Policy Evaluation Versus Explanation of Outcomes in Education:

I.e. Is it the teachers? Is it the parents?

A policy brief

Richard Startz*

Revised February 2012

Abstract

Education reform advocates who base policy decisions on empirical research often face the argument that because background variables explain so much of student outcomes, it follows that policy interventions cannot be effective. This policy brief explains the logical fallacy in the argument, illustrating with two examples, one taken from the teacher quality literature and one taken from the class size literature.

* Department of Economics, University of California, Santa Barbara, CA. 93111. email: startz@econ.ucsb.edu. Brief, popularized versions of the arguments here have appeared online at http://profitofeducation.org/?p=708 and http://www.nctq.org/p/tqb/viewStory.jsp?id=26260. Comments from Dan Goldhaber, Arthur McKee, Doug Steigerwald, Kate Walsh, the UCSB Econometrics Working Group, and two anonymous referees are particularly appreciated.

1. Introduction

Consider the following two statements:

- Most of the variability in student outcomes is explained by out-of-school factors (e.g., socio-economic status).
- Policy reforms, for example measures to raise the quality of teachers, are effective means of improving student outcomes.

Both of these statements are true. Unfortunately, the widespread (correct) belief in the importance of out-of-school factors is often used to (incorrectly) argue that this implies there is relatively little room left for effective policy measures. This brief will explain how the importance of out-of-school factors and the effectiveness of policy are mutually consistent, and therefore explain that the commonly advanced dichotomy between “fix poverty” and “fix schools” approaches is false.

Advocates of a variety of educational reforms often face opposition based on the argument that success in school is overwhelmingly explained by background measures: socio-economic status, parental contributions toward education, community conditions, etc. The argument goes that if background does a much better job of explaining student outcomes than does teacher quality or class size or school organization, or other inside-the-system items on a reform agenda, then there is little point in pursuing a school-based reform agenda. In point of fact, research quite often does show that background variables explain a much higher fraction of student outcomes than is explained by policy or public school resources. However, decomposing explanation of outcomes into component parts depends on both the impact of each input and on how much that input varied in the observed data. In choosing among policies we care about the impact of each input (and their respective costs), but how much the inputs have varied historically isn't relevant.

Consider the following example. Suppose we looked at student outcomes in a school with uniformly good teachers but large variation in family backgrounds. All the variation in student outcomes would be attributable to family background. But this doesn't speak to the question of what would have happened if the school had even better teachers.

Now suppose we had a school in which both teacher quality and family background varied and that we were allowed to randomly match teachers and students so that each classroom had students with the same range of family backgrounds. All the variation of student outcomes within a classroom would *still* be attributable to different family backgrounds. But the average difference in outcomes across classrooms would tell us how much difference a good teacher makes (subject to the usual issues of statistical variation and the possibility that some other than the quality of the teacher also varied across classrooms.) If a good teacher makes a big difference, and if we have an effective way of raising average teacher quality, then it's worth investing in teacher-oriented policies.

The need to untangle the confusion between policy effects and analysis of variance seems to be a permanent feature across a variety of social science topics; see, for example, Manski (2011). In education, there is a long history of respected policy advocates and social scientists focusing on the extent to which variation is explained by non-school factors. Hanushek and Kain describe this as the "no-school-effect conclusion."[1] An early example is

> It is known that socioeconomic factors bear a strong relation to academic achievement. When these factors are statistically controlled, however, it appears that differences between schools account for only a small fraction of differences in pupil achievement. (Coleman, et. al. 1966).[2]

Hanushek and Kain go on to explain

> The closest the [Coleman] *Report* comes to identifying policy instruments are its estimates of the …contribution of individual variables to explained variance….The proportion of explained variance does not identify policy instruments and gives little indication of the policy leverage provided by different variables. Parameter estimates are much more useful in this respect.…Explained variance…is simply not a very interesting concept…to the policy-maker…[3]

I present first a verbal exposition of the fallacy of equating policy evaluation with the decomposition of explanatory power into different components. Put simply, policy evaluation asks questions about the relative benefit/cost ratios of potential interventions. In contrast, the decomposition of explanatory power relies on breakdowns of statistical correlations (formally

[1] Hanushek and Kain (1972), footnote 3.

[2] Coleman et. al. 1966, pp. 21-22 and cited in Hanushek and Kain (1972), footnote 3. Note that "differences in schools" presumably references differences by race. Ehrenberg and Brewer (1995) show that teacher ability mattered even in the original Coleman data.

[3] Hanushek and Kain, op. cit., page 136.

or informally). The two measures are not entirely unrelated, as both are affected by the measured effect of inputs on outputs. But other elements also enter the two calculations in different ways. In the field of education reform, the practical matter is that the two calculations often lead to different conclusions as to which inputs are important. For those interested in school change, policy evaluation measures are what should count.

I next review two examples from the literature, one looking at teacher quality and one looking at class size reduction, that illustrate important policy effects even though the great bulk of student outcomes are explained by background variables. In the last section I present the mathematical version of the verbal argument, which for those who are used to formulating models mathematically is probably simpler and which may also be useful for teaching purposes. An appendix provides a review of how breakdowns of explanatory power can be done and why they may or may not be unique and why they may or may not be causal.

2. Explained variation versus policy impact

School outcomes depend on school inputs, some of which are under the control of policy makers and others of which can be considered as background variables. Empirical researchers measure the relation between inputs and outcomes using two quite distinct concepts: policy impact and decomposition of variability in outcomes amongst the various inputs. “Policy impact” answers the question “If we change policy by so much, by how much would a specified outcome change?” “Decomposition of variability” answers the question “How much of the variation in outcome can we “explain” by each input?” For example, in a multiple regression policy impact is measured by the size of the estimated coefficient on the particular policy. The R^2 of the regression decomposes the variability of the outcome into the fraction jointly explained by all observable variables and the remaining fraction which is unexplained.

Schools produce many different outcomes, some easily measured and some on which we have very little data. For ease of illustration, I focus on a single outcome, for example, value-added measures (VAM) of student achievement. Inputs relevant to the outcome for a particular student include the quality of that student's teachers (a policy variable), family income (a background variable[4]), etc. Researchers might estimate the link between outcome and all inputs using a multiple regression or some other statistical technique which gives an estimate of how much a change in an explanatory variable would change the measured outcome.[5] For example, if family income is measured in thousands of dollars and student VAM scores are measured in months-of-schooling equivalents, then an "family income impact" measure of 0.02 would mean that a $1,000 increase in family income on average raises a VAM score by the equivalent of two hundredths of a month of normal grade progression.[6] Suppose that teacher "quality" is measured by a hypothetical national percentile ranking, so that a "teacher quality impact measure" of 0.05 would mean that having a teacher ranked at the 50th percentile as

[4] I use "family income" as short-hand that is easy to quantify. Mayer (1997) argues that a variety of family characteristics and behaviors, while correlated with family income, are more important than income per se.

[5] Everything here assumes that the researcher has identified causal impacts. A simple regression of outcomes on teacher quality might merely pick up an association due to students from higher socioeconomic status families being assigned to better teachers. Similarly, measures of family income or socioeconomic status may be proxies for associated family behaviors that are actually causal. Identifying causal relations is necessary for the kind of stylized policy exercise discussed here to be meaningful. Most modern economic research confronts this risk directly by careful attention to identifying assumptions and/or use of experimental or quasi-experimental data.

[6] This is simply a verbal description of the economist's "production function," using "impact" for what economists term "marginal product." I assume here that marginal effects are constant even though researchers occasionally have good enough data to identify nonlinearities and interaction effects.

compared to a teacher ranked at the 49th percentile would on average improve a student's VAM score by the equivalent of five hundredths of a month of normal grade progression.

Turn now to the breakdown of variation in VAM into the part explained by family incomes and the part explained by teacher quality. Suppose we compare a well-off school district to a not so well-off district. Assume for purposes of illustration that in the former district average family income is $100,000 and the average teacher ranks at the 60th percentile on the national scale, while in the latter district average income is $50,000 and the average teacher ranks at the 50th percentile. The hypothetical teacher percentile rankings and family income differences are chosen as round numbers for convenient illustration. Nonetheless, it may be interesting to note that in 2008 $50,000 was about the 44th percentile of family income and $100,000 was the 72nd percentile.[7] So in the illustration the difference in family income (28 percentile points) in the two districts is larger than the difference in teacher quality (10 percentile points) when each is measured according to their positions in their respective distributions.

On average, students in the wealthy district will have a grade-equivalent progression 1.5 months greater than that of students in the less wealthy district. Two-thirds of the difference $(1.0 = 0.02 \times [100 - 50])$ is attributable to family income and only one-third $(0.5 = 0.05 \times [60 - 50])$ to a factor controllable by the educational system. In other words, it is true in this illustration that background variables account for most of the difference in student outcomes.

[7] *Statistical Abstract of the United States,* Table 695.

Since background variables dominate teacher quality as a source explaining student outcomes, does it follow that improving teacher quality is the less attractive route for policy intervention? No, decomposing explanation of outcomes into component parts depends on both impact–which in this example is greater for teacher quality—and variation in the input—in this example family background has greater variation. Policy intervention depends on impact, but how inputs varied in the data sample is not relevant.

What is relevant for policy intervention, but not for decomposition of explanatory power, is the cost of varying the relevant input by policy intervention. Suppose average teacher quality could be improved in the lower-performing district by 10 points on the national percentile scale, a change that many reformers would argue is feasible. This would not bring outcomes up to the level of the wealthier district, but it would improve VAM scores by half a month—which would be a very great improvement. Improving outcomes by the same half a month through background variables would require family income to be increased in the lower income district by \$25,000 $(= 0.5/0.02)$, a change on an order that seems unlikely to be brought about.

The numbers above have been chosen for expositional convenience in order to illustrate two points that are quite general. First, the best policy is the policy with the highest benefit/cost ratio.[8] Because the benefit/cost ratio has nothing to do with the variation of inputs in the data sample, there is no particular reason why the factor that explains most of the

[8] In this example, the benefit/cost ratio is the ratio "impact" factor to the marginal cost of another unit of that input. In principle, those emphasizing out-of-school factors might implicitly be saying that the cost of improved teaching quality or lowering class size (etc.) is very high, although that still would not make decompositions of variation relevant.

variation in outcomes should be a desirable target for policy intervention. Second, while there may well be extramural reforms that pass the benefit/cost test, we know that many background variables that explain much of educational outcomes, parental education for example, are essentially immutable over relevant policy horizons.[9,10] The claim that background variables explain a greater fraction of outcome variation than can be explained by policy variables is often true, so it is especially important to understand that the claim is true-but-irrelevant.

To drive the point home, note that successful policy can mask its own success if measured by decomposition of explanatory power. Carrying on with the illustration, suppose that in the low performing district we were able to carry out the 10 point improvement in teacher quality called for above. This improves the lot of students in the low performing district by eliminating the difference due to the initial difference in teacher quality. Since there would no longer be a difference in teacher quality, all remaining differences in student outcomes would be accounted for by the unchanged difference in family income. In terms of explanatory power, the effect of the *successful* policy reform would be to increase the fraction of the outcomes difference explained by family income from two-thirds to 100 percent!

[9] More is known about the marginal product of various interventions than about the marginal cost. A number of authors estimate the economic return to increased teacher effectiveness for example, but less is known about the cost of increasing teacher effectiveness. There are relatively few cost/benefit analyses comparing in-school versus out-of-school interventions. Levine and Zimmerman (2010) provide some examples. Mayer(1997) offers estimates of the effect of changing family income on a number of child outcomes.

[10] There are some background interventions, early childhood intervention or community-based programs, that proponents would argue are cost effective. Since such interventions are rare (often experimental) they explain little of the variability of current educational outcomes. As with in-school interventions, the fair test of such experiments is their effectiveness.

3. Empirical examples

While the previous section gave a hypothetical illustration, that illustration is typical of many research findings. Goldhaber et. al. (2010) offers a particularly clear example. Using data from Washington State, Goldhaber and coauthors decompose the variance of student mathematics achievement into student background characteristics and within-school teacher effects, as well as a component of between school differences that may be due to teacher quality or may be due to other factors, and a large unexplained component. The decomposition is shown in Figure 1.[11,12]

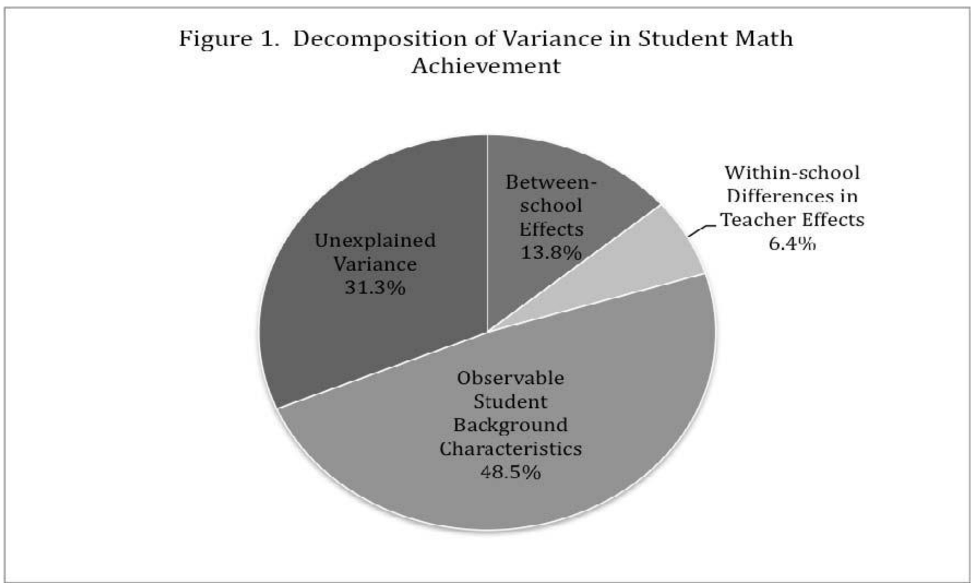

Figure 1

[11] Reproduced with permission from Goldhaber et. al. (2010).

[12] As a practical matter researchers generally make attributions of explanatory power using a "variance decomposition," as is the case in Figure 1. Technical issues about variance decomposition are discussed in the next section.

Note that half the variance in student achievement is attributed to background characteristics and a third of the variance is unexplained by observables.[13] The fraction of variance clearly attributable to teacher effects is quite small, only 6.4%. In fact, if we wanted to "understand" the causes of differential student achievement in mathematics we might well regard the role of teachers as being negligible.

If we are interested in policy intervention, the role of teachers is huge because the marginal impact of having a better teacher is huge. The authors present the marginal impact estimates in easily understood units. The difference between a median teacher and a teacher ranked at the 84th percentile (one standard deviation of the teacher quality distribution) is approximately 2.6 months of learning (18 percent of a standard deviation of student achievement). As is true in most studies of student achievement, direct information on the cost of changing teacher quality is not available. However using related (but not identical) data, Goldhaber and Theobald (2011) simulate student outcomes due to teacher quality changes if Washington State were to base teacher layoffs on VAM measures of effectiveness rather than teacher seniority. They find that this policy switch, which may have considerable political costs but which most likely has little financial cost, would improve student outcomes on the order of 2.5 to 3.5 months of student learning. So this is precisely an example where teacher inputs explains very little of student outcomes but where policy intervention in teacher inputs is potentially very promising.

[13] In principle, there is not a unique way to decompose the variance; a topic discussed in the next section. In this particular data set, the way the variance is decomposed makes relatively little difference (source: Dan Goldhaber, personal communication.)

As a second example, consider Krueger's (1999) study of class size effects in the Tennessee STAR experiment. In STAR, students were randomly assigned to small classes (15 students) or regular size classes (22 students). (Additionally, some regular size classes had teacher aides.) Because of the random assignment, the chance of being assigned to a small class or a large class was essentially uncorrelated with background variables such as socioeconomic status and overall school quality.

Note that the STAR experiment used a relatively large class size difference, where large classes had half again as many students as small classes. It is useful to distinguish which measures are affected by having a large variation in class size. First, if the effect of class size is linear, i.e. the marginal impact is roughly constant, then the measured effect does *not* depend on the size of the experimental variation. This experimental design does help statistically identify the effect of changing class size, leading to more precise estimates.[14] Because the experimental variation in class size is larger than the usual within-district variation "in the field," we might expect to see more explanatory power attributed to class size reduction in the experiment than we would usually see in observational data. In other words, the variation in class size in a particular set of comparisons does affect the attribution of variation to different explanatory factors, but doesn't affect the size of the policy effectiveness.

Table 1 provides an excerpt from Krueger (1999).[15] Reducing class size has a large—five percentile point—effect on student achievement.[16] Krueger reports that the five points is

[14] In other words, the experimental design would not change the estimated coefficient size, but should lead to lower standard errors.

[15] The source is the first panel of Table V in Krueger (1999). Coefficients are OLS coefficients with robust standard errors in parentheses.

equivalent to about two-thirds of the white-black test gap. Decreasing class size has a direct and more or less easily measured effect on raising costs, since half again as many teachers would be needed, plus more classrooms, etc. Krueger's rough estimate is that the net present value of the cost of smaller class size "is in the same ballpark" as the economic benefit.

Explanatory variable	Effect on average percentile of Stanford Achievement Test, Kindergarten students		
Small class	4.82 (2.19)	5.37 (1.26)	5.36 (1.21)
Regular size class with aide	0.12 (2.23)	0.29 (1.13)	0.53 (1.09)
School fixed effects	No	Yes	Yes
White/Asian	–	–	8.35 (1.35)
Female	–	–	4.48 (0.63)
Free lunch	–	–	−13.15 (0.77)
R^2	0.01	0.25	0.31

Table 1

Turning to the question of explanation of variance, we see from the R^2 row that class size explains only a trivial portion, one percent, of test score variation. In contrast in column 3, student characteristics plus unidentified, idiosyncratic school characteristics add another 30 percent to the fraction of variance explained. In other words, background variables out-explain class intervention by 30-to-1. Nonetheless, because the impact of class size reduction is estimated to be large, Krueger's calculations suggest that class size reduction is arguably cost effective. The fact that background variables explain more of the variance isn't relevant.

[16] On the substantive issue of the effect of class size reductions, note that the STAR results are on the high end of effect sizes found from other evidence. See Whitehurst and Chingos (2011).

4. The Math Behind the Argument

A mathematical exposition is in some ways more straightforward than the verbal version above. First, suppose that there is a single student outcome, S, and that the single measureable determinant of those outcomes is a policy variable, such as teacher input or class size, P. We can write

$$\textit{student outcome} = \textit{marginal policy effect} \times \textit{policy input} + \textit{unexplained part}$$

or, in symbols using e for the unexplained part and β for the value of the marginal policy effect as estimated by a regression,

$$S = \beta P + e \tag{1}$$

If β is large relative to the cost of adjusting P, then improved teacher quality (etc.) is sound policy.

When we want to partition student outcomes into a part attributable to policy and a part that isn't explained, statisticians usually ask how much of the variance in S is attributable to each factor. This variance decomposition depends on the variance of policy in the data sample as well as the value of β. The part of the variance explained by the policy input is $\beta^2 \times \mathrm{var}(P)$, the unexplained variance is simply $\mathrm{var}(e)$, and the total variance in outcomes is $\mathrm{var}(S) = \beta^2 \times \mathrm{var}(P) + \mathrm{var}(e)$.[17] The fraction of the total variance explained by policy is the R^2,

[17] In a least squares regression the variance decomposes neatly into an explained part and an explained part because the least squares residuals are orthogonal to the regressors. So it is always true that $\mathrm{cov}(P, e) = 0$. This property is not necessarily true when other estimation techniques are used.

$$R^2 = \frac{\beta^2 \times \mathrm{var}(P)}{\beta^2 \times \mathrm{var}(P) + \mathrm{var}(e)} \tag{2}$$

What is true is that $R^2 = 0 \Leftrightarrow \beta = 0$. Past that, that marginal impact can be large while the explanatory power is small if unexplained factors are large (or vice versa). This is precisely the case in the first column of Table 1 above, where $\beta \approx 5$ is large even though $R^2 = 0.01$.

Equation (2) is useful for one more thought experiment that shows successful implementation of policy does not necessarily increase the explanatory power of policy compared to other factors . Suppose that we went ahead and implemented a policy raising teacher quality for every single student by the amount ΔP. For every student, outcomes would improve by $\beta \Delta P$. Now suppose researchers went out and drew a new data sample. Would teacher input look any more important in the new sample than it had originally—where "important" is thought of in the variance decomposition sense? No, while the entire distribution of P has shifted to the right, $\mathrm{var}(P)$ is unchanged. The variance decomposition would be the same.

More commonly, the relevant variance comparison is between policy variables and background variables. For expositional purposes, suppose there is one policy variable and that background is summarized by the single variable B. We can write

$$\begin{aligned} &\textit{student outcome} \\ &= \textit{marginal policy effect} \times \textit{policy input} \\ &+ \textit{marginal background effect} \times \textit{background} \\ &+ \textit{unexplained part} \end{aligned}$$

or, in symbols

$$S = \beta_P P + \beta_B B + e \tag{3}$$

Assume for convenience that policy and background inputs are uncorrelated and that student outcomes are fully explained so that we can drop e. The partitioning would be given by

$$\mathrm{var}(S) = \beta_P^2 \,\mathrm{var}(P) + \beta_B^2 \,\mathrm{var}(B) \tag{4}$$

When we partition observed student outcomes ($\mathrm{var}(S)$) the contribution of each factor ($\beta_P^2 \,\mathrm{var}(P)$ and $\beta_B^2 \,\mathrm{var}(B)$, respectively) depends on two elements: how much that factor affects outcomes and how much that factor varies. We know that the bulk of student outcomes is explained by background (i.e. family, community, family, genetic, and other non-school) factors. This means that the marginal effect of background, β_B, is large, and/or that background inputs vary, $\mathrm{var}(B)$, a great deal. Neither speaks to the question of whether the marginal effect of teachers or class size, etc., β_P, is large. It's β_P that matters for policy evaluation.

As pointed out in the introduction, policy evaluation and attribution of explanatory power are not entirely unrelated. Both depend on the relative size of the marginal effects. But policy evaluation also depends on a sense of the cost the proposed intervention, where attribution of explanatory power does not. Attribution of explanatory power also depends on how much each of the inputs has varied across a particular data set, which has nothing to do with choosing sound policy. As a practical matter, in the language of equation (4) β_P and $\mathrm{var}(B)$ are each large, leading to the perceived—but incorrect—view that policy interventions can't work.

Equation (4) depended on the special assumptions that policy and background are uncorrelated. Unless the two are uncorrelated, which they are not in general, there is no way to cleanly decompose the outcome variance into a part due to policy and a part due to background.[18] This is true even though the model has no causal interaction effect between policy and background. In both empirical examples in the previous section, the correlation was approximately zero. This was not coincidental. With random assignment we expect the policy variable to be approximately uncorrelated with background variables. And while background variables in the Goldhaber et. al. paper might well be correlated with school attended, it is not surprising that they are not so much correlated with within-school teacher effects.

5. Summary

Understanding explanatory power—decomposition of variance—in a particular data sample is at most loosely related to evaluation of policy impacts. While this is a general point about statistical models, it is particularly important to education reformers. Because most of the variables that explain student outcomes are immune to reform, it is critical to understand the disconnect between variance decomposition and policy effectiveness.

At one level, the distinction between explanatory power and policy effectiveness is a principle that can be understood by thinking through evaluation models carefully. The empirical examples given here illustrate that the distinction is much more than hypothetical. In education research, it is entirely likely to find *effective* policies that explain only a small fraction of the observed variation in student outcomes.

[18] Hanushek and Kain (1972), p 124ff, make precisely this point with regard to the Coleman report.

References

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

Ehrenberg, Ronald G. and Dominic J. Brewer, "Did Teachers' Ability and Race Matter in the 1960s? *Coleman* Revisited," *Economics of Education Review,* Vol. 14, No. 1, pp. 1-21, 1995.

Goldhaber, Dan, "The Mystery of Good Teaching," *Education Next,* Spring 2002.

________, Stephanie Liddle, Roddy Theobald, and Joe Walch, "Teacher Effectiveness and the Achievement of Washington's Students in Mathematics," Center for Education Data & Research, University of Washington, Working Paper 2010-06.

________ and Roddy Theobald, "Managing The Teacher Workforce in Austere Times: The Implications of Teacher Layoffs," Center for Education Data & Research, University of Washington, Working Paper 2011-12.

Hanushek, Eric and John Kain, "On the Value of *Equality of Educational Opportunity* as a Guide to Public Policy," in *On Equality of Educational Opportunity,* Frederick Mosteller and Daniel Moynihan, Eds, Random House, New York, 1972.

Krueger, Alan, "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics,* Vol. 114, No. 2, May 1999.

Levine, Phillip B. and David J. Zimmerman, eds., Targeting Investments in Children, University of Chicago Press, 2010.

Mayer, Susan E., *What Money Can't Buy: Family Income and Children's Life Chances,* Harvard University Press, Cambridge, 1997.

Manski, Charles F., "Genes, Eyeglasses, and Social policy," *Journal of Economic Perspectives, Vol. 25, Number 4, Fall 2011.*

Whitehurst, Grover J. "Russ" and Matthew M. Chingos, "Class Size: What Research Says and What it Means for State Policy," Brown Center on Education Policy at BROOKINGS, May 2011.

______, *Statistical Abstract of the United States, 2011.*

Appendix

Equation (4) depended on the special assumptions that policy and background are uncorrelated and that there is no unexplained part. The more general result is that

$$\mathrm{var}(S) = \beta_P^2 \,\mathrm{var}(P) + \beta_B^2 \,\mathrm{var}(B) + 2 \times \beta_P \times \beta_B \times \mathrm{cov}(P, B) + \mathrm{var}(e) \tag{5}$$

Unless $\mathrm{cov}(P, B) \approx 0$ there is not a unique decomposition of the outcome variance into a part due to policy and a part due to background. Generally, any decomposition that does not offer a separate covariance-related term reflects a decision by the researcher to attribute the part of the overall variance due to the covariance term to one factor or another. For example, in equation (5)**Error! Reference source not found.** the variance explained by teacher input could be presented as either $\beta_P^2 \,\mathrm{var}(P)$ or $\beta_P^2 \,\mathrm{var}(P) + 2\beta_P\beta_B \,\mathrm{cov}(P, B)$. Note that the contribution of the covariance term may well be negative.

Alternatively, the researcher may give a non-causal variance decomposition by deciding on an ordering that maximizes the variance attributed to the first ordered regressor, then decomposes the remaining variance so as to maximize the fraction attributed to the second ordered regressor, and so on. Suppose we are decomposing the variance of S based on k explanatory variables that the researcher chooses to order as $X_1, X_2, \ldots X_k$. Step 1 is to regress the S on X_1, taking the explained variance from the regression as the variance attributed to X_1. Step 2 regresses the residuals from step 1 on a variable constructed as the residuals from regressing X_2 on X_1. The explained variance from step 2 is then attributed to X_2. In general, step j regresses the residuals from step $j - 1$ on a variable constructed as the residuals from regressing X_j on variables $X_1 \ldots X_{j-1}$. The unexplained variance is then the unexplained variance from the k^{th} regression.

By following this procedure, the researcher loads as much explanation as possible on the first variable. As much of the remaining variance as is possible is then loaded unto the second variable, and so forth. To the extent that different orderings give different variance decompositions, the differences reflect choices made by the researcher rather than causal factors in the data.

In the special case where the explanatory variables are mutually uncorrelated, this procedure gives the same variance decomposition as would be calculated from a multiple regression, and the variable ordering is irrelevant. In the general case however, the ordering chosen by the researcher does affect the results and the decomposition is not causal. The first regression attributes as much of the explanation due to confounding variables as is possible to the first variable in order. For example, if P is ordered first then a researcher might attribute a large variance explanation to P even if $\beta_P = 0$.