Review of Educational Research
Fall 1996, Vol. 66, No. 3, pp. 397–409

A More Complete Picture of School Resource Policies

Eric A. Hanushek
University of Rochester

Meta-analytic techniques for summarizing information about the relationship between school resources and student performance are capable only of addressing the narrow and uninteresting hypothesis that resources are never used effectively by schools. The data show clearly that resources are sometimes used effectively, although this happens infrequently and there is no description of the circumstances under which resources are used effectively. This article relates the basic evidence on school effectiveness to the specific application of meta-analytic methods employed by Greenwald, Hedges, and Laine (1996). Their analysis, suffering from the narrowness of the inquiry inherent in their statistical methods, is also based on a very highly selected sample of results that biases their analysis precisely toward their conclusions. As a result, their summary of existing work provides a distorted and misleading view of the potential implications of school resource policies. Both detailed econometric evidence and aggregate performance of U.S. schools point toward serious problems with inefficient use of resources. This evidence in turn suggests that lack of resources is not the largest problem facing schools and that more fundamental reforms are needed in schools.

Discussions of school resource policies have been marked by confusion and controversy. The result has been waves of popular and highly regarded policies that have been generally ineffective and wasteful. The article by Greenwald, Hedges, and Laine (1996) furthers the confusion and offers a rationalization for those who would—to the detriment of the nation—extend the policies of the past.

Concern about the continuing misallocation of education resources led the Panel on the Economics of Education Reform (PEER) to call for fundamental changes in views about education policy. The PEER report, *Making Schools Work* (Hanushek, 1994), concludes that developing more effective schools is crucial to the future health of the U.S. economy. At the same time, the current structure of schools, with a lack of consequential performance incentives and with a tradition of not learning from the alternative approaches and programs that are tried, offers little reason for optimism unless there is a real change in focus.

The conclusions in *Making Schools Work* (Hanushek, 1994) are partly based on the evidence that Greenwald, Hedges, and Laine considered in their article, so it is useful to see how their manipulations and interpretations systematically distort the conclusions that should be drawn from the evidence. The central substantive positions running through their article are (a) that U.S. schools have been working

This discussion benefitted from helpful comments by James Heckman.

quite well, (b) that schools have been providing a good return on expenditure, (c) that any performance problems of students are best attributed to poorer students and parents and not to the schools, and (d) implicitly, that more resources devoted to the current schools would be productive and would be a wise investment for society to make. Greenwald, Hedges, and Laine's statistical analysis appears to lend scientific credence to other recent attempts to extol the current and continuing successes of the American elementary and secondary schools. Unfortunately, the evidence does not support any of their central positions.

This note describes how their evidence relates to issues of school policy. There are two primary themes. First, they misinterpret the implications of their analysis. Second, through a series of analytical choices, they systematically bias their results toward the conclusions they are seeking.

Ultimately, the fundamental problem with their analysis derives from a flawed statistical approach for investigating issues of how and when resources affect student performance. Their specialized meta-analytic approach to combining data is applicable to circumstances very different from the present ones. They assume that all of the schooling situations are identical, when in fact most people believe for good reason that they are very heterogeneous. They further assume that all of the studies should receive equal weight, when in fact the studies are also heterogeneous. If there were a series of independent laboratories conducting exactly the same experiment, and if there were a desire to combine the separate statistical tests in the absence of the original data, then the approaches of Greenwald, Hedges, and Laine might be appropriate. That is not close to the situation faced in the study of the effectiveness of resource usage. By forcing homogeneity onto the data about effectiveness, they both introduce powerful biases into their analysis of the results and distract decision makers from the important issues.

Focus of Analysis

The most basic problem with their statistical analysis is that it addresses a completely uninteresting question—one that has little relevance from a policy viewpoint. The central hypothesis of their analysis is never explicitly discussed, although they attempt to suggest that it 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—that is, H_0: $\beta_1 = \beta_2 = \cdots = \beta_n = 0$, 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, then the null hypothesis is false (that is, somewhere there is an effect on student performance). Their statistical procedures are designed in such a case to reject the null hypothesis, which leads to acceptance of the alternative that at least one study indicated somewhere that the resource was related to performance. 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."

Virtually everybody who has looked at schools is convinced that some schools use resources more effectively than others, an observation central to the entire policy discussion in *Making Schools Work* (Hanushek, 1994). The analysis of Greenwald, Hedges, and Laine, after considerable effort at manipulating the data from the underlying studies, finally rejects the hypothesis that resources never matter—and they attempt to suggest that this is a revelation based on the power of their statistical approach. In reality, it is obvious from the underlying distribution of prior results. Table 1 displays a summary of results from a complete set of studies published through the end of 1994.[1] Underlying this table are 377 separate estimates of the effects of one or more basic resources on student performance. They are found in 90 individual published articles or books.

That resources sometimes appear to matter is found, for example, by looking at the 277 separate estimates of the effects of teacher-pupil ratios on student performance. Fifteen percent find a positive and statistically significant effect of variations in teacher-pupil ratios. If all underlying parameters were identically zero (the basic null hypothesis of Greenwald, Hedges, and Laine), one would expect only 2½% to be positive and statistically significant by chance. This is exactly what *significant at the 95 percent level* means: By chance, 2½% of the time one erroneously rejects the null hypothesis with a large positive estimate, and 2½% of the time one erroneously rejects that null hypothesis with a large negative estimate even though the true effect is zero. Thus, most people who look at these data, including me, conclude without going further that the evidence demonstrates that resources are used productively in some circumstances.

It is a simple fact, however, that neither Greenwald, Hedges, and Laine nor anybody else can readily describe when resources will be used effectively. The estimated effectiveness of teacher-pupil ratios in Table 1 indicates that 13% of the studies find a statistically significant negative relationship with student performance. Fully 85% of the estimates are found either to have the "wrong" sign or to be statistically insignificant, and they are quite evenly distributed around zero. If we knew what distinguishes the 15% from the 85%, we might be able to craft policies that ensured that reductions in teacher-pupil ratios were accompanied by improvements in student performance (and not just the increases in costs that

TABLE 1
Percentage distribution of estimated effect of key resources on student performance, based on 377 studies

		Statistically significant		Statistically insignificant		
Resources	Number of estimates	Positive (%)	Negative (%)	Positive (%)	Negative (%)	Unknown sign (%)
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
Teacher salary	119	20	7	25	20	28
Expenditure per pupil	163	27	7	34	19	13
Administrative inputs	75	12	5	23	28	32
Facilities	91	9	5	23	19	44

automatically follow increasing the teacher-pupil ratio). We do not know how to ensure improvements. We do know from this and other evidence that the normal choices and operations of schools provide little reason to expect increases in class sizes to yield improved performance.[2] The same conclusions hold for the other resources identified in Table 1.

Greenwald, Hedges, and Laine would say that this evidence shows that money matters, because more studies indicate a positive relationship than would be expected by chance. They even go so far as to say that evidence showing more negative relationships than expected by chance also indicates that money matters, and represents a contradiction to my general conclusion that there is "no strong or systematic relationship" between resources and student achievement. They are, of course, free to develop their own terminology, but this choice is both very unnatural and very unfortunate from a policy perspective.

Sample Selection

It is actually a bit surprising how much effort it appears to take to reach the obvious conclusions from the data. Table 2 compares the distribution of results from the full set of estimated parameters that is available in Table 1 and the distribution of results from the sample that Greenwald, Hedges, and Laine ultimately used for their combined significance analysis. As can be seen in the second and third columns, Greenwald, Hedges, and Laine's analysis relies on between 17% and 30% of the available data; that is, their main analysis begins by eliminating 70% to 83% of the available studies. More importantly, it is a very selective sampling of available results. Table 3 shows the selection percentages, reflecting the proportion of available studies (by results) that are used by Greenwald, Hedges, and Laine. 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 3). This action by itself eliminates 13% to 26% of the available data. Clearly, since they are out to show that there is a statistically significant relationship, the preliminary elimination of substantial evidence to the contrary biases the results in favor of their perspective. Second, the sample selection process uniformly retains a higher proportion of the statistically significant positive results than of the overall results. In the cases of teacher education and of per-pupil expenditure, the selection rate for statistically significant positive results is literally double the overall selection rate. While they retain just 22% of the available estimates of the effects of teacher education, they retain fully 44% of those that show a positive and statistically significant effect. Similarly, for per-pupil expenditure, 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, Hedges, and Laine retain a lower proportion of statistically significant negative results than of the overall results. Moreover, among the insignificant results, the selection again tends to retain a relatively higher proportion of the positive estimates than of the negative estimates (with the minor exception of essentially equal selection rates for per-pupil expenditure). Substantial sections of the discussion of the underlying statistical theory behind meta-analysis in Hedges and Olkin (1985) and Hedges (1990) are devoted to the pitfalls of incomplete and nonrandom selection of results. The overall selection of results in Greenwald, Hedges, and Laine's study

TABLE 2

Total number of studies available and number of studies used in Greenwald et al.'s combined significance analysis

	Total number of estimates		Statistically significant				Statistically insignificant					
			Positive		Negative		Positive		Negative		Unknown sign	
Resources	Avail.	GHL	Avail.	GHL	Avail.	GHL	Avail.	GHL	Avail.	GHL	Avail.	GHL
Teacher-pupil ratio	277	64	41	13	37	7	74	32	70	12	55	0
Teacher education	171	38	16	7	9	6	56	15	46	10	44	0
Teacher experience	207	60	60	18	10	2	63	25	50	15	24	0
Expenditure per pupil	163	27	44	15	11	1	56	5	31	6	21	0

Note. Avail. refers to studies in the complete tabulation of results in Table 1. *GHL* refers to actual number of studies used in the full analysis of combined significance tests by Greenwald, Hedges, & Laine (1996).

TABLE 3
Selection rates for studies employed by Greenwald et al., total, and by results (percentages)

		Statistically significant		Statistically insignificant		
Resources	Total estimates (%)	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

is dramatically biased toward retaining both statistically significant positive and insignificant but positive results—just the direction that leads to support for their general conclusions.

They further reduce the sample for their "robustness samples" by eliminating the tails of the distribution, a procedure that has no statistical justification in this application. This eliminates another 10% of the data on resource effects.[3] The samples used for the analysis of effect sizes are almost always even smaller (but just as selective) as the comparable samples for joint confidence tests.

The sample selection rules that they apply represent a combination of arbitrary decisions based on their particular choice of methods and choices explicitly or implicitly based on the data themselves. While these selection rules tend to have some surface plausibility, none is based on explicit analysis of the underlying statistical properties of the data and appropriateness for the actual statistical methods employed. The only thing that is absolutely clear is that the selection rules systematically bias the results toward their conclusions.

Distribution of Results

Some consideration must be given to the distribution of the overall results. Greenwald, Hedges, and Laine, as discussed, consider that there is a single parameter (for each resource) that is being estimated across different studies. As noted by them, if this parameter were truly zero and if the statistical tests were appropriate, the hypothesis should be erroneously rejected 5% of the time (i.e., the size of the Type I error). Why is it that there are both more positive *and* more negative rejections than the 2½% called for by the simple case? It cannot just be that the common parameter is greater than zero, implying that there are more positive rejections than would be expected with pure randomness. In such a case, the larger proportion of positive rejections would be balanced by fewer than 2½% negative rejections; that is, the distribution would simply be shifted.

My interpretation is that there is actually a distribution of underlying parameters, that is, that there is an underlying heterogeneity in the use of resources. In certain circumstances resources are used effectively. In many they are not used well at all. And in some they are employed in ways that are actually harmful to achievement.[4] In this case, the policy question is how to identify or select situations that involve effective use of resources and discard others. But there are also other possible explanations.

Two other explanations are important, and probably contribute to the overall distribution of estimates in Table 1. First, Hedges and Olkin (1985) and Hedges (1990) stress the possible biases of both effect size estimates and of joint confidence tests that arise because of publication bias. For this, Hedges's (1990) own 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. (p. 19, references omitted)

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, even independently of Greenwald, Hedges, and Laine's subsequent sample selection, the published results underlying the summaries in Table 1 would overstate the magnitude and significance of each of the resource effects.

Second, the results may also arise from systematic biases in the underlying parameter estimates. One possible source of such effects is the uniform omission of any measures of state education policies. Each state in the United States maintains its own independent policy toward schools, as expressed by separate school regulations and laws, by different financing formulas, by varying graduation requirements, by general labor policies, and the like. If these policies have an impact on student performance, their omission from modeling could bias the results. The form of the bias is particularly important, however, because analyses within an individual state—where all schools face the same policy environment—will not be affected, but analyses across states will. There is no a priori reason to expect this omission to have a given upward or downward bias (because the bias depends on the unknown correlation of policy and resources). The empirical fact nonetheless is that this omission leads to a systematic upward bias in estimated resources effects. Hanushek, Rivkin, and Taylor (in press) both develop the theory behind this and indicate how the evidence from Table 1 is affected. Table 4 displays the results for the estimated effects of teacher-pupil ratio and per-pupil expenditure to illustrate how model specification dramatically affects the overall statistical results. The individual estimates of resource effects displayed in Table 1 come from databases that sometimes are drawn entirely from within a single state and at other times are drawn from across state boundaries. The results for data drawn from schools entirely within a single state are noticeably less likely to find statistically significant resource effects and are much more likely to find

TABLE 4
Percentage distribution of estimated effect of teacher-pupil ratio and expenditure per pupil by state sampling scheme and aggregation

		Statistically significant		Statistically insignificant		
State sampling scheme	Number of estimates	Positive (%)	Negative (%)	Positive (%)	Negative (%)	Unknown sign (%)
		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
		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	24

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.

statistically significant negative effects than those drawn from schools found in multiple states. Table 4 demonstrates that the positive and statistically significant results for both teacher-pupil ratios and per-pupil expenditure come disproportionately from studies that analyze resource effects across state boundaries but that do not include any measure of state policy factors. In such a situation, the estimated resource effects are biased, so there is no reason to believe that the overall distribution of test statistics matches the theoretical underlying distribution (that is derived on the assumption of unbiased parameter estimates). Other specification problems would similarly lead to biased results and could contribute to the excessive proportion of significant estimates in Table 1.

It is important to distinguish among the various explanations for finding too many statistically significant results, because each has different implications for interpretation of the results and for education policy. In all cases, however, the particular analytical choices of Greenwald, Hedges, and Laine lead to distinct upward biases in their statistical tests and in their estimates of effect sizes.

Social Capital

The available aggregate data for U.S. schools provides troublesome evidence counter to Greenwald, Hedges, and Laine's claim that resources and expenditure have a powerful effect on student performance. As Table 5 indicates, U.S. schools have noticeably expanded total (real) spending along with the key resources analyzed in the many studies of student performance. In the quarter century from 1965 to 1990, spending per pupil (after accounting for inflation) more than doubled. Yet, as Greenwald, Hedges, and Laine point out, achievement is at best flat. SAT scores dropped precipitously from the mid-1960s through 1980, and the

TABLE 5
Public school resources, 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
Percentage of teachers with master's degree	23.1	23.2	27.1	37.1	49.3	50.7	52.6
Median years teacher experience	11	8	8	8	12	15	15
Current expenditure/ ADA (1992–93 $s)	$1,903	$2,402	$3,269	$3,864	$4,116	$4,919	$5,582

Note. Data from *The Condition of Education, 1994,* by U.S. Department of Education, National Center for Education Statistics, 1994, Washington, DC: U.S. Government Printing Office.

subsequent recovery leaves scores noticeably below the 1965 level. SAT scores have been subject to selective test taking, but the National Assessment of Educational Progress (NAEP) has not. On the NAEP tests, from 1970 to 1994 mathematics performance went slightly up, reading was constant, and science went down.

Greenwald, Hedges, and Laine claim these trends show the effectiveness of spending. Their claim is two-fold. First, kids have gotten worse, so the fact that performance is flat must reflect effective spending. Second, the Black-White score differential has narrowed, again showing the effectiveness of spending.

The argument that kids have gotten worse—which they somewhat pretentiously call social capital—is based on the observation that female labor force participation rates and the proportion of single parent families rose over the period, implying to them poorer family inputs to kids' education. At the same time, they ignore the fact that the education level of parents rose over the period and the average family size fell dramatically. These and other offsetting factors are presumably also relevant to social capital. Netting out the separate effects is difficult, because it is necessary to have estimates of the effects of various family inputs on achievement in order to weight the trends properly. One attempt to do so (Grissmer, Kirby, Berends, & Williamson, 1994) suggests that on average the positive factors outweigh the negative factors, which implies that the ineffectiveness of spending is even worse than shown in the raw aggregate trend data. This type of analysis is difficult to do with any precision, but their results blunt the casual assertions that things have necessarily gotten worse.[5]

There is little evidence that the growth in spending has been disproportionately aimed at Blacks. The narrowing of Black-White differences then does not provide an answer to the aggregate trends either, unless one presumes that spending on Blacks has been much more effective than spending on Whites. Direct analysis of the narrowing of Black-White differences in NAEP scores indicates that differences in spending across districts is not an important explanation (Cook, 1995).

Greenwald, Hedges, and Laine go to considerable lengths to estimate the effect sizes for per-pupil expenditure and the other resources in Table 5. They calculate significant positive effects. But for the past three decades we have been running the precise spending experiment that they would want to be future policy. There is no indication that student performance has increased at all, let alone by the magnitude they would predict.

Longitudinal Studies

Researchers investigating the determinants of school achievement have advocated longitudinal designs that consider how an individual student's achievement grows and changes over time. This approach has distinct advantages because it does not require complete knowledge of past family and school inputs and because it provides a method of correcting for differences in individual abilities (Hanushek, 1979; Murnane, 1981). Greenwald, Hedges, and Laine allude to these rationales and produce separate results according to whether individual studies are longitudinal or quasi-longitudinal. In doing this, however, they neglect to provide a complete description of the studies they include in this category and of their interpretation.

Greenwald, Hedges, and Laine place great emphasis on the findings for per-pupil expenditure, but this is a mistake (both for longitudinal studies and for the total sample). Per-pupil expenditure is not a statistic that is ever calculated for individual classrooms. It is only very rarely calculated for individual schools. This measure is a financial variable for school districts. Moreover, it cannot readily be divided even by primary or secondary school level. Therefore, none of the studies of per-pupil expenditure involve analyses of resources at the individual classroom level. Most commonly they are aggregate studies at the school district level.[6] If these studies consider individual outcomes as opposed to aggregate student achievement, they still do not employ measures of the resources available to the individual student but instead the average for the district. This causes a variety of problems (Hanushek et al., in press), but it also highlights the improper summaries of Greenwald, Hedges, and Laine. The general problems are worsened when they consider longitudinal and quasi-longitudinal studies. In aggregate studies, longitudinal gains by individual students are seldom if ever available. Instead, some researchers have included average test scores for a different group of students at some earlier grade. Thus, for example, models explaining today's sixth-grade reading performance might include a measure of reading performance of today's third graders. Greenwald, Hedges, and Laine would call this quasi-longitudinal, but it bears no relationship to the justification for longitudinal analysis. It is not surprising, given the flawed methodology, that truly longitudinal models yield negative resource effects but that Greenwald, Hedges, and Laine can turn this around by including quasi-longitudinal studies.

The consideration of longitudinal analyses could be viewed as an attempt by Greenwald, Hedges, and Laine to weight the results differently by quality of the underlying study were it not for the fact that they mix together both the best and the worst studies through their selection criteria. When done appropriately, the overall conclusions about the lack of a consistent relationship are unchanged (Hanushek, 1996).

Specific Meta-Analytic Methodology

The obviously preferred way to combine the results of the many studies of student achievement would be to begin with the raw data for these studies and to reestimate joint models of performance. This approach is impractical, so both Greenwald, Hedges, and Laine and I have attempted to summarize the findings from the published results. Using just published information is clearly limiting, which leads to my simple presentation of distributional information without any further refinement (Hanushek, 1986, 1989). Greenwald, Hedges, and Laine go a different route by choosing a very specialized statistical approach to the aggregation of results. They attempt to provide formal statistical tests (albeit of a relatively uninformative set of hypotheses). They calculate combined significance tests using the Fisher approach. A key element of this method is that it requires independence of the separate estimates. This requirement of the methodology they chose lies behind some of the sample selection discussed earlier. But simply because it is required by their specialized procedure does not imply that there is no information in the results they discard along the way. Instead, one might simply conclude that their specific choice of a statistical approach is inappropriate.

The really unusual aspect of their approach is their attempt to deal with the large proportion of both positive and negative results. The inescapable fact from Table 1 is that results are rather evenly distributed around zero, even if there are more significantly negative and more significantly positive results than would occur by chance with homogeneity of effects. Their methods cannot handle two-sided tests, so they must invent an approach to deal with an obvious characteristic of the data. Their choice is applying separate one-tailed testing to positive and to negative estimates. One-tailed tests in statistics are used when there is outside information that indicates only positive or only negative results are appropriate. Greenwald, Hedges, and Laine have no such information, and their approach completely lacks justification. Their testing procedure involves selecting the test statistic on the basis of the observed data, a procedure that cannot yield unbiased tests. Interestingly, no mention of such an approach appears in the comprehensive treatment of meta-analysis by Hedges and Olkin (1985), and the approach cannot be found in standard statistical texts. Greenwald, Hedges, and Laine are driven to this procedure because the data for estimated results do not conform to the underlying distributions assumed in their statistical procedures. Their determination to apply the procedure they know—even when inappropriate for the data—seems misguided. An alternative approach that incorporates the fundamental heterogeneity and the dependencies in the data, such as empirical Bayes methods, would be much more productive and might actually provide useful guidance to policymakers.

Policy Issues

If there is a low likelihood that increased resources will be effective in any specific circumstance, it would be foolish policy to employ pure resource policies. Instead, as stressed in *Making Schools Work* (Hanushek, 1994), policies that point toward effective resource use should be the focus of attention. The inability to identify why resources count at some times and not at others suggests that more use should be made of decentralized performance incentives. In any event, the key conclusion remains: How resources are used will be more important than how many resources are used, at least within the context of current levels of basic resources for schools.

Those people who can look at the data of large increases in resources without improvements in student performance and still call for further continuation of pure resource policies are surely leading us astray.

Notes

[1]To be included, an underlying study must minimally include separate measures of family background (in order to distinguish between the effects of families and those of schools); must relate to an objective measure of student performance, such as standardized test scores or subsequent wages; must investigate the effect of one of the key resource measures included below; and must provide information about the statistical significance of any estimated effect (see Hanushek, 1989). These criteria, which are close but not identical to those of Greenwald, Hedges, and Laine, yield a slightly larger sample of publications than that used by Greenwald, Hedges, and Laine. Greenwald, Hedges, and Laine exclude some studies that use aggregate data and miss some others. They also include some that are not included here because they do not require that underlying studies incorporate direct measures of differences of family background if there is a prior achievement measure.

The data in Table 1 are from *Assessing the Effects of School Resources on Student Performance: An Update* (Working Paper No. 424), by E. A. Hanushek, 1996, Rochester, NY: University of Rochester, Rochester Center for Economic Research. [Editor's note: That paper was made available after Greenwald, Hedges, and Laine had written their rejoinder, which appears on pages 411–416 of this issue of *Review of Educational Research.*]

[2]Some controversy exists about the correct interpretation of the Tennessee STAR experiment in class size reductions. In large part the controversy comes from the failure to observe children who were first in small classes and later in larger classes. See Hanushek (1994) and Word et al. (1990).

[3]When the discussion turns to effect sizes, there is interest in analyzing whether or not outliers are influencing the results. In this case the trimming would look more plausible if it were not for the fact that they use medians of the estimates (presumably to limit the influence of outliers). Using both together makes no sense.

[4]It may be possible to think in terms of an underlying distribution of true parameters with an interest in estimating the mean of that distribution. This approach would, nonetheless, require different statistical procedures than employed by Greenwald, Hedges, and Laine.

[5]Grissmer et al. (1994) estimate cross-sectional models of the effects of individual family factors on achievement. They then use the weights from these models to examine the trends in various family factors. They find that family factors have improved somewhat for Whites and the opposite for Blacks. This implies that the overall results, which are most reflective of the majority White population, should have improved on the basis of just family factors. Their analysis, however, does not explicitly measure school factors. They attribute the difference between actual performance and that predicted on the basis of family factors to schools, where it may also reflect other unmeasured factors.

[6]Greenwald, Hedges, and Laine do exclude analyses aggregated to the state level.

References

Cook, M. (1995, November). *Families or schools?: Accounting for the convergence in White and Black test scores since 1970.* Paper presented at the meeting of the Association for Public Policy Analysis and Management, Washington, DC.

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

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

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educa-

tional production functions. *Journal of Human Resources, 14,* 351–388.
Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24,* 1141–1177.
Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45–51.
Hanushek, E. A. (with others). (1994). *Making schools work: Improving performance and controlling costs.* Washington, DC: Brookings Institution.
Hanushek, E. A. (1996). 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 Institution.
Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (in press). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics.*
Hedges, L. V. (1990). Directions for future methodology. In K. W. Wachter & M. L. Straf (Ed.), *The future of meta-analysis* (pp. 11–26). New York: Russell Sage Foundation.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.
Murnane, R. J. (1981). Interpreting the evidence on school effectiveness. *Teachers College Record, 83,* 19–35.
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, Wallis Institute of Political Economy and Department of Economics, University of Rochester, Rochester, NY 14627-0158; hanu@troi.cc.rochester.edu. He specializes in public finance and labor economics.

Received May 14, 1996
Accepted June 20, 1996