, given better incentives, school personnel can be motivated to search out what will work in their specific situations. Under current incentives, they appear to devote more of their attention and energies elsewhere.

Take the specific example of policies to reduce teacher-pupil ratios and class sizes. Many people find it difficult to believe that lowering class sizes will not lead to improved student performance, because teachers could devote more attention to the needs of each individual student if there were fewer students. While the overall evidence provided earlier pointed to no clear relationship between teacher-pupil ratios and student performance, my own interpretation is that there are almost certainly some teachers, some specific classes, and some groups of students for whom smaller classes can lead to real performance gains but that these circumstances are balanced by others where there are no obvious advantages to smaller classes. Without performance incentives, the question of class size policy is often viewed from the vantage point of fairness, which is frequently interpreted as calling for lowering all class sizes uniformly. In other circumstances, the teacher-pupil ratio may rise without actually affecting class sizes—through the addition of special programs or simply from negotiations to lower teacher contact time in the classroom. Such circumstances offer plausible explanations for the lack of effect on student performance of overall differences in class size or teacher-pupil ratios because well-considered reductions in class size are generally mixed in with overall, across-the-board reductions. On the other hand, if there were direct incentives to consider improving student performance, there could well be more surgical use of reduced class size—balanced perhaps with some increases in class size so that student performance could be increased for a given spending on programs. Indeed, it is conceivable that some of the best teachers were put into larger classes, where they could influence more students. These kinds of decisions seldom occur today, given the lack of direct rewards and incentives and the perspective of making overall, centralized decisions. Instead, objectives and goals other than enhanced student achievement are more readily considered and pursued.

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). The alternative incentive structures include a variety of conceptual approaches to providing rewards for improved student performance and range from merit pay for teachers to charter schools to privatization to vouchers. These are contentious proposals, in part because introduction of performance incentives might lead to having a variety of people other than current school personnel making decisions and even providing educational services. They are also proposals that could work well or poorly, depending on the details. The purpose here, however, is not to consider the pros and cons of alternatives, but to emphasize the radically different perspective on policy that is embedded in each. Performance incentives recognize that there might be varying approaches by teachers and schools that are productive. Thus, they avoid the centralized "command and control" perspective of much current policy. At the same time, they recognize that simply decentralizing decision-making is unlikely to work effectively unless there exist clear objectives and unless there is direct accountability.[32]

Given the current lack of knowledge about the design or implications of performance incentives, an aggressive program of experimentation and evaluation seems very appropriate (Hanushek with others, 1994). Nonetheless, the lack of direct information about alternatives should not be taken as support for more of what we are doing now. We actually have considerable experience with the current organization, and current approaches appear to offer little hope for general improvement.

The existing work does not suggest that 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 policymakers. We know that they have not been accurate in their past identification.

APPENDIX

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

The conclusions of the statistical testing of Greenwald et al. (1996) have received considerable attention, in part because they appear to follow careful statistical procedures. Unfortunately, their testing is dependent on choosing a 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 10 shows the selection percentages, reflecting the proportion of available studies (by results) that are used by Greenwald 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 10). This action by itself eliminates 13% to 26% of the available data. The preliminary elimination of substantial evidence against significant resource effects biases the results toward finding statistically significant results. Second, additional loss is caused by the fact that their methodology cannot deal with any dependencies among the estimates, such as those caused by analyzing different students who are enrolled in a common school system. Thus, they employ rather arbitrary rules for dropping results from correlated studies by given authors (although they ignore correlations from different authors who employ a common data set). Dropping studies, even if the samples are related and the estimates from them will be correlated, clearly leads to a loss of

TABLE 10
Selection Rates for Studies Employed by Greenwald, Hedges, and Laine (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

Note: Source—Author's tabulations.
[a] The number of studies by results employed in the statistical analyses of Greenwald et al. (1996) are compared to the total number of studies available, as found in Table 3.

information. Their specific 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 expenditure per pupil, the sampling rate for statistically significant positive results is double the overall sampling rate. While they retain just 22% of the available estimates of the effects of teacher education, they retain 44% of those that show a positive and statistically significant effect. Similarly, for expenditure per pupil, they retain only 17% of all studies but 34% of those with positive and statistically significant estimated effects. At the same time, with the exception of the teacher education results, Greenwald et al. (1996) retain a lower proportion of statistically significant negative results than of 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 Greenwald 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.

Notes

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.

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

[2]The tabulations do include results in Hanushek, Rivkin, and Taylor (1996) because 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 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.

[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.

[5]New analyses have also appeared, but they are not included because the systematic search of available journals and books went just through the end of 1994. Without systematically surveying all available sources, inclusion of some studies could lead to selection biases. Among these newer studies are Betts (1995), Ehrenberg and Brewer (1995), Ferguson and Ladd (1996), Grogger (1996), Lamdin (1995), and Staley and Blair (1995). It is also the case that, given the number of sampled studies, a few added results could not affect the overall conclusions here even if they all uniformly pointed in the same direction.

[6]Some studies include expenditure per pupil along with measures of the real classroom resources. In such a case, because 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.

[7]The individual studies tend to measure each of these inputs in different ways. For example, while many studies include an indicator variable for whether 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.

[8]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.

[9]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 sit-

uations of multivariate models. (See Hanushek, Rivkin, & Taylor, 1996, for a general discussion of aggregation and the potential biases.)

[10]The federal government has always had a rather limited role, directed largely at specific programs and populations. Its largest elementary and secondary programs involve funding for compensatory programs (such as Title 1), vocational education, and 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 statistical analyses of the educational production process.

[11]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.

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

[13]A related group of studies employs synthetic cohorts. These studies do not match current and past performance of the same students, but instead either add performance of current students in earlier grades or of students of the same vintage in prior grades (e.g., Ferguson & Ladd, 1996, Sebold & Dato, 1981). The first approach has none of the features that lead to preferring value-added studies because 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 socioeconomic and school factors because these are related to mobility behavior. Greenwald, Hedges, and Laine (1996) demonstrate that these synthetic cohort studies tend to find more significant expenditure effects. (See Hanushek, 1996a, for a discussion of these results.)

[14]The clearest evidence comes from a series of covariance, or fixed-effects, estimates of performance differences across teachers (e.g., Armor et al., 1976; Hanushek, 1971, 1992; Murnane, 1975; Murnane & Phillips, 1981). These analyses consistently show large and significant differences among teachers. 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 (Hanushek, 1992)). Moreover, the consistency of individual teacher effects across grades and school years indicates that the estimated differences relate directly to teacher quality and not the specific mix of students and the interaction of teacher and students.

[15]It is possible to ignore publication bias in the interpretation here because publication bias works against the "no systematic effect" conclusion. The same is not the case when one is working to establish that resource variations are important, as in Hedges, Laine, and Greenwald (1994) or Greenwald, Hedges, and Laine (1996). In their work, the inherent biases push the results toward their conclusions.

[16]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, Johnson and Stafford (1973), Wachtel (1976), and Welch (1966). An insightful review of past studies that considers underlying characteristics of the studies is Betts (1996).

[17]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 & Kim, 1996.)

[18]An important specification issue is that Card and Krueger (1992a) attempt 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, Layne–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 & Krueger, 1992b). The level of resources is lower and the differences in resources are higher in that study than in current situations, again suggesting that resources may matter at low levels.

[19]The major focus of that article 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.

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

[21]If, on the other hand, there are important measurement errors in the school resources, aggregation

could be beneficial because this would tend to average out any measurement problems. A central concern of Hanushek, Rivkin, and Taylor (1996) is distinguishing between the harmful effects of aggregation and model misspecification and the beneficial effects of aggregation and measurement error. That analysis rejects the hypothesis that measurement error is a primary element in the apparent importance of resources in the more aggregated studies.

[22]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 large samples frequently available.

[23]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."

[24]The actual application of the specific tests they employ requires a number of severe restrictions. One crucial aspect is the reliance on selective samples that are biased toward resource effects. They employ a series of arbitrary, but far from innocuous, selection rules in an attempt to make the data fit their methodology, which requires independence of the estimates. The sampling is discussed in the appendix.

[25]Note that the precise testing depends crucially on their specific choice of statistical methods and on their selective sampling of available results. (See the appendix to this article.)

[26]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.

[27]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 to participate 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. And some re-assignment of students occured after the first year of the experiment.

[28]A series of questions about the effects of initial randomization, of sample attrition, and of student mobility do remain. Unfortunately, the data from the STAR experiment have not been made generally available to researchers, so the analysis and interpretation of the results have had to rely on the published reports of the original researchers.

[29]In 1996, the state of California moved to a statewide program of providing significant additional funds to all schools that lowered class sizes in primary grades to state-prescribed levels. This program appears to have been the policy implementation of perceived results from the STAR experiment. Having done this on a statewide basis, the state has no effective way to evaluate the results of such an initiative, so that neither California nor other states can learn from this program. The existing evidence provides little reason to be optimistic about the future achievement effects of this policy.

[30]The overall performance of 17-year-olds on the National Assessment of Educational Progress (NAEP), while varying slightly by subject, indicates achievement in the mid-1990s is about the same as in the early 1970s (Hanushek with others, 1994). These trends could be complicated by nonschool factors, although these do not seem to be plausible explanations for the overall results (Hanushek 1996a, 1996b).

[31]One somewhat different reaction to the lack of consistent determinants of educational performance (as seen from the existing production function work) has been a concern that that research has been a failure because it does not clearly indicate what policies should be mandated. Again, this view accepts a level of comfort with centralized decision-making that has been discarded throughout most sectors of most economies in the world.

[32]While the decentralization considered here really refers to pure resource policies and general funding, the evidence supports this conclusion even at the level of school-based management (see Summers & Johnson, 1996).

Sources of Tabulated Results

Akin, J. S., & Garfinkel, I. (1977). School expenditures and the economic returns to schooling. *Journal of Human Resources, 12*(4), 460–481.

Anderson, G. M., Shughart, W. F., II & Tollison, R. D. (1991). Educational achievement and the cost of bureaucracy. *Journal of Economic Behavior and Organization, 15*, 29–45.

Andrews, D. R., Fayissa, R., & Tate, U. S. (1991, Summer). An estimation of the aggregate educational production function for public schools in Louisiana. *Review of Black Political Economy*, 25–47.

Armor, D., Conry–Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools.* Santa Monica, CA: Rand Corp.

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*(2), 157–178.

Behrendt, A., Eisenach, J. A., & Johnson, W. R. (1986). Selectivity bias and the determinants of SAT scores. *Economics of Education Review, 5*(4), 363–371.

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

Berger, M. C., & Toma, E. F. (1994). Variation in state education policies and effects on student performance. *Journal of Policy Analysis and Management, 13*(3), 477–491.

Bingham, R. D., Heywood, J. S., & White, S. B. (1991). Evaluating schools and teachers based on student performance. *Evaluation Review, 15*(2), 191–218.

Boardman, A., Davis, O., & Sanday, P. (1977). A simultaneous equations model of the educational process. *Journal of Public Economics, 7*(1), 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*(1), 31–39.

Bosshardt, W., & Watts, M. (1990). Instructor effects and their determinants in precollege economic education. *Journal of Economic Education, 21*(3), 265–276.

Bosshardt, W., & Watts, M. (1994). Instructor effects in economics in elementary and junior high schools. *Journal of Economic Education, 24*(3), 195–211.

Bowles, S. (1970). Towards an educational production function. In W. L. Hansen (Ed.), *Education, income and human capital* (pp. 11–60). New York: National Bureau of Economic Research.

Brown, B. W. (1992). Achievement, costs, and the demand for public education. *Western Economic Journal, 10*(2), 198–219.

Brown, B. W., & Saks, D. H. (1975). The production and distribution of cognitive skills within schools. *Journal of Political Economy, 83*(3), 571–593.

Burkhead, J. (1967). *Input-output in large city high schools.* Syracuse, NY: Syracuse University Press.

Butler, M. R., & McNertney, E. M. (1991). Estimating educational production functions: The problem of multicollinearity. *Social Science Journal, 28*(4), 489–499.

Caldas, S. J. (1993). Reexamination of input and process factor effects on public school achievement. *Journal of Educational Research, 86*(4), 206–214.

Card, D., & Krueger, A. B. (1992a). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy, 100*(1), 1–40.

Card, D., & Krueger, A. B. (1992b). School quality and black-white relative earnings: A direct assessment. *Quarterly Journal of Economics, 107*(1), 151–200.

Cohn, E. (1968). Economies of scale in Iowa high school operations. *Journal of Human Resources, 3*(4), 422–434.

Cohn, E., & Millman, S. D. (1975). *Input-output analysis in public education.* Cambridge, MA: Ballinger.

Deller, S. C., & Rudnicki, E. (1993). Production efficiency in elementary education: The case of Maine public schools. *Economics of Education Review, 12*(1), 45–57.

Dolan, R. C., & Schmidt, R. M. (1987). Assessing the impact of expenditure on achievement: Some methodological and policy considerations. *Economics of Education Review, 6*(3), 285–299.

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: Ballinger.

Dynarski, M. (1987). The scholastic aptitude test: Participation and performance. *Economics of Education Review, 6*(3), 263–274.

Dynarski, M., Schwab, R., & Zampelli, E. (1989). Local characteristics and public production: The case of education. *Journal of Urban Economics, 26*, 250–263.

Eberts, R. W., & Stone, J. (1984). *Unions and public schools: The effect of collective bargaining on American education.* Lexington, MA: Lexington Books.

Edwards, L. N. (1975). The economics of schooling decisions: Teenage enrollment rates. *Journal of Human Resources, 10*(2), 155–173.

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), 1–17.

Ehrenberg, R. G., Ehrenberg, R. A., Rees, D. I., & Ehrenberg, E. L. (1991). School district leave policies, teacher absenteeism, and student achievement. *Journal of Human Resources, 26*(1), 72–105.

Farkas, G., Grobe, R. P., Sheehan, D., & Shuan, Y. (1990). Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban school district. *American Sociological Review, 55*(1), 127–142.

Ferguson, R. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28*(2), 465–498.

Fowler, W. J., Jr. & Walberg, H. J. (1991). School size, characteristics, and outcomes. *Educational Evaluation and Policy Analysis, 13*(2), 189–202.

Grimes, P. W. (1994, Winter). Public versus private secondary schools and the production of economic education. *Journal of Economic Education, 25*(1), 17–30.

Grimes, P. W., & Register, C. A. (1990, Summer). Teachers' unions and student achievement in high school economics. *Journal of Economic Education, 21*(3), 297–306.

Grimes, P. W, & Register, C. A. (1991). Teacher unions and black students' scores on college entrance exams. *Industrial Relations, 30*(3), 492–500.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review, 60*(2), 280–288.

Hanushek, E. A. (1972). *Education and race: An analysis of the educational production process.* Cambridge, MA: Health–Lexington.

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84–117.

Hanushek, E. A. (1993). Can equity be separated from efficiency in school finance debates? In E. P. Hoffman (Ed.), *Essays on the economics of education* (pp. 35–73). Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics, 78*(4), 611–627.

Harnisch, D. L. (1987). Characteristics associated with effective public high schools. *Journal of Educational Research, 80*(4), 233–241.

Heim, J., & Perl, L. (1974). *The educational production function: Implications for educational manpower policy.* Ithaca, NY: Cornell University.

Henderson, V., Mieszkowski, P., & Sauvageau, Y. (1976). *Peer group effects and educational production functions.* Ottawa, Canada: Economic Council of Canada.

Jencks, C. S., & Brown, M. (1975). Effects of high schools on their students. *Harvard Educational Review, 45*(3), 273–324.

Johnson, G. E., & Stafford, F. P. (1973). Social returns to quantity and quality of schooling. *Journal of Human Resources, 8*(2), 139–155.

Jud, G. D., & Walker, J. L. (1977). Discrimination by race and class and the impact of school quality. *Social Science Quarterly, 57*, 731–749.

Katzman, M. (1971). *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), 1–24.

Kiesling, H. J. (1967, August). Measuring a local government service: 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*(4), 341–350.

Kurth, M. M. (1987). Teachers' unions and excellence in education: An analysis of the decline in SAT scores. *Journal of Labor Research, 8*(4), 351–367.

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

Levin, H. M. (1976). Concepts of economic efficiency and educational production. In J. T. Froomkin, D. T. Jamison, & R. Radner (Eds.), *Education as an industry* (pp. 149–190). Cambridge, MA: Ballinger.

Link, C. R., & Mulligan, J. G. (1986). The merits of a longer school day. *Economics of Education Review, 5*(4), 373–381.

Link, C. R., & Mulligan, J. G. (1991). Classmates' effects on black student achievement in public school classrooms. *Economics of Education Review, 10*(4), 297–310.

Link, C. R., & Ratledge, E. C. (1975). The influence of the quantity and quality of education on black-white earnings differentials: Some new evidence. *Review of Economics and Statistics, 57*(3), 346–350.

Link, C. R., & Ratledge, E. C. (1979). Student perceptions, I.Q., and achievement. *Journal of Human Resources, 14*(1), 98–111.

Lopus, J. S. (1990). Do additional expenditures increase achievement in the high school economics class? *Journal of Economic Education, 21*(3), 277–286.

Maynard, R., & Crawford, D. (1976). School performance. *Rural income maintenance experiment: Final report.* Madison, WI: University of Wisconsin.

Michelson, S. (1970). The association of teacher resources with children's characteristics. In U.S. Office of Education (Ed.), *Do teachers make a difference?* (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*(3), 283–306.

Montmarquette, C., & Mahseredjian, S. (1989). Does school matter for educational achievement? A two-way nested-error components analysis. *Journal of Applied Econometrics, 4*, 181–193.

Morgenstern, R. D. (1973). Direct and indirect effects on earnings of schooling and socio-economic background. *Review of Economics and Statistics, 55*(2), 225–233.

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

Murnane, R. J., & Phillips, B. (1981). What do effective teachers of inner-city children have in common? *Social Science Research, 10*(1), 83–100.

Namboodiri, K., Corwin, R. G., & Dorsten, L. E. (1966, October). Analyzing distributions in school effects research: An empirical illustration. *Sociology of Education, 66,* 278–294.

Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Cadigan, D. (1987, October). Children who do exceptionally well in first grade. *Sociology of Education, 60,* 257–271.

Perl, L. J. (1973). Family background, secondary school expenditure, and student ability. *Journal of Human Resources, 8*(2), 156–180.

Raymond, R. (1968). Determinants of the quality of primary and secondary public education in West Virginia. *Journal of Human Resources, 3*(4), 450–470.

Ribich, T. I., & Murphy, J. L. (1975). The economic returns to increased educational spending. *Journal of Human Resources, 10*(1), 56–77.

Sander, W. (1993). Expenditure and student achievement in Illinois. *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*(2), 111–121.

Sebold, F. D., & Dato, W. (1981). School funding and student achievement: An empirical analysis. *Public Finance Quarterly, 9*(1), 91–105.

Sengupta, J. K., & Sfeir, R. E. (1986). Production frontier estimates of scale in public schools in California. *Economics of Education Review, 5*(3), 297–307.

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.

Stern, D. (1989). Educational cost factors and student achievement in grades 3 and 6: Some new evidence. *Economics of Education Review, 8*(2), 149–158.

Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5*(1), 41–48.

Summers, A., & Wolfe, B. (1977). Do schools make a difference? *American Economic Review, 67*(4), 639–652.

Tuckman, H. P. (1971). High school inputs and their contribution to school performance. *Journal of Human Resources, 6*(4), 490–509.

Wachtel, P. (1975). The effect of school quality on achievement, attainment levels, and lifetime income. *Explorations in Economic Research, 2,* 502–536. (Occasional papers of the National Bureau of Economic Research)

Wachtel, P. (1976). The effect on earnings of school and college investment expenditures. *Review of Economics and Statistics, 58*(3), 326–331.

Walberg, H. J., & Fowler, W. J, Jr. (1987). Expenditure and size efficiencies of public school districts. *Educational Researcher, 16*(7), 5–13.

Walberg, H. J., & Rasher, S. P. (1974, September). Public school effectiveness and equality: New evidence and its implications. *Phi Beta Kappan,* 3–9.

Walberg, H. J., & Walberg, H. J, III. (1994). Losing local control. *Educational Researcher, 23*(5), 19–26.

Watts, M. (1985, Fall). School district inputs and biased estimation of educational production functions. *Journal of Economic Education, 16*(4), 281–286.

Wendling, W., & Cohen, J. (1981, Summer). Education resources and student achievement: Good news for schools. *Journal of Education Finance, 7,* 44–63.

Winkler, D. (1975). Educational achievement and school peer group composition. *Journal of Human Resources, 10*(2), 189–204.

References

Armor, D., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools.* Santa Monica, CA: Rand Corp.

Betts, J. R. (1995). Does school quality matter? Evidence from the National Longitudinal Survey of Youth. *Review of Economics and Statistics, 77*(2), 231–247.

Betts, J. R. (1996). Is there a link between school inputs and earnings? Fresh scrutiny of an old literature. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 141–191). Washington, DC: Brookings.

Bishop, J. (1991). Achievement, test scores, and relative wages. In M. H. Kosters (Ed.), *Workers and their wages* (pp. 146–186). Washington, DC: The AEI Press.

Borland, M. V., & Howsen, R. M. (1992). Student academic achievement and the degree of market concentration in education. *Economics of Education Review, 11*(1), 31–39.

Bowles, S., & Levin, H. M. (1968). The determinants of scholastic achievement—an appraisal of some recent evidence. *Journal of Human Resources, 3*(1), 3–24.

Burtless, G. (1996). Introduction and summary. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 1–42). Washington, DC: Brookings.

Cain, G. G., & Watts, H. W. (1970). Problems in making policy inferences from the Coleman Report. *American Sociological Review, 35*(2), 328–352.

Card, D., & Krueger, A. B. (1992a). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy, 100*(1), 1–40.

Card, D., & Krueger, A. B. (1992b). School quality and black-white relative earnings: A direct assessment. *Quarterly Journal of Economics, 107*(1), 151–200.

Card, D., & Krueger, A. B. (1996). Labor market effects of school quality: Theory and evidence. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 97–140). Washington, DC: Brookings.

Coleman, J. S., Campbell, E. Q., Hobson, C. J, McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1996). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review, 14*(1), 1–21.

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 265–298). Washington, DC: Brookings.

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

Greenberg, D., & McCall, J. (1974). Teacher mobility and allocation. *Journal of Human Resources, 9*(4), 480–502.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*(3), 361–396.

Grogger, J. T. (1996). Does school quality explain the recent black/white wage trend? *Journal of Labor Economics, 14*(2), 231–253.

Grogger, J. T., & Eide, E. (1993). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources, 30*(2), 280–310.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review, 60*(2), 280–288.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources, 14*(3), 351–388.

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

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141–1177.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45–51.

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84–117.

Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries. *World Bank Research Observer, 10*(2), 227–246.

Hanushek, E. A. (1996a). A more complete picture of school resource policies. *Review of Educational Research, 66*(3), 397–409.

Hanushek, E. A. (1996b). School resources and student performance. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 43–73). Washington, DC: Brookings.

Hanushek, E. A., & Kain, J. F. (1972). On the value of "Equality of Educational Opportunity" as a guide to public policy. In F. Mosteller & D. P. Moynihan (Eds.), *On equality of educational opportunity* (pp. 116–145). New York: Random House.

Hanushek, E. A., & Kim, D. (1996, December). *Schooling, labor force quality, and the growth of nations.* Rochester, NY: Department of Economics, University of Rochester.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics, 78*(4), 611–627.

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

Heckman, J. S., Layne–Farrar, A., & Todd, P. (1996a) Does measured school quality really matter? An examination of the earnings-quality relationship. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 192–289). Washington, DC: Brookings.

Heckman, J. S., Layne–Farrar, A., & Todd, P. (1996b). Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. *Review of Economics and Statistics, 78*(4), 562–610.

Hedges, L. V. (1990). Directions for future methodology. In K. W. Wachter & M. L. Straf (Eds.), *The future of meta-analysis* (pp. 11–26). New York: Russell Sage.

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. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

Hoxby, C. M. (1996). How teachers' unions affect education production. *Quarterly Journal of Economics, 111*(3), 671–718.

Johnson, G. E., & Stafford, F. P. (1973). Social returns to quantity and quality of schooling. *Journal of Human Resources, 8*(2), 139–155.

Lamdin, D. J. (1995).Testing for the effect of school size on student achievement within a school district. *Education Economics, 3*(1), 33–42.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2), 113–127.

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

Murnane, R. J. (1981). Teacher mobility revisited. *Journal of Human Resources, 16*(1), 3–19.

Murnane, R. J., & Phillips, B. (1981). What do effective teachers of inner-city children have in common? *Social Science Research, 10*(1), 83–100.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics, 77*(2), 251–266.

Neal, D. A., & Johnson, W. R. (1996). The role of premarket factors in Black-White differences. *Journal of Political Economy, 104*(5), 869–895.

O'Neill, J. (1990). The role of human capital in earnings differences between Black and White men. *Journal of Economic Perspectives, 4*(4), 25–46.

Peltzman, S. (1993, April). The political economy of the decline of American public education. *Journal of Law and Economics, 36,* 331–370.

Sebold, F. D., & Dato, W. (1981). School funding and student achievement: An empirical analysis. *Public Finance Quarterly, 9*(1), 91–105.

Smith, M. S., Scoll, B. W., & Link, J. (1996). Research-based school reform: The Clinton administration's agenda. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp. 9–27). Washington, DC: National Academy Press.

Staley, S. R., & Blair, J. P. (1995). Institutions, quality competition and public service provision: The case of public education. *Constitutional Political Economy, 6,* 21–33.

Summers, A. A., & Johnson, A. W. (1996). The effects of school-based management plans. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp. 75–96). Washington, DC: National Academy Press.

U.S. Department of Education. (1990). *Digest of education statistics, 1994.* Washington, DC: National Center for Education Statistics.

Wachtel, P. (1976). The effect on earnings of school and college investment expenditures. *Review of Economics and Statistics, 58*(3), 326–331.

Welch, F. (1966). Measurement of the quality of schooling. *American Economic Review, 56*(2), 379–392.

Word, E., Johnston, J., Bain, H. P., Fulton, B. D., Zaharies, J. B., Lintz, M. N., Achilles, C. M., Folger, J., & Breda, C. (1990). *Student/teacher achievement ratio (STAR), Tennessee's K–3 class size study: Final summary report, 1985–1990.* Nashville, TN: Tennessee State Department of Education.

Author

ERIC A. HANUSHEK is professor of economics and public policy at the Wallis Institute of Political Economy at the University of Rochester, Harkness Hall 107, Rochester, NY 14627-0158. His areas of specialization are public finance and education policy.

Received September 15, 1996
Revision received January 15, 1997
Accepted January 30, 1997