DOCUMENT RESUME

ED 073 009 SO 005 314

AUTHOR Michelson, Stephan
TITLE The Existentialist Reality of Educational Production Functions (Their Existence Has Preceded Their Essence).
INSTITUTION Harvard Univ., Cambridge, Mass. Center for Educational Policy Research.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
PUB DATE 1 Aug 70
CONTRACT OEC-0-70-5108
NOTE 110p.

EDRS PRICE MF-$0.65 HC-$6.58
DESCRIPTORS Cost Effectiveness; *Educational Accountability; *Educational Economics; Educational Finance; *Input Output Analysis; *Mathematical Models; Statistical Analysis; Systems Analysis
IDENTIFIERS *Educational Production Function

ABSTRACT

This publication is a discussion of the concept of an educational production function, a mathematical formulation of the relationship between inputs and outputs in education. A description of how a production function can direct decision making toward economic efficiency precedes a theoretical discussion of the nature of actual estimation. Several statistical estimations which demonstrate the empirical counterpart of the theoretical discussions are made using one sample of data. The author concludes the paper with a summarization of the work. (Author/SHM)

$$Y = a + b_1'X_1 + b_2'X_2 + c_1X_1X_2 \ldots$$

has partial derivatives

$$(b_1' + c_1X_2)$$

$$(b_2' + c_1X_1) \ .$$

Here the response of Y to increments of X_1 depends on how much X_2 is present.

Other complications arise when other forms are tested. Non-linear relationships can be approximated with higher order polynomials, such as

$$Y = a + b_1X_1 + c_1X_1^2 + b_2X_2 \ .$$

In this case, $\frac{\partial Y}{\partial X_1} = b_1 + c_1X$; i.e., the response of Y to X_1 depends on how much X_1 there is to begin with. Typically, the exponent c_1 in such estimates is negative but small. The result is that for small values of X_1, b_1 dominates, and Y responds positively to increases in X_1. As X_1 increases, the effect of added X_1 diminishes.[5]

The mathematical form of the production equation, then, is crucial for determining its partial derivatives. These, in turn, give the information we are seeking: an estimate of the change in output given a specific input change. In Part II, therefore, this discussion on functional forms is continued.

[5]Examples with this property will appear in Part IV.

INTRODUCTION

Since the entry of sociologists and economists into educational research--which dates from the late 1950's--the concept of an "educational production function" has gained ever increasing mention. Yet nowhere is there a clear discussion of what such a function would do if there were one, what its characteristics would be, and how one would go about finding (estimating) one.[1] In the first Part of this paper, I will discuss the reasons why one might want a mathematical formulation of the relationship between inputs and outputs in education. In Part II I will briefly discuss some of the properties of such a formulation. That discussion will be the most meager of all, thus failing to fill a gap that has existed from the beginning of the attempts to fit these functions. The reasons for this continued failure will be discussed.

Third, I will present some theoretical discussion about the nature of actual estimation. Much of this discussion is a straightforward translation of traditional production theory into terms directly relevant to educational production. Some, however, is a novel attempt to deal with a number of problems which have been skirted by previous researchers. These include the problem of identifying frontier ("best practice") institutions, the problem of multiple outputs, and the problem of simultaneous determination of interrelated outputs.

The fourth Part will present some statistical estimations demonstrating the empirical counterpart of the theoretical discussions of Parts II and III.

---

[1] Kershaw and McKean do sketch clearly uses for such a function in (12), and Bowles discusses estimation in (1) and (2). These papers do not discuss many of the issues covered below, however.

However, no useful educational production function will be presented here. Indeed, it is my contention that there never has been even a reasonable mathematical formulation, not to mention an adequate estimate of such a function. Nor will there be for some time to come. The reasons for this are found in Parts I and III below, although the crucial element may be determining a functional form as discussed in Part II. A fifth section will summarize the work of the paper. The reader might be advised to skip to Part V first, and then return to Part I.

I

THE USES AND ABUSES OF EDUCATIONAL PRODUCTION FUNCTIONS

In general, a functional relationship between inputs and outputs in production is expressed thus:

$$Y = f(X_1, X_2, \ldots X_n).$$

Y is a measurable output or index of outputs;[2] the X_i are inputs into the process. Since production adds value to raw materials, the inputs are the factors of production (labor and capital, in quantity and quality), and the output is the value added by these inputs. No account is taken of the initial value of the materials in this formulation. The initial value is expressed in the same units as the output value, and if the initial value is the same for all observed production units, then it makes no difference if one thinks of Y as $Y_t - Y_o$ (output value at the end of the process less output value at the beginning), or as Y_t (output value at the end of the process). The difference is a constant term in the expression $f(\ldots X_i \ldots)$.

Since the raw materials in education are pupils whose initial values (in output terms) differ, some account must be taken of these differences in education functions. However, this is an estimation problem, which poses no difficulty in the conceptualization of the value added function. The educational production function, then, though in estimation requires adjustment for critical values, in presentation should appear as value added being a

[2] The conditions under which an index of outputs may be formulated is discussed in Part III of this paper.

function of production inputs only.

The X_i are elements of the production process during the time period being considered. Since a student in non-boarding schools spends most of his time not at school, then any production period exceeding one school day must account for value added outside the school. As an example, consider the output Y to be the increment to vocabulary between the 9th and 12th grades. The conceptually correct educational production function would adjust inputs for differences among pupils in vocabulary at the 9th grade, and consider items outside the school--say literacy of parents--as an input to the production process during the high school years. Thus variables describing the "social class" of pupils serve two conceptually separate functions: They might correct for differences on entry to the production period, or for output production during the production period, but not at school. This distinction is crucial. To the extent that output differences are due to differences in production during the period under consideration, then programs to get more resources to children who have few outside-school resources, preferably during the times the other children are getting the outside-school resources, would have an obviously good chance of success. To the extent that differences in final output are due to differences in initial value of the output measure, a different production process entirely may be called for.[3] And we know little about this process.

[3] This simple point seems little appreciated. It could be more cost-effective, for example, for a public school charged with getting children to some minimum reading level, to launch pre-school programs than to place remedial teachers in elementary schools. (In areas with great mobility, of course, it might not be cost effective for any one school to do this, since it will still have children without pre-school experience in their elementary schools. For the system, however, this policy would still be most efficient.)

The next step in specifying the production function is to indicate the signs of its first partial derivatives, $\frac{\partial Y}{\partial X_i}$:

$$Y = f(\overset{+}{X_1}, \overset{+}{X_2}, \overset{-}{X_3} \ldots) .$$

A partial derivative indicates the rate of change of Y when X_i is incremented by a small amount, other variables staying the same. A negative sign indicates that an increase in only X_i produces a loss in Y. If many outputs are to be investigated, then it would not be surprising to find negative derivatives for some variables with respect to some outputs. Thus increasing the average verbal facility of teachers might produce a reduction in manual skills; increasing the brawn of assistant principals might reduce some kinds of creative expression, etc. Yet, of course, such losses might be an acceptable "price" to pay for gains in other outputs.

The last important feature of the production function is actual estimates of the partial derivatives.[4] Thus we have to know the functional form of the input-output relationships. For example, a linear function

$$Y = a + b_1X_1 + b_2X_2 + \ldots$$

has partial derivatives b_1, b_2, etc. But a linear function with multiplicative interaction terms:

---

[4] It is easier to discuss changes in terms of derivatives than in terms of differences between initial and final Y values. Where the derivatives themselves are not constant, then an estimate of the change in Y should proceed by inserting initial and final values of the X_i into the equation.

$$Y = a + b_1'X_1 + b_2'X_2 + c_1X_1X_2 \ldots$$

has partial derivatives

$$(b_1' + c_1X_2)$$

$$(b_2' + c_1X_1) \; .$$

Here the response of Y to increments of X_1 depends on how much X_2 is present.

Other complications arise when other forms are tested. Non-linear relationships can be approximated with higher order polynomials, such as

$$Y = a + b_1X_1 + c_1X_1^2 + b_2X_2 \; .$$

In this case, $\frac{\partial Y}{\partial X_1} = b_1 + c_1X$; i.e., the response of Y to X_1 depends on how much X_1 there is to begin with. Typically, the exponent c_1 in such estimates is negative but small. The result is that for small values of X_1, b_1 dominates, and Y responds positively to increases in X_1. As X_1 increases, the effect of added X_1 diminishes.[5]

The mathematical form of the production equation, then, is crucial for determining its partial derivatives. These, in turn, give the information we are seeking: an estimate of the change in output given a specific input change. In Part II, therefore, this discussion on functional forms is continued.

---

[5]Examples with this property will appear in Part IV.

Production Alternatives

Suppose we have estimated a production function for a school organization. That is, we have defined an output, and have estimated partial derivatives of that output with respect to the inputs affecting that output during the time of production. We are now considering changing X_1, or, alternatively, changing X_2. That is, we are considering some changes in inputs, say a curriculum change on the one hand, a class size change on the other. If the alternatives are constructed so as to cost the same, then one might simply choose that which is the most "effective." If, in fact, only one output (or an output index) is to be considered, this is precisely what should happen. So in the simplest example, one wants to know the derivatives of an educational production function to choose the most effective among equal cost alternatives.

$dY = \frac{\partial Y}{\partial X_1} dX_1$ is compared with $dY = \frac{\partial Y}{\partial X_2} dX_2$, where dX_1 and dX_2 are the (small) equal cost changes just described. Whichever calculation is larger indicates the estimate of the better--more productive--option.

Seldom are the alternatives before us so clear cut. In the first place, there are many outputs, and the "preferred" equal cost change might be different depending on which output is considered. In theory, increments to these outputs are weighted by a set of preferences (though whose preferences is not clear). If each output is Y_j, and its index weight is W_j, then

$$I_k = \sum_j W_{jk} Y_{jk}$$

can be calculated for every equal cost proposal, k, and the largest I_k can be chosen. Such an abstract, precise procedure, of course, does not describe reality. But it does demonstrate a real problem: suppose different schools (or districts) prefer different outputs. What does this do to our ability to estimate production functions? This question will reappear in Part III.

I will continue to assume that the problem of judging multiple outputs is taken care of.[6] Still assuming that we have a well-defined function, with quantitative estimates of its partial derivatives, we can investigate the more general question of choice of program option. Consider two options, P_d and P_a, which do not cost the same. P_d we will call the option of hiring teachers with more advanced degrees than the present average. P_a is the alternative option of adding teacher aides to less "qualified" teachers. Consider the ratio of outputs from the two options:

$$\frac{Y_d}{Y_a}$$

where Y represents an output. Consider the case in which the ratio of outputs is greater than the ratio of costs, $\frac{C_d}{C_a}$. In the first example discussed above,

---

[6] I will note that this subject deserves much more attention than is ordinarily paid it. To what extent are high schools unable to choose different outputs, because the prime goal of the parents is admission to college and college entrance is based on a very limited set of skills? Indeed, Katzman (11) considers admission to Boston Latin School (which, with Boston English, comprises essentially a "college track") as one output of Boston elementary schools. Schools wanting to maximize that figure--entrance being determined by a competitive examination--are extremely limited in output choices at very early grades. Though it is true that the more limited the output set, the more use can be made of well estimated production functions, it is also true that this structure seriously constrains the freedom of individual parents and children to develop in alternative ways.

$$\frac{C_d}{C_a} = 1, \text{ and when } \frac{Y_d}{Y_a} > 1,$$

we said "buy P_d." We now have the more general case. When the ratio of the outputs is greater than the ratio of the costs, then more value per dollar is achieved by buying the project of the numerator. When the inequality is the other way around, the denominator is the better buy.

Suppose, then, that aides are a more efficient, but more expensive proposal.[7] Remembering that $Y_a = \frac{\partial Y}{\partial X_a} dX_a$, where X_a is the input "aides," and $Y_d = \frac{\partial Y}{\partial X_d} dX_d$ where X_d is "degrees," we can estimate a loss in output if we reduce the degree level of teachers (by replacing retiring teachers with new ones less "qualified" than the average) enough to save the extra money. Now we can consider equal cost proposals in which, where $\Delta Y_{d/a}$ indicates the change in Y from reduction in d given a, the aides would produce an additional

ΔY_a	$-$	$\Delta Y_{d/a}$
(gain in output due to aides)	less	(loss in output due to less qualified teachers)

and hiring new teachers with higher degrees would produce Y_d. If $(\Delta Y_a - \Delta Y_{d/a}) > \Delta Y_d$, then the aides are a better solution, and the in-

[7] One might be tempted to want to adjust the quantity of the aide program down to the cost of the advanced degrees program. But it could well be the case that a program has a minimum feasible size, or that the costs considered here are for a range of sizes, all higher than that equal to the cost of higher degrees. For a small project, the cost of aides might be higher, and blind obedience to the ratio rule above would lead to preferring more "qualified" teachers. This would be a poor solution, as we shall see. Thus one should compare such projects at efficient sizes first, which is why we need the general formulation which does not assume equal costs.

struction staff should be adjusted downwards in nominal qualifications.

Economic vs. Technical Efficiency

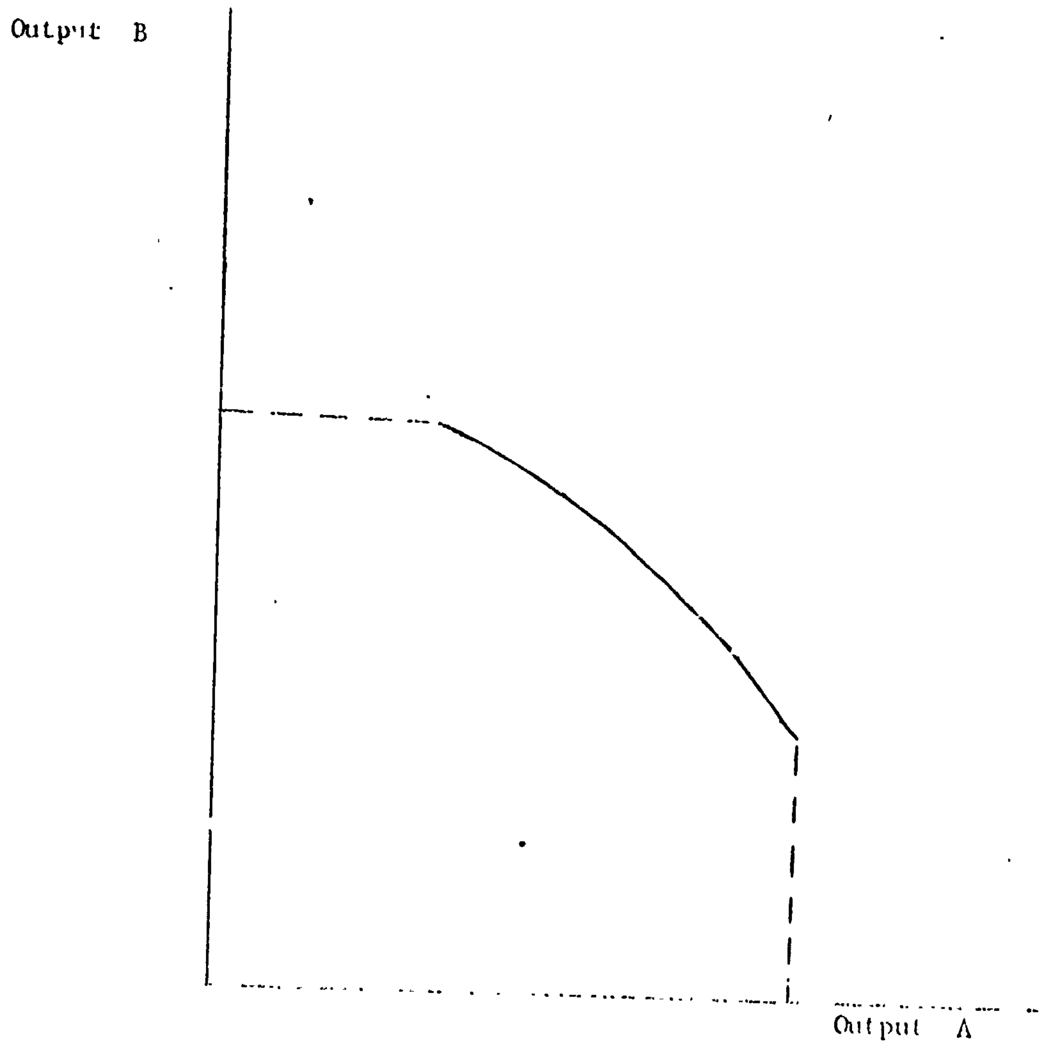
The problems of choice discussed above have implicitly assumed that all options were utilized to produce the maximum output, or outputs in some determined proportions. In other words, the question asked was an economic one: what is the best mix of well used resources to produce given outputs. We will have to probe into the possibility that resources are not always well used. A school committee, of course, has to balance the gain from greater efficiency against the cost of better management, in determining its economically efficient production mix.

Maximum output for a given set of resources is technical efficiency. Two schools with equivalent resources could produce different outputs with technical efficiency. If we consider two outputs, Y_1 and Y_2, then for a given set of resources, a production frontier defines the locus of efficient output combinations. For a different set of resources, a different frontier is defined. In Figure I-1, one such frontier is pictured. It is somewhat different from an ordinary production frontier in that it assumes that for low levels of either output, there is no trade-off of outputs. That is, improvement in one output does not necessitate loss in the other. Where both are beyond minimum levels, the trade-off does occur. More resources devoted to one necessitates a reduction in the other output.

Technically efficient production simply means production on the production frontier. Technical inefficiency, then, is a condition where more of

FIGURE I-1

A Production Frontier



one output could be gained without loss of any other output.[8] A frontier could be convex to the axes. This would mean that as additional equal size sacrifices are made in one output, larger and larger gains can be made in the other. Although this is not impossible, it leads only to specialization in very few outputs. I do not think this kind of trade-off exists in public schools, although it may be the case in graduate schools.

Technical efficiency, then, means maximum (frontier) production, given an input structure. Economic efficiency means choosing the best input structure, given input-output relations and prices of inputs. Production functions are economic tools: they are used to purchase the correct inputs, assuming they will be managed correctly. It should be obvious that if some inputs are typically mismanaged, then an *ex post* production function analysis may advocate purchase of other inputs. Yet the truly efficient solution might be to use the current inputs differently, depending on the cost of better management. This could be the case, for example, where mathematics and science teachers are scarce, and other teachers are used in their place. There is no reason to believe that production of mathematics output is technically efficient, given the resources. We would "find" that, to produce math skills, we must go into the market and "buy" prepared math teachers. But it could be true that another combination (say, more materials with inferior teachers) is more cost-effective. Since principals "cover" the math class, but do not add non-human resources to substitute for inferior human resources, it is likely that mathematics production is technically inefficient given the resources of the

[8] Production on the dashed lines, then, is inefficient, and these lines are not actually part of the frontier.

schools. Not observing efficient production, we cannot estimate its effect.

It is then asked of the production function analyst to recommend production changes. That is, he is asked to make <u>technical</u> recommendations about substitution of resources. I want to stress that this is not the function of such analyses, and by and large there is little to say for a recommendation proceeding from such an approach.[9] Where there is technically inefficient production--and more importantly, where some outputs (say, math) are produced (in public schools) with less technical efficiency than others (say, reading)--the real management alternatives are much greater than the production function analysis implies.

For this reason, one ought to observe only technically efficient schools. One can then determine both what a cost-effective input mix would be like with good management, and what returns could be expected from better management (of inefficient schools) itself. In production theory, this is called observation of "best practice" firms.[10] In Part III I will present a method of isolating best practice schools, and in Part IV I will attempt to actually pick some from my data. The exercise is merely illustrative of the problem, and not a good solution to it.

---

[9] As we shall see in Part IV, there is a role to be played by a well specified production function in promoting technical efficiency. The problem, again, is knowing how well-managed the schools are which provide the data for the production function estimate.

[10] See Salter (20).

Production Function Estimation

I have shown so far that production functions could be a management (or policy) tool to promote more precise definition of alternatives, and more economic efficiency in production. But estimation of such a function must follow detailed technical knowledge of input-output relationships which does not yet exist. A complete technical description is not necessary: we do not estimate production functions from blueprints, but from data. Still, the production function must appear reasonable to the experienced educator, or there must be a good explanation of why it does not appear so. A short discussion of properties of functions appears in Part II, so this discussion about reasonableness can begin.

Empirical estimates are not without value absent the knowledge of forms (and measurement, and a host of other technical items) one would like to have. I am merely distinguishing regression estimates as they have appeared in the literature from production function estimates, which have never appeared. A linear additive regression on all observations (not just best practice ones), not adjusted for initial values of the output measure, may give a good estimate of average relationships. Thus the typical equation:

$$Y = a + \ldots + b_j X_j + \ldots + c_j Z_j + \ldots$$

where X_j are school inputs, and

Z_j are pupil background variables

gives an average relationship between the output, Y, and the inputs X_j and Z_j.[11]

[11] For a non-technical summary of seventeen regression studies, see Guthrie, et al. (8), Chapter 5. These studies agree that there is some average relationship between school resources and cognitive skills.

How much of Y is produced during school years is seldom distinguished from Y produced before school. Nor, even given this distinction, has any attempt been made to distinguish between the effect of home resources during non-school periods, vs. the effect of home resources as producing an interaction during school.[12] The knowledge that home resources explain more variance than school resources was predicted well before any empirical results were known.[13] What is important is to discover the production relationships which do occur in schools as presently constituted.

All of these important research topics can be investigated with current theoretical and statistical tools. Determining these average relationships would be vital for enlightened educational policy. I only wish to stress that they would not help in making detailed production decisions, which involve marginal estimates. Congress and the Office of Education want to know what area, on the average, deserves support (say, between summer care, pre-school, after-school programs; teacher training in skills or in behavior, etc.). School systems and particularly school principals want to know what to do on the margin. Congress and the Office of Education need to deal with students in general, but principals deal with very particular students. A general

---

[12] For a stimulating attempt to distinguish between school year and summer learning, see Hayes and Grether (9). I know of no attempt to separate, within the school year, in-school from out-of-school effects. One might begin to do so with a sample of children who had been deprived of school for some period of time and a comparable sample in boarding school.

[13] "It is probable that all the other variables taken together will have only a fraction of the impact on achievement that either IQ or the intellectual atmosphere of the pupil's home will have." Kershaw and McKean (12), p. 27.

solution is bound to be incorrect in many places. The aim of production function analysis is to be able to determine those places before the damage is done (or the opportunity wasted). Average regression estimates are important tools, but they don't help to answer that particular question.

II

FUNCTIONAL FORMS IN FACT AND FICTION

The need for discussing the mathematical structure of an educational production function has already been demonstrated. To estimate changes in output which are expected to occur with input changes, one needs to know the actual equation expressing Y as a function of the inputs. For small changes in inputs, we can approximate estimates of output changes from the derivatives of the output with respect to the inputs. There are a number of other properties of functions which are important. These properties exist implicitly in any functional form. Hence, whether a researcher wishes to discuss these properties or not, his "model" of input-output relationships in education contains estimates of them, which may or may not be reasonable. In some forms, in fact, the values of these properties are independent of the values of the parameters (the values discovered empirically by fitting the functional form to the data). Thus important properties may be unknowingly specified *a priori*. I will discuss the following properties: derivatives, elasticity of output, elasticity of substitution.[1]

Derivatives

The first partial derivative of Y with respect to X_i, $\frac{\partial Y}{\partial X_i}$, denotes the instantaneous rate of change of Y with respect to changes in X_i, where all $X_j \neq X_i$ are held constant. As above, with small changes in X_i, one can ignore the fact that this rate of change may itself change, and estimate a discrete

---

[1]There are many other areas deserving of investigation: what the equation says about returns to scale, for example.

change in Y, dY, by

$$dY = \frac{\partial Y}{\partial X_i} dX_i .$$

When Y is a simple (no interactions) linear function of the X's, this formula is exact. The coefficient of X_i, b_i, is its derivative. Thus with linear estimations, researchers give estimates of changes in Y from the above formula without having to calculate Y_t (after change) and Y_o (before change). And this presentation is valid no matter what the values of the X_{oi} and Y_o -- inputs and output before hypothetical adjustments.

The independence of the effect on output of an input adjustment from the initial value of that input is an important property of an educational production function. We have already seen two functions of which this is true:

(1) $Y = a + b_1X_1 + b_2Y_2 + \ldots$

(2) $Y = a + b_1X_1 + b_2Y_2 + c_1X_1X_2 + \ldots$

In equation 2, changes in X_1 may affect Y differently depending on the values of X_2 etc., but not depending on its own initial value.[2] One might ask, is this a reasonable characteristic of an educational production function?

The answer is equivocal. It depends, largely, on the definition of

[2] The multiplicative interaction term shown here will be the only interaction discussed. It is not the only possible interaction, however: for example, it is symmetric: two high values have equally strong effect as two low values (though in opposite directions). This might not be the case. Interactions which work in limited ranges can be defined to test for this possibility.

inputs. Thus increasing teacher training by one semester may have diminishing effects, the higher the training level initially. It is unreasonable to assume that training has a linear relationship to output. However the "training" variable can be partitioned. Define two variables: the first was the value 1 if the teacher has credits beyond the bachelor's degree, the second is 1 if the teacher has credits beyond the Master's degree. Each variable is zero otherwise.[3] There are a number of good statistical reasons why non-linearities are better estimated by non-linear forms than by sets of binary coded variables, but this does not obviate the fact that sometimes one can have level dependent estimates within the confines of a simple linear equation. The crucial element here, then, is whether the definition of the variables is such that the equation defines only one or more than one derivative for a particular policy. A program of "training" can have two different estimated results, but a program of "credits beyond an M.A." can have only one. There is no reason to be critical of a linear additive form, *per se*, on the question of whether effects of increments to X_i should depend on the level of X_i.

Using a free definition of variables, and the interaction terms of Equation 2, we can then have an additive form which accounts for dependence of changes in Y on both its own level and the level of other variables. Since a number of other properties are not accounted for so simply, we should in-

[3]The complete set of variables is

	less than BA or just BA	BA + credits	MA + credits
var. 1	0	1	1
var. 2	0	0	1

Thus these two variables define three cases, and they are not linear dependent.

vestigate at least one other form:

$$(3)\quad Y = aX_1^{b_1}X_2^{b_2} \quad \ldots$$

An extension of this form is

$$(4)\quad Y = aX_1^{b_1}X_2^{b_2}e^{b_3X_3} \quad \ldots$$

Form 3 seems to impose a severe interaction restruction: when X_1 or X_2 is zero, there is no output. However, this form cannot be estimated in the first place where X_1 or X_2 is ever zero. Form 4 takes care of this problem by allowing X_3 to be zero, in which case $e^{b_3X_3} = 1$.[4] These equations are estimated by taking logarithms:

$$3')\quad \text{Log } Y = a + b_1(\log X_1) + b_2(\log X_2) + \ldots$$

$$4')\quad \text{Log } Y = a + b_1(\log X_1) + b_2(\log X_2) + b_3X_3 \quad \ldots$$

The derivatives are dependent on the values of all variables:[5]

$$\frac{\partial Y}{\partial X_1} = b_1aX_1^{b_1-1}X_2^{b_2} \quad \ldots$$

[4]This form is most useful when X_3 denotes time. Then b_3 is an estimate of (1 + r) where r is a compound interest rate. In an education production function where there is time series data, this term could usefully estimate the general educational effect of growing up in a literate society.

[5]Needless to say, this expression does not hold when X_2 is a function of X_1. But this whole discussion assumes that only Y is a function of the X_i: they are not functions of each other, or of Y. A brief discussion on estimation when X_i are not exogenous--i.e., do not take on values independently of the other variables--appears in Part III of this paper.

$$\frac{\partial Y}{\partial X_3} = b_3 a X_1^{b_1} X_2^{b_2} e^{b_3 X_3}$$

The change in Y when X_1 is changed is, then, a function of the original value of X_1, and of all other independent variables. Alternatively, it is assumed that each variable interacts multiplicatively with all other variables. This is fairly general, though one could complain that the way in which $\frac{\partial Y}{\partial X_1}$ is a function of Y is restrictive:

Suppose, for example, that a school increased reading scores by 10% by hiring more qualified teachers. The predicted increment to scores from now hiring more experienced teachers has necessarily increased, if the effect of experience was positive to begin with. That is, if the sign of the direct effect of two variables is positive, the sign of their interaction must be positive.[6] I do not wish to judge this property, merely to expose it.

Elasticity of Output

The elasticity of output with respect to some input variable X_i, which I will denote ϕ_i, is a straightforward extension of the first derivative. One might want to define a measure which is independent of the units in which X_i and Y are measured. Thus

---

[6]Bowles (1, footnote, p. 11) explains this as the requirement that "an increase in the quality of teachers be more effective on children of well educated parents than on the children of illiterate parents."

$$\phi = \frac{dY/Y}{dX_i/X_i} = \frac{dY}{dX_i} \cdot \frac{X_i}{Y}$$

This gives the proportionate change in Y when X_i changes by a small amount. For example, if $\phi_i = .3$, then a 10 percent increase in X_i leads to a 3 percent increase in Y. If $\phi_j = -.3$, then a 10 percent increase in X_j leads to a 3 percent decrease in Y. Just as it was true that the simple linear form defined a constant derivative, the simple multiplicative form (equation 3, above) defines a constant elasticity of output. The equation

$$Y = aX_1^{b_1}X_2^{b_2}$$

forces an estimate of the percentage increase in Y, given an increase in X_1, to be b_1 times that percentage increase in X_1. Where a constant output elasticity is assumed as a matter of production theory, this gives a convenient way of estimating it. Elasticities given for linear equations are usually calculated by setting the variables at their means. Though the presentation may mention "the elasticity," it really means an elasticity. One might want to take extreme values of his data, and calculate the range of elasticities implied. Of course, the value of Y used in this calculation depends on the values of other variables. Unless other important variables in the equation are negatively correlated with X_1, then the elasticity of output with respect to X_1 will be found to decrease as X_1 increases. This should be obvious: since increments of X_1 have a constant absolute effect on increments to Y, then (barring extreme negative correlations in the data) a linear function will estimate a small elasticity of output with respect to X_1 for a resource-rich school compared with a resource-poor school.

If this is a sensible property, then defining the elasticity of output as the same for resource-rich and resource-poor schools is not sensible. On the other hand, it may be that there are more interactions in rich schools, and the output elasticity actually *increases* (i.e., resource-rich schools are better able to utilize a small addition to resources than poor schools). If this is the case, the linear form could give even worse estimates of the increase in Y for schools at one extreme or the other than the multiplicative form.

The general conclusion about elasticity must be the same as about rates of change: For large policy purposes, an average estimate is good enough. It would not seem to matter materially whether one estimates elasticity from a linear function, with variables at their mean values, or from a constant elasticity function. For statements about educational *production*, however, and for an idea of what to expect in schools with input characteristics different from the mean, the difference does matter. Detailed studies of extreme schools would be necessary to determine which functional form best described the changes in output elasticity over ranges of inputs.

Elasticity of Substitution

The Marginal Rate of Technical Substitution (MRTS) between two inputs is the ratio of the respective first derivatives of the production function. Recalling that this function was, in general,

$$Y = f(X_1 \ldots X_n),$$

let us denote the partial of f with respect to variable X_i, as f_i. The marginal rate of substitution of inputs j for input i is

$$\frac{f_i}{f_j}$$

This is, as the name implies, the rate at which one can substitute two inputs without sacrifice of output.[7] This is an obviously important concept. For example, from equation 1 above, the simple linear form, the MRTS between i and j is

$$\frac{b_j}{b_i}.$$

Since the derivatives were independent of the levels of the variables, so is the marginal rate of technical substitution. No matter how much or how little X_i there is, the same amount of X_j can substitute for a unit of X_i, leaving output unchanged. Denoting the relative change in the MRTS as the ratio of inputs changes as the Elasticity of Substitution, σ, it is clear that this measure is infinite in the simple linear form, regardless of the parameters:[8]

---

[7]The condition is $dY = \frac{\partial Y}{\partial X_i} dX_i + \frac{\partial Y}{\partial X_j} dX_j = 0$. Therefore $\frac{\partial Y/\partial X_i}{\partial Y/\partial X_j} = -\frac{dX_j}{dX_i}$, where dX_j and dX_i denote the amounts of X_j and X_i which are substituted under the condition that output remain constant.

[8]The Elasticity of Substitution, σ, is actually the reciprocal of the formulation implied by this statement, in order to make its values accord with common sense.

$$\sigma = \frac{d\left(\frac{X_i}{X_j}\right) / X_i / X_j}{d\,(f_j/f_i)/f_j/f_i}$$

$$= \frac{d(X_i/X_j)}{d(f_j/f_i)} \cdot \frac{f_j/f_i}{X_i/X_j}$$

Before discussing other forms, let us pause to think what this concept implies, and what $\sigma = \infty$ implies, in educational production functions. The reason we want to substitute inputs is because their prices vary from place to place, or time to time. The variation in nominal input mix, however, is not observed to vary in adjustment to these prices. This is prima facie evidence that schools are not making economically efficient adjustments.[9] On the other hand, it means that the range of substitution possibilities which we observe is very limited. To say that the MRTS is constant ($\sigma = \infty$) over this range may not be a bad approximation at all. If we had more precise definitions of inputs, so that we could accurately assess the real range of variation of instructional quality, from facilities, peer influence, teachers and other adults, then infinite substitution between inputs at a constant rate would be unacceptable.

The introduction of interaction produces an elasticity of substitution which depends on the estimates of the parameters of the equation. For Equation

[9] Levin (14) persuasively argues that the market adjusts whether schools do or not, the difference being in the quality of the inputs. Thus, for example, where the market price of mathematics teachers is high relative to English teachers, the school should perhaps pay more for the math teacher, perhaps substitute non-human resources. Substituting a non-math teacher and calling the position "filled" is using an inferior quality resource. There is no reason to believe that this is a good solution to the scarcity of mathematics teachers. Owen (19) also presents evidence that school boards do not consciously attempt to purchase quality inputs.

2 above:

$$MRTS_{X_1, X_2} = \frac{X_2 + c_1 X_1}{X_1 + c_1 X_2}$$

$$\sigma_{X_1, X_2} = \frac{b_1 X_1 + b_2 X_2}{2c_1 X_1 X_2} + 1$$

This elasticity tends to be smallest in absolute value when X_1 and X_2 are present in about equal amounts, and largest when there is more of one input than the other. When c_1 is negative, then some values of X_1 or X_2 could produce a negative rate of substitution. This means that rather than give up, say, X_1 for an increment to X_2, one needs additional X_1 just to maintain the output. Surely no school wants to have such an input structure. Once again, given the range of variation we observe, this possibility is unlikely to occur. One might want to take extreme data values from a data sample to calculate the range of rates of substitution implied. This is not possible, of course, in the simple linear form.

The linear form with interactions then can produce a positive or negative elasticity of substitution, depending on the value of the interaction term (where X_1 and X_2 are defined so that b_1 and b_2 are positive). Substitution possibilities tend to be greatest when there is an abundance of one input relative to the other. This is generally considered a more reasonable characteristic of production than assuming no dependence of substitution on the levels of inputs.

The multiplicative forms have elasticities in general related to the equation's parameters.[10] A high σ indicates that substitution can occur at a rate which does not vary much over a large range. This would seem to be a desirable property of schools. If we had refined data on inputs, we would not want to _define_ this situation into existence. On the other hand, as before, we could define variables as ranges of input measures, and derive different substitution rates for the different ranges. The linear form with interactions and higher powers of variables is capable of assuming parameter-dependent values of important statistics, though of course the type of dependency is restricted.

Conclusions

This brief discussion of functional forms of educational production functions has barely scratched the surface of this subject. It has gone far enough to show the great importance of the way in which variables are defined. In general, the simpler the functional form, the more dependent statistics derived from the estimated equation will be on the particular definition of variables. This is true from the simplest adjustments, like multiplicative interaction, to more complicated adjustments, like determining relevant ranges of input measures which should be expressed as separate variables. In regression studies to date, little attention has been paid to the scaling or range of the variables--though a great deal to their construction in terms

[10] A particular form of multiplicative equation, the Cobb Douglas, has a constant $\sigma = 1$. This does not occur in general, however.

of educational sense--while simple equational forms have been utilized. I have tried to show that effort in these two areas, the form of the function and the definition of the variable given an input measure, are somewhat substitutable. We ought to observe one or the other. We generally observe neither.

The reason for this is obvious: There is no theory which would lead a researcher to prefer one form over another, one scaling over another. There are theories of how children develop, but none about what we can do to help them develop.[11] Indeed, it is not too strong to say that not only do we not know if schools work, and not only do we not know, if they do, how they do; we do not even have a theory about it. We have taxonomies,[12] and average relationships between inputs and outputs which might be useful for some policy purposes. But it is too much to ask of a production analyst that he go unguided in choice of functional form, when that form defines certain characteristics of schools which may be known to be incorrect. Thus I suggest that educators lend a hand by discussing production aspects of their schools with the analysts. A theory of school production could then proceed by suggesting not only which inputs might be important, but how these inputs are related in production. Until some progress is made in this direction, there is little hope of seeing estimates which have any pretension of being education production functions.

---

[11] See Bruner (4) for this distinction, and a start at developing a theory of instruction as opposed to learning.

[12] See, for example, Bloom (1), or Mood (17).

III

PROBLEMS IN PRODUCTION FUNCTION ESTIMATION

The purpose of this Part of the paper is to discuss some theoretical issues which are vital to good production function estimation. This is not an elaborate discussion, because by and large there is little I can do about the problems raised here. Some differences between one city and multi-city studies will be exposed. The following topics will be briefly considered: (1) multiple school outputs with single regressions, (2) simultaneous systems, and (3) non-frontier observations.

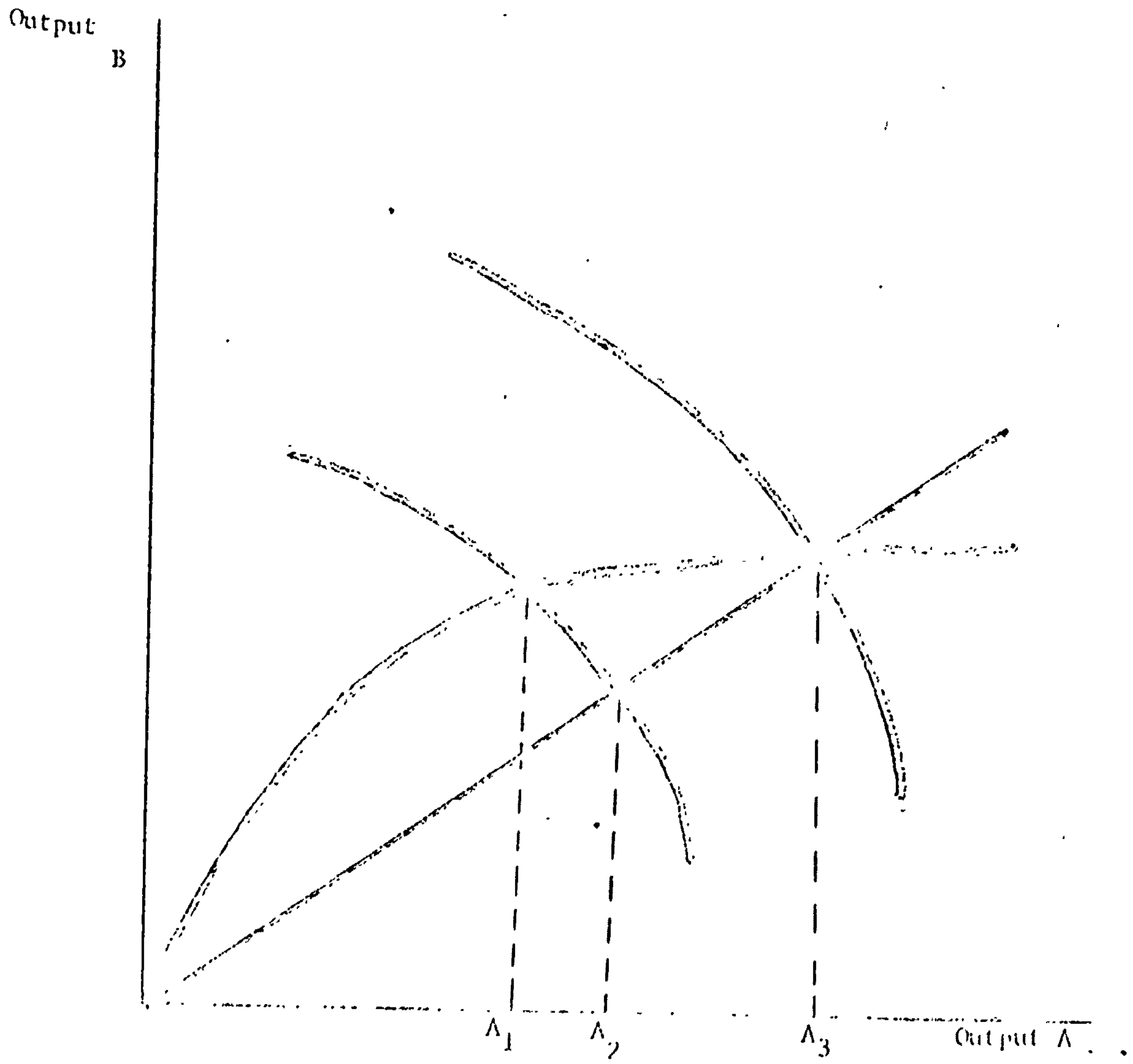
Multiple Outputs and the Single Equation

Since it is obvious that schools produce many outputs, one should raise the question: Under what conditions might it be satisfactory to analyze school production with relation to one input? I will append to this question the possibility of using one fixed weight index of output. Thus, for example, one might have used factor analysis to derive a weighting of the three EEOS academic tests (Verbal, Reading, and Mathematics), and used the factor weighted sums (or averages) of these tests as output observations.[1] The condition can be simply examined with production frontiers, as introduced in Part I.

If schools are structured to specialize in one output or another, then the input characteristics presumably will favor the output desired. For ex-

[1] For this sample I did, at one time, derive such weights. They were in the proportion 1:1:.[illegible] for the tests in the order in the text, and were essentially equivalent to just adding the scores. As will be seen below, factor analytic weights have no relevance to the required index. Thus I did not utilize these weights for further work.

FIGURE III 1



$A_2 - A_1$ indicates loss of output A in lower resource school from managerial discretion (stress on output B).

$A_3 - A_2$ is correct measure of output difference between these two schools.

can be devoted to one use only at some sacrifice to the other, let us ask what these two schools are likely to produce. Educators, for reasons which I have never understood, think that a certain kind of classroom decorum is a necessary precondition for learning. Therefore they are likely to expend their resources on producing that behavior first, and only after some success will they produce more academic achievement. Between cities, where the amounts of resources are certainly correlated with social class,[3] we can talk of an expansion path of outputs. This is drawn as the curved line from the origin in Figure III-1. It describes the locus of points on successive frontiers, that is, the relationship between choice of outputs and total resources. More resources are associated with more emphasis on output A.

Now consider a regression estimation using, as an output, only output A (if you will, achievement). The appropriate output measure from the nearer school, _for a production function estimation_, would not be the observed test score had that school produced outputs in the same proportions as the more endowed school. The distance $A_2 - A_1$ shows the amount of academic output which was not produced only because of management discretion, as opposed to the amount which could have been produced with the resources at hand. _Only when the expansion path is linear can a single output measure production._ Otherwise, we would have to know the amount of output B produced, and correct for it, before determining the effect of the resources in producing output A.

The condition under which a fixed weight index of outputs can be used as a single measure is that the production frontier itself be of constant

---

[3]See, for example, Owen (19).

slope (i.e., frontiers are parallel). In that case, when one output was sacrificed for another, the index would not change, using the denominator of the slope as the vertical axis weight, the numerator as the horizontal axis weight. This is clearly possible when the production frontier is linear, but it may also be true if the curves are "parallel" and the expansion path is linear. Thus a linear expansion path or parallel straight line production frontiers are alternative conditions for using a single output measure, the latter case allowing a non-linear expansion path, if we index the outputs appropriately.

The outputs considered here, behavior and academic achievement, are exaggeratedly different. Within the sphere "academic achievement," however, output differences are easy to observe. Shaycroft, for example, gives us correlations among 49 different output measures in both the ninth and twelfth grades (as well as test-retest correlations) separately for boys and girls.[4] Among all the tests, I have looked closely at the three which she brackets as "Mathematics Test" and the five bracketed as "English Test." The highest correlation within the mathematics test battery for an age-sex subgroup is .74 for ninth grade boys. At the twelfth grade, the highest correlation is .64 for the same two tests, the arithmetic reasoning and intermediate high school math. Within the English battery, the highest correlations occur between punctuation and English usage, at .63 for ninth grade girls, .60 for ninth grade boys, and .54 for twelfth grade boys and girls.[5] Between individual tests in

[4] Shaycroft (21), Tables 6.1a and 6.1b. In this peculiar Project Talent data there is no information on the race of the children.

[5] Twelfth grade girls actually related the punctuation test slightly better with spelling, correlation of .55.

the English and mathematics batteries, the highest correlations per subgroup occurred always between punctuation and either arithmetic reasoning or intermediate high school math, all highest scores in the .60 to .62 range. Thus variations in scores on one test are not extremely well related to variations in scores on another. (Average correlations were considerably lower, in the .40-.50 range.) Whether this is explainable by inherent "talents," by background, or by resource specialization in schools, it is a good indication that, at the margin, resources (home and school) produce one output or the other. Of course an addition to resources can produce more of all outputs. The curved expansion path in Figure III-1 indicates that more resources produce more of both outputs A and B. Thus one can expect considerable correlation between output scores. But this correlation will be reduced to the extent that different children are in systems with different expansion paths (even if linear), in systems with schools with varying resources and a single, but non-linear expansion path, or just in homes which stress different outputs.

The question remains, however, if <u>within a city</u> resources are distributed randomly enough that an expansion path is essentially linear. Even in this case schools may, indeed, choose to produce different outputs. Suppose we observe two well-managed schools with the same inputs producing different amounts of the output. The output measure in the regression is the horizontal axis, and the same amount of school resources are observed to produce different amounts of that output. Suppose the reason for the managerial discretion appears in our data as a "background" variable. Then some part of the estimated production relation between background and academic achievement actually indicates that behavior was acceptable, and therefore the output of the school was focused on the measure we are using.[6] The social class measure picks up the effect ascribed to managerial discretion in Figure III-1, and its coefficient is biased. However, the school coefficients may not be affected, according to the assumption that this discretion is random with respect to these resources.

This, of course, overstates the case. It is unlikely that resources are distributed equally within a city.[7] If one can argue that they are more equally distributed within than among cities, then this at least argues that a one-city analysis will be less biased. Since it is difficult to make that argument until one knows what is a resource, and how much of a resource it is; and since the effort here is to make that determination by estimating production relationships; the whole process seems circular. I will therefore flatly claim that resources are more equally distributed within than between cities, by social class of child. This makes the one city analysis seem viable, though not admirable. As noted in Part 1, the best data sample would have already ascertained the output focus of the schools, and chosen those along a single ray from the origin, covering a wide range of resources.

In conclusion, it generally appears inadmissible to investigate one output or output index with a single equation regression. I will indicate in the

[6] Of course the background of the school might be the relevant measure, but since that is higher the more high social class children there are, the social class of the child is correlated with that of the school. But not all students are treated alike, so that even within a lower class school a high class student might receive an academic focus, if it is institutionally possible.

[7] I have given evidence that within Eastmet resources are correlated with social class. See Michelson (15).

next section how one might accommodate several outputs. Since the major effort of this paper is directed at estimation under different specifications of the production relationships, and since these points are valid whatever other estimating procedures are employed, no more mention of simultaneous estimation of multiple outputs will be made after the next section.

Simultaneous Estimation

In a recent U. S. Office of Education volume, Henry Levin and I produced simultaneous equations estimations of several outputs, using the Eastmet data.[8] The major focus of our models was an attempt to incorporate attitude variables into the production estimation.[9] In this case, attitude measures produce test scores, and test scores produce attitudes. Although this no doubt does not actually occur simultaneously, a simultaneous estimation is required if it occurs within the time period of our investigation. As these are models of school production, from first through the measured grade, certainly attitudes and outcomes have interacted, and are considered "simultaneous" within the production period.

The need for a simultaneous estimation of separate outputs which are not inputs into each other's production functions is somewhat more complex. One

[8] See Levin (15), and Michelson (16).

[9] Bowles had previously correctly noted that inclusion of attitude measures into a single regression equation produced biased estimates. He directly estimated reduced form, excluding the attitude measures (1). The advantage of simultaneous estimation is to separate the direct effect of school resources on academic achievement from those operating indirectly through attitudes as well as to estimate the effects of resources on the attitudes, which could be the more important output measure.

would want to control for the production of some other output in assessing the influence of the resources on the output of interest. Thus a negative relationship would be expected between some outputs and others, net of the influence of the total amount of resources which induces a positive relationship between outputs. The best way to do this, as has been indicated, is to choose data points along a linear expansion path. Otherwise, the procedure for unbiased estimation involves a two-stage regression equation, in which alternative outputs are considered endogenous in an estimate of the output of interest.

The relationship between outputs in this system is not a production relationship of the sort "a positive self-image produces higher reading score" and "a higher reading score produces a positive self-image." Rather, "given the resources observed, and the amount of output B which these resources ordinarily are associated with, ____ amount of output A is produced." Two stage least squares was the algorithm used to solve the simultaneous systems in the references given above. The reader is referred there for more explanation. The point here is that, despite the different interpretation, several outputs can be inserted into one equation with proper estimation techniques to derive resource effects on one output net of the other.

It is doubtful that this technique could be usefully employed to determine empirically the marginal trade-off's between outputs until better output measures are available. There will certainly be a good deal of work on school output measures in the future, motivated by the lack of relationships between current output measures and later-life success.[10] With these better measures

[10] See Bowles (2), for example, for some evidence on this question.

we will also be further along in estimating production functions, and will be able to attempt to talk about a production system in which several outputs are produced by their identified equations, and a solvable set of equations is estimated.

Isolating Frontier Schools

We observe output in children from several schools, not all technically efficient. I assume away the trivial case where the difference between the efficient output and our observed output is a constant for all schools.[11] I will discuss the following two exclusive and exhaustive cases:

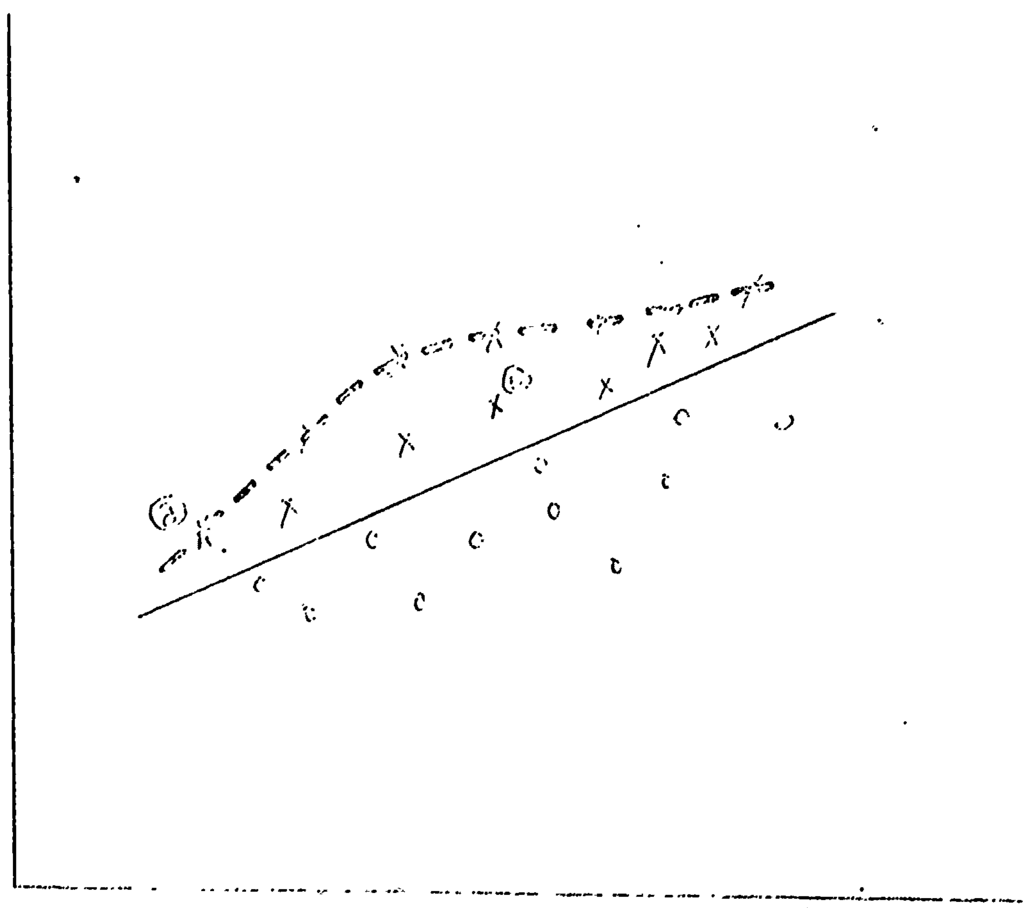
(1) Inefficiency is random with respect to all the variables we measure. Inefficiency just strikes some schools, or some school districts, independently of the nominal characteristics of teachers and principals, and independently of the social class, race, nativity, etc., of the school population.

(2) Inefficiency is related to some characteristic which we measure.

In the first case, estimates of the "frontier" of production from a given set of inputs will strictly speaking not be the frontier at all. It will be a kind of average output attainable with an average amount of inefficiency. This is inherent in estimation techniques in which the "best fit" places the fitted hyperplane within the data observations, minimizing a measure of error both above and below the observed output. The solid line in Figure

[11] Where the linear form is used on transformed variables, and as logarithms of observations, I assume the technical inefficiency is not described by a similar transformation.

FIGURE III-2



III-2 illustrates a typical regression fit.[12] However, the dashed line indicates a smooth locus of maximum observations. All "error" must be reductions from a true frontier.[13]

This frontier is not found by taking all high output observations. Point "a," for example, has a lower Y value than point "b," yet "a" is on the frontier, whereas "b" is not. An easy way to find these frontier schools is to estimate the solid line, and consider only schools with positive error:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \ldots$$

$$e = Y - \hat{Y}$$

defines error.

Divide each X_i variable into ranges, and find the schools with the largest error in each range. This gives a series of schools which do better than expected, where by assumption this is not because we have omitted some important variable, but because these schools use their resources most efficiently.

Since by construction inefficiency was random with respect to the X_i characteristics, the sample of efficient schools should be a random subsample of the entire data set. Differences in both the level of the frontier and its slope with respect to any X_j variable indicate an advance in precision of

---

[12] This should be considered a partial relationship where the other inputs are held constant. A one input production function would not be very interesting.

[13] The problem of measurement error, which can still vary around the frontier, is neglected here. We are considering differences in efficiency as creating all the observed error. If measurement error is large relative to efficiency differences--a real possibility with educational data--then this procedure fails to isolate frontier schools.

the estimation.[14]

One could follow the same procedure under case number 2, where inefficiency is related to a measured characteristic, say X_j. Now, however, there are two major problems. First, to the extent that X_j is in fact a proxy for technical efficiency, then the remaining observed error must be measurement error or due to variables we have failed to measure. As mentioned above, this error is legitimately found on both sides of the "frontier." Thus choosing positive error schools is a matter of chance, not of precision. Furthermore, the sample that will result may not be a random subsample of the original population, but may be those schools with high values of X_j.[15] A regression estimate on these schools will hopelessly confuse efficient management with the specific abilities of these pupils to progress with or without efficient management, or anything else that stratifying on X_j might accomplish. Thus if X_j were social class--i.e., if the upper class schools had better resource management (for the outputs considered) than lower class schools--then we would have no way of estimating a frontier for lower class schools.

If, in addition, the schools with better resource management also have more inputs, then the existence of these inputs will appear more highly correlated with outputs than they would be under average management. It is difficult to tell whether there is a real difference among principals' abilities

[14] Heteroscedasticity, which is commonly analyzed as not affecting the slope of a regression estimate, would markedly affect the slopes here. If var(e) increased with X_j, for example, then the original b_i estimate would be lower than the revised b_i estimate. Nonetheless, one could not say that error was correlated with X_j--a violation of the assumptions of construction--if the original heteroscedastic variance was truly symmetrical around the fitted regression line.

[15] This will be true if the true relationship between X_j and output is multiplicative, but the estimate is linear.

to manage their resources, in terms of producing outputs. To the extent that this is true, and known, it will be reflected more in analyses involving more than one district, than in analyses confined to one district. The arguments for this problem of multi-district analyses have already been presented. If the real resources purchased vary between districts, then the variables representing social class (or, possibly, race) will incorporate the managerial gains in these districts. A multi-district study, then, corresponds to case number 2, where technical efficiency is correlated with variables in the analysis, probably with social class variables. A frontier cannot be determined, nor can the effects of "social class" variables be interpreted.

Within a district, as I have argued, the situation appears to be closer to case number one. Principals are probably approximately randomly distributed with respect to their technical managerial ability. An attempt to locate true frontier schools appears in Part IV of this paper.

IV

EXPERIMENTS WITH FUNCTIONAL FORMS

In this Part of the paper I will present regression estimates of relationships between inputs and outputs in one city. I call this city "Eastmet," and have been studying it and another city, Northmet, for some period. This paper presents the first results using weighted regressions on this sample, and also the first results using a variety of functional forms. In the sections that follow I will (1) discuss the origin of the data, and definition of the variables which will appear; (2) discuss the effect of using one city; (3) present regression estimates of simple linear equations for two different output measures; (4) add interaction terms to the simple linear form; (5) consider transformations of the variables; (6) consider squared terms and interactions together; (7) attempt to isolate frontier observations; and (8) assess the results and conclusions to be drawn from this experimentation.

The Data Sample

The data employed here come from the original data tapes of the Equal Educational Opportunity Survey (EEOS), which was conducted in September, 1965. The original presentation of results from this survey appeared in *Equality of Educational Opportunity*, by J. [illegible] Coleman, Ernest Campbell, and others.[2]

[1] The data sample here was created by myself and Henry Levin, for joint work. Some discussion of earlier data and results has appeared in Levin (13) and (14); Michelson (15) and (16).

[2] For full citation see (7).

The largest section of that document was concerned with presentation of summary statistics showing how various measures were allocated over the population. The most controversial section, and therefore the one on which the report has gained notoriety, was an analysis of the relationships between school measures and pupil achievement. A steady flow of criticism has followed this analysis.[3] One thing must be said, however, strongly in its favor: at no point did the authors contend that they were estimating functional causal relationships between school inputs and their output measure, a verbal facility test. They reported average associations, not estimates of production functions.

The major difference between the work reported here and most other work with this data is the decision to use one city only. The effect of this decision is discussed in the next section. There are a number of other data refinements which I consider vital to proper estimation of input-output relationships. For example:

> One would have to make sure that the school characteristics associated with each pupil were those of the school actually attended most of the time. School A could hardly have influenced a student who just transferred to it after spending the previous six years in School B. This problem might be solved by using the characteristics of School B, or possibly by including in the analysis only pupils who had been in School A for several years. Similarly it might be advisable to exclude those schools whose characteristics altered markedly during recent years.[4]

[3] See, for example, Bowles and Levin (5), Kain and Hanushek (10), and Cain and Watts (6).

[4] Kershaw and McKean (12), pp. 23-24. I include this entire paragraph because this report, dated 1962, appeared years before the EEO study was even envisioned. Yet in four years of analysis and reanalysis, only Levin and myself have paid attention to this advice.

There is no way to know in what year a pupil transferred into the school he is now in, and of course tracing pupils to other schools would not only have been expensive, but in most cases virtually impossible.[5] We do know, for our sixth grade sample, how many children had been to other schools; the present sample eliminates those that answered that they had been in more than one school.[6]

We divided the sample by race, eliminated those children who reported no sex, those in schools with incomplete records, and those in suburbs of Eastmet.[7] From a city and suburb sample of 4505 children, this left a sample of 1055 black and white children, of which the 597 whites are used here. Of the original 36 Eastmet city schools, 35 could have appeared in the white sample, as only one school was all black. However only 30 schools survived the pruning. Several schools, in fact, are represented by only one child. In previous regressions this has not been a major concern, but in this paper it is.

A major problem when one wants to estimate a production relationship is determining the appropriate production unit. If each unit is an observation, then each unit should have equal weight. When we are talking about average relationships facing children, then children are the appropriate unit. Each child is equal weighted, and his situation is recorded. If many children are in the same situation, but respond differently--i.e., have different output scores, but the same inputs--then the correlation between inputs and outputs

---

[5]The major question here is what if the student came from a school not in the original sample? Should school-wide data be collected for that one child?

[6]The question read: "How many different schools have you gone to since you started the first grade?" The first possible answer to this question was: "One--only this school." Only children who checked this answer remain in our data sample.

[7]Some schools sent one set of records, say, from children, but not another,

is reduced. Although researchers who use individual data as opposed to grouped data complain that their R^2 are low (I have been known to be among them), it properly is so, for it says that these children are not subject to a firm relationship indicating their test score.[8] Although statistically one likes to have a perfect fit of his regression equations, one does have to wonder morally what kind of a world it would be if we could predict perfectly a child's reading score from knowledge of his social class and school resources. The $\bar{R}^2$ we get are high enough. I would be frightened by a more determinate world.[9]

The task at hand is not to estimate relationships averaged over children, but to estimate technical relationships of production. In this case, the production unit is the school. Each school should have equal weight. On the other hand, each child should be allowed to enter his own background. That is, some correction can be made on a per-child basis for differences in quality upon entering the production process, and differences in ability to respond to the production process. As has already been indicated, I can correct for both of these effects to some extent, but I cannot easily distinguish between them. The following formula weighted each school equally, and each child equally within each school:

---

[8]On the other hand, there is a great deal of error in individual scores which is reduced by grouping. If error dominates the individual child regressions, they are of no advantage. We do not know that this is the case, however, and until someone shows that it is, I will continue to accept the logic which calls for using the individual variations which we can observe.

[9]It would also be frightening to find that the world is, after all, linear. I will give enough evidence below to dispel that notion, however.

$$W_{ij} = \frac{597}{30 \cdot N_j}$$

where N_j = number of children in school j

W_{ij} took the extreme values of 19.90 (for the one-child schools), and .13 (for the largest school). Some children were therefore weighted over 150 times others. As shockingly explicit as this weighting is, it is not unusual. Those researchers who use data grouped by school are doing the same thing, except they ignore intra-school variation indicated by background variables. Unless they weight their observations by N (or one could argue for $\sqrt{N}$), they are using per-pupil weights similar to those used here.

Production function estimates, then, require a different kind of data set than was collected here. I am correcting as well as I can for that deficiency. In a survey designed to produce data for production function estimation, we would want to take a representative sample from each observation, to try to get approximately equal-sized samples per school. In one school this might mean sampling 1 in 100, though in another school the sample might be 1 in 10. There is no need to get more observations just because the school is bigger, if it is treated as a production unit. If there are economies or diseconomies of scale, this should be indicated by a scale factor. If different kinds of children go to big than to little schools, this should be corrected for by covariance techniques. But the theory of sampling for production information is different from that used in the FLOS report, which was investigating average characteristics of children. Mobile children must be treated separately, in production function estimation, or not at all.

Four kinds of variables have been defined. First there are control variables for sex and age. These are binary coded variables. The children are in the sixth grade, where it is a well-known phenomenon that girls do better than boys on achievement tests. Students who reported that they were 12 years old or older are separated by a binary variable.[10]

Second, there is a set of student background variables. These control for the quality of the input upon entering the production process, for continued production during the period of schooling, and interaction with resources in school. An Index of Possessions, the child's report of his Father's Education level, the child's report of the number of People Living in his Home, and his report of whether he attended Kindergarten appear as background variables. Four school variables are used as production estimates: Teacher Test Score, Teacher's full-time Teaching Experience, Teachers' Racial Preference, and the Principal's report of whether the school engaged in tracking. Teacher's Racial Preference is, as it says, a question which asked what racial composition the teacher preferred. A higher answer indicates preference for whites. The Test Score was a 30 question vocabulary quiz.

The teachers were selected for this sample if they indicated that they taught in the third through fifth grades. Their individual responses were averaged, and the average applied to each student.[11] Teachers in the third

---

[10]Other forms for the age variable were experimented with, but in a binary coded classification a variable denoting exceptionally young children was not significant, whereas the one for old children was. Thus this one binary variable suffices.

[11]In future work by Levin and myself we will discuss the implications of this averaging, with some estimates of the kinds of observational error it can imply.

grade who reported they had not been in this school when the sixth grade students were third graders are not eliminated. There is a bias either way: to eliminate them would make the sample too old and experienced, as there is a good deal of turnover of young teachers. Including them makes the sample too young and inexperienced, as to some extent young teachers replace old. Furthermore, the biases are different in different schools. This means that a really careful data collecting job for production estimates should collect data on teachers who were there, in lower grades, when the students were.

The final variable type is a binary variable which describes a certain amount of interaction. By listing the schools and their characteristics, I was able to discern four in which the school resource measures were somewhat low relative to the social class of the students, and four in which the resources were quite high though the students were of quite low social class.[12] For each set of schools I defined a binary variable if the student was himself of above average social class, and another variable if he was below. Thus the student's class is interacting with a general description of the match between his school's resources and his peers' social class. Of all these variables, two survive into this exposition: HiSes-LoRes-MdPeer indicates an above average social class student in a school with low values of resources and middle range of peers. LoSes-HiRes-LoPeer indicates a below average student in a high resource school, with low class peers.

Means and standard deviations are given in Table IV-1 for the two output variables and the school variables. The means in the first two columns are calculated per child. That is, this is not the average teacher characteristic, nor

[12] Another group had high resource values and middle social class.

the average school characteristic, but the teacher or school characteristic which, on the average, is faced by an Eastmet sample child. The "Total Sample" columns refer to the sample of white children in Eastmet city and suburbs with complete school records. The "Regression Sample" columns are averaged over the 597 children who attended one central city school from the first grade. The statistics presented here are so close because the one-school only sample in the central city represents higher performing children than the city children in general. Thus the mean scores are lower in the city than in the suburbs, but higher among one school children than more-than-one school children, and these differences about balance out. Similarly, Teacher Test Score is lower in the suburbs than in the city, but higher in the city for these children than for the entire sample. Experience is lower in the suburb, and this difference is not corrected by taking this select sample. Similarly, Tracking is more prevalent in the central city.

The third column of Table IV-1 contains the means of the weighted variables as they actually entered the regression equation. These are averages over schools. Apparently the higher scoring children are in larger schools, in this sample, than the lower scoring children. Since, as the comparison of columns 1 and 2 shows, the pupils are representative of the Eastmet sample, the question arises whether the schools are representative of the Eastmet schools. I have not been able to determine the extent of representativeness for this paper. Since there is no test of the representativeness of the entire Eastmet sample, it is not clear how much information would be gained by knowing how like the Eastmet sample the sub-sample is.

Additional variables will be defined below, but in all cases they are transformations of these variables. What a variable means should not be confused with its name. The tracking variable, for example, defines two groups of schools. Twenty-two schools had the value 2 (track for all students), seven the value 0 (no tracking), and only one between. But what characteristic about these two groups makes them different is not necessarily the degree of

TABLE IV-1

Means and Standard Deviations
of Production Variables

Eastmet City Whites

	Regression Sample N = 597		Total Sample N = 1727		Weighted Sample N = 597	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Verbal Score	35.1	10.1	35.4	10.4	25.3	11.9
Reading Score	23.6	7.1	23.4	7.3	17.8	8.2
Teacher Test	24.7	1.8	24.6	1.7	23.0	2.4
Teacher Experience	15.1	5.1	13.6	5.0	10.8	4.6
Teacher Race Preference	6.9	1.0	7.0	1.0	5.8	1.1
Tracking	1.5	.8	1.2	1.0	1.5	.8

tracking. Indeed, all these schools probably engage in a form of tracking. The difference could be in the principals who admit it and the principals who do not, or in any number of unimagined characteristics which happened to be picked up by this classification. For this reason, I will follow this convention: When I am referring to the variables in the equations, I will capitalize them, no matter what the context. Thus Experience always means the Experience Measure in my schools, whereas experience means real teaching experience. I will simply avoid beginning a sentence with the word Experience if I mean experience. At times I use different words to mean the same thing, such as Preference for Whites instead of Racial Preference. The meaning should be clear, and the capitalization rule will always apply.

The One City Sample

Most analyses using EEOS data use many cities. Some, in fact, like the original presentation by Coleman and his allies, do not even investigate the city representation.[13] There seem to be good reasons to include a number of cities in one analysis, and good reasons on the other hand to study one city only. Some of the differences in results are revealing, and deserve some exposition.

In Part III the importance of knowing the focus, the object of the school's education was explained. Thus, for a crude example, one would not mix academic and vocational or technical high schools in the same production

[13] We are never informed, for example, how many schools are represented in Coleman's sample of 1000 students, nor, besides large regions, where these schools are.

function analysis.[14] One might expect that different kinds of cities aim at different kinds of outputs. That is, at the high school level--and therefore indirectly at lower levels--the politically dominant group determines the focus of the school. That this focus is real and definable and political can easily be seen in cities which change their nature. Brockton, Massachusetts, for example, in changing from a blue-collar to a white-collar city, as the industry changed from shoes to electronics, has had corresponding changes in the focus of its high school with reasonably open political debate.

To the extent that the input structure of a school reflects the aim of that school, then the only problem with including schools of different aims in one analysis is one of interpretation. As a production function, the measure would still be incorrect. As a determinant of average relationships, it would not be bad. For example, suppose schools which tried to place their students into prestigious colleges deliberately hired teachers with academic majors in college, and schools which tried to place their graduates in the labor force deliberately hired teachers with education majors.[15] As an average relationship, we would find that academic majors of teachers are associated with college-type skills--say, Verbal Score. It would be wrong to assume that a school which hired more academic majors would necessarily produce as much of these skills, as indicated by their association, without a concomitant change in policy. The estimates, as production estimates, would be biased

---

[14] Burkhead (5), for example, specifically excludes 12 technical and vocational schools, and one school for the physically handicapped, from his analysis of Chicago high schools. On the other hand, he excluded only two vocational high schools from his Atlanta sample, apparently not recognizing that the five Negro schools were also "technical" schools in that peculiar euphemism of the South. The EEOS high school data does not identify the school as academic or vocational.

[15] Other types of schools are assumed to fall in the middle.

upwards because they had not accounted for managerial discretion. On the other hand, these particular types of teachers presumably are hired because they have an effect in the direction which the school is emphasizing. Therefore it would still be correct to assume that there would be some effect on college-type skills from hiring more academic-major teachers, even with no managerial change.

This explanation, however, asserts more rationality and technical competence on the part of school authorities than probably exists. Indeed, John Owen has recently attempted to identify supply and demand characteristics of teachers, to see if he could find if schools deliberately sought "quality" teachers by offering higher pay. He found that this was not the case.[16] Thus between cities we might expect the relationship between the focus of schools and their input structures to be essentially random. This could lead to unbiased estimates of real production relationships.

The problem with this argument is that, though as far as school board demand is concerned the characteristics of teachers is not a function of the focus of the school, they still may be so. A blue collar school on a blue collar budget has a smaller supply of academic major teachers, and therefore has a structure dominated by education-majors, even if the school would like (but cannot afford) more academic majors. There will be variation within that school district, and a school with more academic majors may produce more college type outputs. But if this system was pooled with others, then in some other system this same percentage of academic-major teachers might produce more college type

[16] See Owen (19).

output; and education majors would produce less blue-collar output than found in the first school. The result could be no academic effect found for academic majors, or blue-collar effect from education majors even though both effects occur in both districts.[17]

It would be convenient to argue that this dilemma disappears in a one-city analysis, because the focus of the schools is constant. As we know, however, this is not the case. In fact, we can be reasonably certain that within the 30 elementary schools in the Eastmet white sample, some schools are more oriented towards producing the skills tested by the EEOS tests than others. The best analysis, clearly, would involve choosing from several cities those schools with common skill goals, and testing the production of those skills. There are, nonetheless, some further arguments for using a one city sample, which can be reproduced from Burkhead:[18]

> A great many of the variables whose influence on output is difficult to isolate are held constant for a single city. The labor market for teachers and the market for other factor inputs is reasonably uniform for the city as a whole and . ... [therefore] a given outlay will purchase inputs of similar quality.
>
> Some aspects of "administrative responses" may also be uniform within a single large-city system.

Since in preparation of the Eastmet sample I experimented with regression estimates including suburban schools with the Eastmet city schools, I had a chance to notice some differences in results. One will be stressed here, be-

---

[17] For a diagrammatic exposition of this problem, see Michelson (16), pp. 123-125.

[18] Burkhead (5), p. 39.

cause it points to an important problem of interpretation. The importance of Teacher Experience--importance defined as regression coefficient, Beta coefficient, or increment to R^2--is greater in the single city sample than in the city-suburb sample. There seems to be a good explanation for this, which casts considerable doubt on an interpretation of the Experience coefficient as indicating a production relationship.

Suburban systems tend to have younger teachers than central city systems. That is, among districts, social class and experience are negatively correlated. This is only partly explained by the rate of population growth of the suburbs, i.e., the relative newness of the positions occupied, naturally, by relatively new teachers.[19] The rest of the explanation presumably lies in deliberate policies of suburbs to maintain a turnover in staff so as to minimize the cost of expensive experienced teachers. Given a surplus supply of teachers to suburban systems, they can operate this way.

Within any system, city or suburb, a positive correlation exists between teacher experience and pupil social class. This is due, at least in part, to the well-known seniority choice system: the more senior teachers can choose to fill vacancies in schools before the new teachers are assigned. This is virtually universally true; and it is observed to occur. Thus the experienced teachers[20] are found in the schools which produce--or from which emanate--academic skills, at least partly because they associate themselves with children who will, with them or without them, acquire these skills.

---

[19] See Owen (19) for evidence on this point.

[20] Strictly speaking, experience in present system, not over-all experience, is the correct measure of seniority. These measures are highly correlated, however.

Since this seems to be a description of the system at work, we would expect that regression results would show Experience more associated with output within than between systems. As noted, this is exactly what we do find. The inter-city estimate would therefore be better for determining the actual production effect of Experience than the one-city estimate. For this reason I will de-emphasize the strikingly significant relationships between Experience and the test score outputs found below.

This argument does not pertain to any variable other than Experience. To the extent that teachers with higher Test Score want to move to higher class schools, they are perhaps more able to do so between than within districts. By their personal appeal in interviews, they may be preferred in new hiring by the suburban schools. But they would not have seniority in a single district. So the association between Test Score and output Scores of children probably better describes production in one-city estimation, and has more of a component of deliberate and prior association in the multi-district analyses. This argument holds for Race Preference also, and probably *a fortiori*.[21]

Simple Linear Equations

The "simple linear" equational form has been implicitly defined as that where the variables are linearly additive, not transformed, and not involved in interactions. Since this does not exclude dividing variables into categories and entering these categories separately, the simple linear form does not necessarily mean that the relationship between output and the ori-

[21] By a similar argument, school management efficiency is likely to be random within a city, but associated with social class between cities.

ginally coded variable is linear. For example, Experience as a scaled variable might be independent variable X_1 in the following equation:

$$Y = a + b_1X_1 + b_2X_2 + d_1Z_1$$

where Y is a school output

X is a school resource

Z is a background measure

The effect of an increase in one unit of X_1 is to increase Y by b_1 units. However Experience could be a categorized variable, where:

X_1 is 4-8 years of experience

X_2 is more than 8 years of experience.

We can no longer consider "a unit increase" in Experience, but must consider a shift of categories. The constant, a, includes the effect of having 0-3 years Experience. The effect of moving into the 4-8 category is b_1, and the effect of moving into the 8+ category is b_2. Thus in terms of the scaled Experience variable, non-linear effects are allowed for. In the equations presented here, I do not take advantage of this possibility, and present scaled variables for inherently scaled measures.

As mentioned in the section describing the data sample, I will add two binary coded "interaction" variables. Being in the category defined by the interaction gives a child a value of 1, and not being in that category gives him the value of 0. The coefficient, then, indicates the gain or loss in output, other things equal, from being in that category, i.e., from having that particular type of interaction.

Two equations are presented in Table IV-2, in a format which will be continued throughout this Part of the paper. The coefficients are listed in a column, and underneath each coefficient is its standard error. The standard error is essentially an estimate of the standard deviation of the random error around the regression coefficient. If this standard deviation is "large," then there is a great deal of error, and the coefficient may well not be what the estimate says it is. "Large" can be defined in terms of the coefficient itself, implicitly testing whether the error is such that the true coefficient may likely be zero, the estimated coefficient being a result of sampling error. A convenient rule of thumb is to reject a coefficient if it is smaller in absolute value than its standard error (i.e., if zero lies within one standard error of the estimate of the coefficient). Almost all coefficients presented will be larger than their standard errors.

Random error is assumed distributed according to the normal distribution, or "bell-shaped" curve. The bulk of the error lies close to the mean. In fact, less than 5 percent of the error is more than two standard errors from the mean. Thus if the coefficient (in absolute value) is more than twice its standard error, it is highly improbable that the true coefficient is zero. "True" here does not refer necessarily to the real production relationship, but the coefficient which would be found if what we measure as error is truly random, and we sampled 100 percent of the relevant population.

The dependent variables, or outputs of production, are Verbal Score and Reading Score of these sixth grade children. The first two columns of Table IV-2 list the equations for these two variables, the specification of the equations (i.e., which variables are included) differing because of the

TABLE IV-2

Simple Linear Regressions

	Regression Coefficients		Beta Weights	
	Verbal	Reading	Verbal	Reading
Sex	2.54 (.69)	3.35 (.48)	.106	.205
Age 12+	-8.82 (1.17)	-4.12 (.87)	-.207	-.141
People in Home	-.221 (.12)		-.056	
Possessions	1.06 (.17)	1.66 (.11)	.217	.495
Father's Ed.	.572 (.14)	.219 (.10)	.123	.069
Kindergarten	3.26 (.79)		.131	
Teacher Test	.999 (.18)	.347 (.11)	.201	.102
Teacher Experience	.490 (.076)	.329 (.06)	.189	.185
Racial Preference	1.38 (.39)		.126	
HiSes-LoRes-MdPr	4.63 (1.86)	3.08 (1.34)	.074	.072
LoSes-HiRes-LoPr	-10.94 (1.87)		-.190	
Tracking		-1.63 (.28)		-.170
Constant	27.52	-9.84		
R^2	.611	.525		

statistical considerations (the coefficients relative to their standard errors) just discussed. I see no theoretical reason why People In Home should be related to the Verbal Score of the children, but not to Reading Score; nor Teacher Race Preference nor whether the child went to Kindergarten. One could argue, after the fact, that this makes some sense, but I had no particular expectation a priori about these variables. Similarly, whatever it is that the Tracking variable indicates seems related to Reading, but not to Verbal Score of these children, although I have no theoretical explanation for this.[22] At the bottom of these first two columns is the constant of the equation (a), and the R^2, or measure of the percentage of variation in Y which has been accounted for by the X_i and Z_j. Although the specification of the variables in the equation did not proceed in an attempt to maximize R^2, the algorithm which estimates the parameters (the coefficients), given the specification, does do so. These coefficients then are the set of coefficients which best explains variations in Y, for the given list of input variables.

The third and fourth columns present "Beta Weights." These are the regression coefficients weighted by the relative size of the standard deviations of the input and output variables. One can say "a one standard deviation increase in X_i will produce a β_i standard deviation increase in Y" (assuming β_i is positive). Since the units of the variables sometimes have

[22] These results show, however, the extreme weakness in most presentations of these analyses which concentrate on one output. The justification often is that all the EEOS output measures are correlated. This is certainly true: For this sample, the correlation by pupil between Reading and Verbal Score was .84; for the entire Eastmet sample, not stratified by race, number of schools attended, etc., .85. Nonetheless, high correlation does not mean that input-output specifications will be identical, nor even that inputs will have the same relative effect on both outputs.

no particular intuitive meaning, expressing them in terms of Beta weights can be quite helpful. With an unrepresentative sample, as this surely is, the Beta weights are considerably more suspect than the regression coefficients. In addition, one may not care about historical variation: Not "what is the reaction of Y to one standard deviation increase in Teacher Test," but "What would Y be if Teacher Test were at 30--the maximum" might be the relevant question. The regression coefficients are used to answer this. However, absent prices--which eventually become the crucial element in judging effectiveness of one input vs. another--the Beta weights give some sense of the relative import of one variable as opposed to another. All equations will present both sets of coefficients, but the discussion will focus on the regression coefficients.

The effects of the control variables for sex and age are striking. Girls are 1/4 a standard deviation ahead of boys in Verbal Score, and nearly 1/2 a standard deviation ahead on Reading Score, adjusted for social class, age, and access to school resources. The 27 exceptionally old children, adjusted for sex, etc., are virtually a standard deviation below the mean in Verbal Score, and 2/3 of a standard deviation below in Reading Score.[23] Of the background variables, Kindergarten attendance is quite important in terms of generating Verbal Score points, Father's Education could be important in

[23] The equation estimates the mean output measure exactly--except for rounding errors--when the input variables are set at their mean levels. The effects reported here consider extreme values of the control variables, as compared to their mean values. The estimate of the difference between boys and girls is obtained by setting this variable at 1 for girls, at 0 for boys. The mean of the age control is .085, and therefore one can think of the differences reported here as differing from other children or differing from the mean of all children with little consequence to the results. The standard deviations come from Table IV-1.

the case of large differences, but Possessions is the most important in terms of observed variation.

Of the school variables, Teacher Test seems the most important. Experience is suspect, for reasons detailed above. But for what it is worth, two years of experience would seem to "trade" for one Test Score point in producing Verbal Score, and one year of Experience is worth one Test point in producing Reading Score. The shift from average (per pupil) to maximum quality teachers would produce an average of about 5 Verbal Score points--one half a standard deviation--and less than two points--about one quarter of a standard deviation--of Reading Score.[24] Using per-school averages, the gains would be somewhat greater. Teachers who preferred all-white schools would presumably produce about four more points of Verbal Score than the mean Racial Preference, or two fifths of a standard deviation. The difference in Reading Score between Tracking and Not Tracking would be nearly one half a standard deviation, though the anticipated improvement of not Tracking over the mean would be three quarters of that amount.

The interaction terms included in the simple linear form as simple variables also have striking values. The combination of having low background and being in a school with low background will produce eleven points less of Verbal Score than would otherwise be predicted, if the school has high resources. That is, these resources seem to affect other children, but not these particular children. That could be because they do not receive the high resources in the school, and this variable corrects for an error in the resource

[24] Since neither equation held constant the other output, the effects of increasing resources are additive between equations. This is, to be honest, an indication that the original coefficients should have been estimated by more involved simultaneous procedures, which were discussed in Part III of this paper.

variables for these children, or because the combination of their own background culture plus a similar dominant school culture simply swamps the effect which resources has on other children. Eleven children had this characteristic, possibly too few to draw any conclusions about. Thirteen children in this sample with higher than average backgrounds were in schools with low resource values and average peers. They did better than the effect of the low resources would have estimated. This is possibly because they in fact had high resources within their schools, masked by the averaging process. It is also possibly because the culture of the school was not such as to prevent them from learning, and their learning was derived from their background characteristics. Since the effect of within-school resource allocation cannot be separated from that of peer-individual interaction, there is no way at present to choose one interpretation over the other. But the magnitude of these coefficients suggests that the assignment of these children into separate classifications did in fact reflect some reality, even if I cannot without being arbitrary explain what that reality is.

Two things which have been emphasized over and over should be more clear now. First, in discussing the "trade" between a year of Experience and a point of Teacher Test Score, no decision could be made on which was a better buy--assuming both represented production estimates--without knowing how much they cost. Second, the idea that there is one rate of trade between these two items, independent of the amounts currently present in a particular school, seems especially far fetched when the numbers are at hand. In this paper I will do nothing to add information about prices. However, let us proceed immediately to consideration of equations which allow for level-dependent estimates of the marginal output effect of input changes.

Interactions

As discussed in Part II above, the multiplicative interaction term allows an estimate of the effect of an increment in one variable to be dependent on the level of another variable present. Needless to say, interaction terms could be defined in terms of a product of more than two variables, or in categories of interactions. For this demonstration, simple products of two variables have been entered, and equations have been refined on consideration of these variables.[25] The interaction equations appear in Table IV-3.

In the Verbal equation, Teacher Test does not appear as a variable except in the interaction terms. We can calculate the rate of change of Verbal score of students with small changes in Teacher Test as follows:

$$\frac{\partial Y}{\partial T} = .165 \text{ (Experience)} \cdot -.116 \text{ (Race Preference)}$$

It is appropriate to consider the value of this expression at the weighted means, or the average of school means of the variables. This value is 1.10. The regression coefficient and hence partial derivative of Teacher Test in the simple linear estimate was 1.0. The standard error of this estimate was .18. Thus the value from this equation with interactions, evaluating at the means, is well within one standard error of the value estimated from the simple linear form, and therefore not statistically distinguishable

[25] One should pay more attention to the units in which one defines multiplicative interaction than I do here. One variable may dominate the interaction either by the size of its units, or by a large variance relative to the other variable. The coefficient will be interpreted as applying to both, however.

TABLE IV-3

Linear Regressions with Interactions

	Regression Coefficients		β Weights	
	Verbal	Reading	Verbal	Reading
Sex	2.77 (.69)	3.61 (.47)	.116	.221
Age 12+	-8.47 (1.16)	-3.79 (.85)	-.199	-.130
People in Home	-.173 (.12)		-.044	
Possessions	1.10 (.17)	.918 (.27)	.225	.274
Father's Education	.532 (.14)	.196 (.098)	.114	.062
Kindergarten	3.65 (.79)		.146	
Teacher Test		-1.30 (.41)		-.383
Teacher Experience	-3.53 (1.18)	-3.62 (.83)	-1.363	-2.045
Race Preference	4.23 (2.26)		.385	.035
HiSes-LoRes-MidPeer	3.42 (1.86)	1.522 (1.36)	.054	-.122
LoSes-HiRes-LoPeer	-10.20 (1.83)		-.177	
Tracking		-1.18 (.33)		1.972
(Experience) (Test)	.165 (.048)	.134 (.036)	1.652	.106
(Test) (Race Preference)	-.116 (.091)	.0252 (.015)	-.336	.483
(Experience) (Possessions)		.0798 (.013)		

from that estimate. In this sense we can see that the simple linear form may give average estimates of parameters.

I have selected two schools with near the lowest and near the highest Experience levels with which to investigate the range estimates. Thus the values which I will give below are neither the highest nor lowest values obtainable, to be conservative in the face of a sample and sampling process which leaves much to be desired. Since my school coding is arbitrary and unrelated to the original EEOS code (which, even so, does not identify schools by name or location), I will use my code numbers for this exposition. School #79, with average Teacher Experience of 3.7 years, and school #86, with average Teacher Experience of 17.5 years, will be the demonstration cases throughout this section.

School #79 would respond negatively, though hardly at all, to a small change in average Teacher Test Score. Its derivative is -.10, indicating that a 1 point gain in Teacher Test Score would produce a .1 loss in pupil Verbal Score, if all other values stayed the same. For example this assumes that Teacher Preference for White Students, which is about at the mean of schools, would be unchanged. If Preference for Whites increased when higher Test Score teachers were hired (these two factors are correlated greater than .4), then the estimated decline in Pupil Score on account of Teacher Test Score would be larger, assuming these teachers with higher Test Score and greater Preference for Whites were as equally Experienced as current school #79 teachers. Of course the direct effect of higher preference is positive. The correlation between Test Score and Experience is lower than that between Test Score and Race Preference (being less than .2), so random selection of teachers on the basis of their Test Scores might seem to indicate a negative effect on pupils in this school (and with regard to the output "Verbal Score," though we shall find a negative relationship between Reading and Test for this school, also).

This is not so, as one can easily see by picturing the distribution of teachers with respect to Experience. Picking "randomly" from a pool of teachers with Test Scores higher than that in school #79--which has an average Score higher than the mean school in Eastmet or, for that matter, the nation[26]--one is picking from teachers also with higher than mean Experience. This is true because of the correlation between Experience and Test Score. Since the school in question is well below the mean in Experience, then the actual net effect of picking teachers with higher than average Test Scores, but otherwise randomly, will be to increment pupil test score.

The point of this discussion is to make clear the limited meaning of the "partial derivative" in assessing policy. It would take a very non-random selection of teachers to actually produce a decline in Pupil Verbal Score: all characteristics but one, Test Score, would have to remain constant. And since these characteristics are extreme values to begin with, there is a natural tendency for them to become less extreme with random selection. If, on the other hand, School 79 is a school which always gets inexperienced

[26] Coleman, et al. (7) do not actually give the mean teacher Test Score, but the score of the teacher of the average pupil, in Table 2.33.1, page 131. My School #79 has a higher average than that facing any group of children listed in this table.

teachers, i.e., teacher selection is not random, a deliberate policy to increase experience will prove beneficial, even if average Test Score would decline.[27] We can see this by taking the derivative with respect to Experience:

$$\frac{\partial Y}{\partial \text{ Experience}} = -.3 \cdot 53 + .165 \text{ (Teacher Test)}$$

At the average Teacher Test for school #79, this value is +.64, or indicating over six times the gain in Verbal Score for a year's experience than the loss in Verbal Score from a point of Teacher Test score.

Once again, however, we should take account of the fact that a <u>randomly</u> selected teacher with higher experience will still have a lower test score than the average teacher in School #79.[28] How low would it have to be to <u>decrease</u> the Verbal Score of pupils? Where the derivative is a function of one variable only, as in this case, we can solve for the level of that variable at which the derivative changes sign (in those cases in which it does). Setting $\frac{\partial Y}{\partial \text{ Exp}} = 0$ and solving the equation immediately above, we find that Test Score must fall below 21.4 before this expression becomes less than zero. Since this figure is below the average Test Score for this sample (or the nation), and since I am sampling only teachers above a minimum level of experience, it would require

---

[27] This discussion, of course, assumes that I have estimated a real production relationship. From preceding comments on Experience above, it should be clear that I give little credence to the estimated relationship between Experience and outputs being a production relationship. This discussion is proceeding, however, to demonstrate the use of production estimates if we had any.

[28] This effect is actually a function of the distributions of the two factors. We are selecting randomly from teachers with more than 3.7 years experience, and asking whether, on the average, these teachers have lower than 25.3, the current School #79 mean, on the Teacher Test.

a negative correlation between Experience and score to produce an expected Test Score low enough to reduce output. As we know, there is a slight positive correlation between Experience and Test Score. So one could randomly select teachers on the criterion that they be more experienced than those in School #79, with confidence that this will improve the output of that school. Furthermore, such a selection will induce a higher payoff to the already high Test Score present in that school, as seen in the derivative of Verbal Score (of students) with respect to Test Score (of teachers).

Note that all this makes some sense with regard to how schools might actually work. School #79, this discussion indicates, could profit greatly from a selection procedure which brought some experience to the school, even at the loss of some Test Score. If one selected randomly among applicants with high Test Score, there might be some improvement also, but this is just because that Test Score is likely to be associated with Experience. If one selected nonrandomly, for teachers like those in this school but with higher Test Score, the improvement would be small or even negative. The school might be characterized as having far above average teachers in terms of their talents, but far below average teachers in terms of their abilities to put that talent to use in producing school output. A couple of Experienced teachers, even if not as capable of scoring well on tests, could direct the talents of the inexperienced teachers. There is, in other words, a real interaction between experience and talent, which corresponds to the equation's interaction between Experience and Test Score. Although on the average the school system would use these equations to look for teachers with higher Score, in this particular case, it should look for teachers with greater Experience.

I should note here that the interactions as defined are the product of school means, not of individual teachers. In fact, the average of teacher interactions will equal the interaction of teacher averages only if there is no correlation between teacher attributes within schools.[29] Although it is important to consider the real-world conditions that might produce the significant interaction coefficients, it is facile and not altogether justifiable to consider them as indicating interaction between different teachers. Without considering the characteristics of the different teachers, there is no way to know if Experience and Test Score interact in that one teacher with both characteristics is a super teacher, or because two teachers, each with one characteristic, complement each other. To determine this difference, I would have had to go back to the original teacher data, and re-aggregate by school, taking interactions for each teacher, and averaging. When the variables are positively correlated, the average of individual interactions is larger than the interaction of the means; and conversely when negatively correlated, it is smaller. I would expect that in most schools the values of Test Score and Experience would be positively correlated. I see no reason to expect that they would be positively correlated between schools, but negatively correlated within. However this might not be the case in all schools, and it might be the case to varying degrees. Thus the average of individual interactions would not be a linear transform of the interaction of averages, and could be expected to produce different results. For the purposes of this paper, I felt this point could not be investigated further. It does bear keeping in mind, however,

[29] This is a well-known probability theorem that the expected value of a product is the product of the expected values only if the elements in the product are uncorrelated.

when it comes to interpreting the results.

We return now to the derivative of the Verbal equation with respect to Test Score, and consider a school with high Experience, #86. Despite a higher Preference for whites, and consequent negative effect, the rate of increase in Teacher Test Score is 1.93. Thus School #86, which has an above mean Test Score, could gain a great boost in output from selection of high Test Score teachers. This selection, if random, will reduce the average experience level of the school (which was selected as having a high experience level), and thus reduce the incremental effect of further increases in Test Score. Nonetheless, the Test Score in this school is lower than in School #79. While the Experienced but lower scoring teachers should be assigned to School #79, the higher Scoring (if less Experienced) new teachers should be assigned to School #86.

This discussion has abstracted from price considerations. As Part I of this paper suggested, the economic efficiency criterion needs prices for a solution. If a year of Experience costs more or less than a point of Test Score, then the simple suggestions made above are not strictly relevant. The method by which prices are accounted for was outlined in Part I. However, to the extent that general productivity and cost considerations have led to hiring those qualities in teachers which are most cost effective, and there is now a pool of new teachers to be assigned, the considerations above can proceed without reference to prices. They are policies of resource allocation <u>given resources</u>. This, we have seen, is the concept of Technical Efficiency. It assumes proper management in Schools #79 and #86, after the policy, and in all schools (for estimation purposes) before the policy is enacted. It is one use to which a good production function could be put. In addition, of course, the derivatives offered here could be combined with price data--which might also be different for different schools--to determine the most effective resource mix for a given budget. If either the prices facing the schools or their output reactions are different, then different resource allocation decisions apply to different schools.

The derivative with respect to Experience became negative at a value of Test Score not far below the mean. In fact, nine of the thirty schools take on a negative value of this derivative, i.e., would lose output given more experience, and nothing else. At the means of the variables, the rate of change of Verbal Score with respect to Experience is .36. Once again, this figure is close to that of the coefficient in the simple linear regression, .49. In Table IV-4 I have calculated the derivatives from the linear function at the mean values of variables, and indicated how far (in terms of standard errors of the coefficients from the simple linear equation) this estimate is from the regression coefficient in the simple linear form. In all cases where school resource variables interact with other resource variables, the difference from the linear estimate is insignificant. Why the interaction with a background variable[30] should so differently affect the resource estimate I do not know; but this issue will be discussed below.

Before leaving the Verbal Score equation, we might look at its last interaction derivative, that with respect to Race Preference. The equation for this derivative appears in Table IV-4. Calculating where it turns negative, we get a Test Score of 36.5--on a thirty question test! In other words,

[30]The interaction between Experience and Possessions in the Reading Equation.

TABLE IV-4

Partial Derivatives from Interaction Equations

Equation	Partial	Evaluated at Means	Comparison with partial from simple linear equations (within--standard errors)
Verbal			
$\frac{\partial Y}{\partial Test}$	= .165 (Exp) - .116 (Race Pref)	1.10	.56
$\frac{\partial Y}{\partial Exp}$	= -3.53 + .165 (Test)	.36	1.7
$\frac{\partial Y}{\partial RacePref}$	+ 4.23 - .116 (Test)	1.57	.49
Reading			
$\frac{\partial Y}{\partial Test}$	= -1.30 + .134 (Exp) + .0252 (Race Pref)	.29	.55
$\frac{\partial Y}{\partial Exp}$	= -3.62 + .134 (Test) + .0798 (Posses)	.01	5.17
$\frac{\partial Y}{\partial Race\ Pref}$	= .0252 (Exp)	.27	n.c.
$\frac{\partial Y}{\partial Posses}$	= .918 + .0798 (Exp)	1.78	1.09

there is no really negative effect induced in the return to output from additional Race Preference, as Test Scores are higher. There is a non-linear relationship such that at higher Test Scores the effect of additional Preference for Whites is diminished; but an increment of the Preference never results in a lower score. At the maximum possible value of Test Score, 30, the return to an additional point of Race Preference is .75 of a point of pupil Verbal Score. At the mean Test Score, the return (remember, these are white children) to additional White Preference is approximately 1.57 points of Verbal Score, insignificantly different from the simple linear estimate.

In terms of comparing equations, note first that the coefficients of the two "interaction" terms which had been entered, because of their form, into the original simple linear equations, are hardly affected by these multiplicative interactions. Of the other variables not involved in the interaction terms, only People in the Home was particularly affected. The coefficient in the interaction equation is not significantly different from zero at the generally accepted 5 percent level. This variable had originally appeared in the Reading equation also, with non-weighted estimation.[31] With estimation by weighted regression this variable became insignificant in relationship to Reading, and now with interaction its importance with respect to Verbal Score is diminished. Given the stability of the effects of the other background variables, one is inclined to ask just what People In The Home is measuring. It is one of the many questions to which I have no answer, however.

The derivatives from the Reading equation can be calculated, and have been presented for those variables with interaction in Table IV-4, above. Once

[31] See Michelson (16), Table 1.

again the partial derivative of the output equation with respect to Teacher Test is negative for School #79 (- .66), positive for School #86 (1.25). It is positive at mean values, and very close to the value of the Test coefficient in the simple linear equation. Race Preference now can be significantly entered into the Reading equation. There is no coefficient to compare it to in the reading equation.[32]

The introduction of an interaction between a school variable and a background variable has more policy significance than might at first be apparent. The previous interactions had refined the notion of a derivative, so that the best policy for a particular school depended on the level of resources already at that school. But there was no mention of requiring a different mix of resources for each school depending on the kinds of children in that school. Yet just as teachers interact with each other, so that one's experience might add to another's talent, so teachers and children interact. It might be that some resources are particularly appropriate to some children, other resources to other children. In a previous paper I have developed this idea, calling it "Resource Specificity." The point I am making here is a further development of that concept.[33]

"Specificity" means that different children react to different resources differently. In general this can be tested by asking whether the same

[32] I did specify a simple linear Reading equation with Race Preference to determine a coefficient: It was .55, and the two derivatives diverge by less than one standard error. If a researcher had a theory about instruction which included Race Preference as an important variable, then he would not pay such obeisance to statistical significance in determining the specification of his equations.

[33] See Michelson (16).

production function describes education for different types of children. These may be by ethnic group, urban-rural background, language in the home, etc. If some grouping produces a different relationship between resources and output, then different resources are specific to these groups.

Specificity implies that an interaction occurs between the characteristics which define the group and all school variables. One is best off simply saying that their production relationships are different. In the present case, I am claiming a much more limited interaction. There is, for example, no conflict in sign: Experience is not a positive resource for some children, a negative resource for others. But there could be some difference between schools which would make more Experienced teachers more effective in some than in others.

If this were a production function, the implication would be that the more experienced teachers function better in interaction with higher social class students. However, notice that the derivative of Experience, at the means, is essentially zero, whereas the derivative with respect to Possessions has increased close to two standard errors (with respect to the simple linear equation). This might be used to strengthen my previous argument that Experience was really a social class variable. On the other hand, two additional factors should be noted here. The first is that in this interaction, unlike the others, there is within school variation. Each pupil has a Possessions index, whereas each child in a school has the same Experience measure. Though this is true, I do not see why this argues that Possessions should take away the effect of Experience in interaction. The second factor is that, given the units of coding, the values for Experience were generally two to three, and at times more than four times the values entered for Possessions. This does strain the task of the single interaction coefficient, to mean the same thing for an increase in one unit of each variable. On the other hand, the difference between Test Score and Race Preference was as severe, without such ill effects.

I have, then, no strong explanation for the difference between this interaction and the others, in terms of estimating, at the mean values, the simple linear coefficient for Experience. However the important point is made that part of the interaction investigated here is among resources themselves, and part is between resources and children. Given this latter interaction, there is no reason to think that all schools within a district should have the same resources. A good production function estimation would help determine which resources are best employed where.

Non-Linear Transformations

Some non-linear forms can be brought within the estimation capabilities of linear regression by transformations of the data. Two such transformations have been discussed, and will be presented in this section: parabolic (i.e., second degree polynomial) and logarithmic (or multiplicative). To justify one form or another, one ought to discuss the type of error assumed, though this is seldom done. For example, the multiplicative form assumes that error also has a multiplicative effect. Error is otherwise assumed additive. I have no theoretical basis for assuming error is additive or multiplicative and, like my predecessors, will say no more about it.

The equations utilizing squared terms appear in Table IV-5, and their partial derivatives in Table IV-6. It can be seen that adding one term in the Verbal equation, three terms in the Reading equation, raises R^2, though not by much. The effects of the background and control variables remain fairly much as they were. The effect of Teacher Test is raised in the Verbal equation. As in the interaction equation, it never gets negative, despite the presence of a negative term in the expression of the partial derivative. At 30 questions correct, the derivative is still +.60. The overall effect of this transformation is to raise the estimated effect of an increase of one point of Teacher Test Score for most schools.

Teacher Test does not have this property with respect to the Reading output. At 30 correct questions, the rate of change of Reading Score with increments to Teacher Test Score is -.90. There is no unique Beta weight given this transformation, because of course the effect of a standard deviation

TABLE IV-5

Weighted Regressions with Non-Linear Transformations

Eastmet City Whites

	Regression Coefficients		Beta Weights	
	Verbal	Reading	Verbal	Reading
Sex	2.43 (.70)	3.19 (.48)	.102	.195
Age 12+	-8.93 (1.18)	-4.12 (.86)	-.209	-.141
People in the Home	-.227 (.13)	--	-.057	--
Possessions	1.15 (.17)	.427 (.52)	.235	.127
Father's Education	.595 (.14)	.164 (.10)	.128	.052
Kindergarten	3.54 (.80)		.142	
Teacher Test	3.36 (1.54)	3.05 (1.22)	.677	.899
Teacher Experience	.545 (.075)	-.534 (.22)	.210	-.301
Hi-Ses-LoRes-MidPeer	6.84 (1.91)	3.56 (1.38)	.109	.083
LoSes-HiRes-LoPeer	-12.90 (1.79)			
Tracking		-2.26 (.31)	-.224	-.234
$(\text{Teacher Test})^2$	-.046 (.035)	-.066 (.027)	-.411	-.857
$(\text{Teacher Experience})^2$		.038 (.010)		.526
$(\text{Possessions})^2$		.107 (.046)		.365
Constant	-50.45	-28.14		
R^2	.603	.546		

TABLE IV-6

Partial Derivatives of Non-Linear Equations

		Evaluated at Means
Verbal		
$\frac{\partial Y}{\partial \text{Test}}$	$= 3.362 - .092\ (\text{Test})$	1.25
Reading		
$\frac{\partial Y}{\partial \text{Test}}$	$= 3.054 - .132\ (\text{Test})$	.025
$\frac{\partial Y}{\partial \text{Experience}}$	$= -.534 + .076\ (\text{Experience})$	.28
$\frac{\partial Y}{\partial \text{Possessions}}$	$= .427 + .214\ (\text{Possessions})$	1.92

change in Test Score now depends on what the Test Score was to begin with. From the mean Score, essentially 23 right, with a standard deviation of 1.8 questions, the effect on Reading Score of a standard deviation increase in Test Score is a .044 standard deviation increase in Reading Score. This is far smaller than the Beta weight from the simple linear regression. The squared term here is clearly compensating for the truncated score, the fact that no score over 30 was possible.

The partial derivative with respect to Experience has a disappointing form, which remains even in the next section, where non-linear transformations and interactions are combined. I expected that the constant term would be positive, and the level-dependent term negative, implying a decreasing relationship between Experience and output. The opposite signs occur. Below slightly over seven years of Experience the relationship between Reading Score and Experience is negative. Six schools have such low Experience levels, but this group does not include the two schools with the lowest Reading Scores.

The results from using $(\text{Possessions})^2$, however, are quite reasonable. The relationship between output and background would not be linear, it seems to me, in any carefully thought out model of education. Part of this is due to the cumulative advantages of spending six years in a home where cognitive skills are stressed and practiced before the schooling production even begins. The whole concept of accumulation of skills is interactive, not additive. Consider, as a simple example, the use of the word "not," or the whole concept of negation. This does not add one word to the vocabulary, but multiplies all other words and concepts by two. An adjective is not just a word, but is as many phrases as nouns it goes with; and adverbs count as many times as they go with adjectives. In fact, the coefficient of Possessions alone is no longer statistically significant. The effect of Possessions on Reading seems to be dominantly multiplicative, even if all it can multiply in this form is itself. It is clearly not multiplicative in relationship to Verbal Score.

Additionally, the effect of background can be expected to be interactive with school resources, and possibly even with the average background level in the school.[34]. Thus another part of the non-linearity of the relationship between background and output should be due to production in the production unit (school). A last part, needless to say, occurs during the production process but not at school.

In Table IV-7 a Reading equation is presented which includes logarithmic transformations. Untransformed, the coefficients in that table refer to the following equation:

$$Y = (.165)S^{.30}e^{-.36A}P^{.56}H^{.24}F^{.06}T^{.83}E^{.17}e^{.29I}e^{-.08K}$$

where Y is Reading Score output
S is Sex
A is Age 12+
P is Possessions
H is People in the Home
F is Father's Education
T is Teacher Test Score
E is Teacher's Experience
I is High SES, Low Resources Mid Peer Interaction Measure
K is Tracking

---

[34] This is a crude way of analyzing "peer effect," however. Without identifying with whom a pupil is placed in classes, and with whom he associates, it seems feeble to attempt to identify the influence of other children on him. If there are enough children like him, he may associate with them and be influenced by them regardless of the mean level of the school.

The constant is the anti-log of -1.801. The three variables which appear as exponents work in the following way: When they have the value zero, the expression can be written without them (since $e^0 = 1$). When they have another value (which can only be 1 for age or interaction, 1 or 2 for tracking), the rest of the expression is multiplied by a constant. They shift the multiplicative constant by a fixed amount. A child age 12+, for example, is given the value 1 for the age variable, and therefore $e^{-.36}$ for the multiplicative shift. This is .698. The equation constant then effectively becomes (.698) (.165) = .115. This implies that, all other things equal, the child 12 years old or older in the sixth grade is scoring at approximately 70 percent of the level of other children. It also must mean that any increase in resources will be 30 percent less effective for this child than for other children.

To see this more clearly, consider the derivatives of this equation, as given in Table IV-8. The derivative, as explained there, retains the constant as multiplier. If we consider the exponential terms as shifting this constant, then they do so for the derivatives as well as for the equation It is then a part of this functional form that anything which reduces the equational value by a constant percentage, as does the age-shift variable, also reduces the incremental effect of any other variable by the same amount.

By the same argument, the effect of having the specified Interaction increases the equation (and effect of increases in any other variable) by 33.6 percent. The effect of Tracking (as opposed to no Tracking) is to reduce the equation by 17.3 percent from the value it would otherwise have. Tracking, under this equational form, reduces the effectiveness of increases in other school resources also by 17.3 percent.

TABLE IV-7

Logarithmic Transformation

Eastmet City Whites

	Reading	
	Regression Coefficients	Beta Weights
Log Sex	-.359 (.063)	.185
Age 12+	.303 (.052)	-.176
Log Possessions	.560 (.032)	.544
Log People in Home	.241 (.038)	.214
Log Father's Education	.057 (.061)	.030
Log Teacher Test	.833 (.183)	.164
Log Experience	.169 (.036)	.145
HiSes-LoRes-MidPeer	.285 (.099)	.095
Tracking	-.079 (.020)	-.118
Constant	-1.801	
R^2	.506	

TABLE IV-8

Partial Derivatives of Logarithmic Equation

	Evaluated at Means
Reading	
$\frac{\partial Y}{\partial \text{Test}} = .0363\ \bar{R}$	.64
$\frac{\partial Y}{\partial \text{Experience}} = .00736\ \bar{R}$	.13

Note: Where

$$Y = aX_1^{b_1}X_2^{b_2}$$

$$\frac{\partial Y}{\partial X_1} = b_1 a X_1^{b_1 - 1} X_2^{b_2}$$

(where X_2 is constant)

$$= \left(\frac{b_1}{X_1}\right) \quad aX_1^{b_1}X_2^{b_2}$$

$$= \frac{b_1}{X_1} \cdot Y$$

when Y is evaluated from values of X_1 and X_2.

At the means, $Y = a\bar{X}_1^{b}\ \bar{X}_2^{b}$. Hence the expressions for the derivative above, which are $\frac{b_i}{\bar{X}_i} \cdot \bar{R}$.

The other coefficients, which are written as variable exponents in the form above, can be interpreted as output elasticities. As noted in Part II, they are here assumed constant. I will compare these elasticities with those calculated from the simple linear equations just below. First, let us look at the value of the derivative at the mean. From Table IV-8 we see that this form leads to low estimates of the derivative.[35] The mean values through which the regression fit must pass are the <u>geometric</u> means of the variables (the means of the logarithms of the variables), whereas arithmetic means are used for the table and are appropriate for the linear form. This makes comparisons at the same values difficult. However, since the geometric mean is lower than the arithmetic mean,[36] the derivative figures would be even lower. Taking the maximum Teacher Test Score, 30, the derivative (1.09) is about equal to the simple linear estimate.

The output elasticity can be easily calculated for the simple linear form. Recall that

$$Y = a + b_1X_1 + \ \ldots\ + b_iX_i + \ \ldots\ ,$$

is the familiar linear form, and

$$\phi_i = \frac{dY/Y}{dX_i/X_i}$$

is the elasticity formula from Part II. Then a one unit change in X_i ($dX_i = 1$)

---

[35] Of course the preceding second degree polynomial did have an even lower partial derivative with respect to Teacher Test.

[36] Averaging over schools, the arithmetic mean Reading Score is 17.62. The geometric mean is 16.45.

causes a b_i change in Y ($dY = b_i$). This formula therefore reduces to

$$\phi_i = \frac{b_i \bar{X}_i}{\bar{R}}$$

where the bar denotes the mean, and R is Reading Score. The values from the simple linear equation are

$$\phi_T = .45$$

$$\phi_E = .20,$$

using the symbols from above. The value of the elasticity with respect to Experience is close to the coefficient from the logarithmic form, but the Test Score elasticity is not close. Using the standard errors from the logarithmic equation, ϕ_E is within one standard error of the constant elasticity estimate, but ϕ_T is over two standard errors low. I would not want to conclude, then, that the simple linear form and the logarithmic form reach similar average estimates of statistics. I would conclude that in general the linear form with interactions or non-linear transformations allows a great deal of flexibility compared with the multiplicative form.[37] Therefore I will present one last set of equations combining these two features.

---

[37]Bowles (1), for reasons which I cannot fathom, chooses the multiplicative over the simple linear form, not considering the linear form with interactions. He does note some of the disadvantages of the multiplicative form, particularly that the sign of the interaction between two inputs (their cross-partial derivative) is determinate, given the signs of the first partial derivatives. He does not go on to explore the linear form with interaction, in which the sign of the interaction is not constrained.

Final Level-Dependent Equations

I have called the general form of an equation in which at least some partial derivatives depend on the levels of some variables, "level-dependent" forms. In the preceding sections I argued that, considering a very limited set of forms, the combination of high order transformations (squares or higher powers of one variable) and interactions in a linear regression format was convenient compared with the multiplicative form. Many more complex considerations have been ignored. Of these, the most obvious is one in which the elasticity of substitution between factors is held constant. This form has become popular in economics, though at aggregate levels (that is, for a mix of products, not for an industry production function). It is not immediately apparent that this would be a good restriction for an educational production function, but at least this demonstrates the direction in which future experimental and theoretical work should look before we will be ready to estimate such functions.

I have estimated general "level-dependent" equations for the Reading and Verbal outputs. These appear in Table IV-9. In the Reading equation, Possessions and $(\text{Possessions})^2$ both appear. In the Verbal equation, no background variable appears to a power greater than 1. $(\text{Teacher Test})^2$ appears in both equations, and in both it has a negative sign. As explained above, this is reasonable, given the truncated Test scores. In addition there is no reason to believe that any relevant talent for teaching which might be measured by such a Test is linearly correlated with that Test Score. Higher scores on the Test might indicate more talent, but not necessarily in equal increments.

The Teacher Test has been interacted with Possessions, as each repre-

TABLE IV-9

Combined Level-Dependent Equations
Eastmet City Whites

	Regression Estimates		Beta Weights	
	Verbal	Reading	Verbal	Reading
Sex	2.675 (.67)	3.393 (.470)	.114	.112
Age 12+	-9.508 (1.14)	-4.413 (.86)	-.223	-.151
Possessions	-10.422 (1.97)	-4.958 (1.40)	-2.124	-1.477
$(\text{Possessions})^2$		.142 (.045)		.484
People in the Home	-.137 (.120)		-.035	
Father's Education	.485 (.13)	.140 (.098)	.104	.044
Kindergarten	4.246 (.77)		.170	
Teacher Test	4.970 (1.52)	1.564 (1.20)	1.002	.460
$(\text{Teacher Test})^2$	-.207 (.039)	-.0979 (.028)	-1.830	-1.268
Teacher Experience	-2.627 (1.102)	-3.403 (.83)	-1.014	-1.920
Tracking		-1.735 (.29)		-.180
HiSes-LoRes-MidPeer	5.355 (1.85)	2.794 (1.37)	.085	.065
LoSes-HiRes-LoPeer	-3.934 (2.044)		-.068	
Experience - Test	.125 (.045)	.151 (.034)	1.253	2.223
Test - Possessions	.506 (.086)	.221 (.06)	2.636	1.681
Test - Race Preference	.0608 (.016)		.176	
Constant	-8.663	20.534		
R^2	.643	.565		

sents the most powerful indicator of quality: Test, of school quality, and Possessions, of background. The coefficient seems large, significant at the .1 percent level in both equations, and positive in both equations. If one believes these measures and the multiplicative interaction, then it seems that children who come from high class homes and go to high resource schools do better than the sum of the high class and the high resource effects. Since this already presumably corrects for the triple interactions of high class, low resources and mid peers, plus low class, high resources and low peers, which work in complementary directions, then the home-school interaction effect is truly spectacular.

The other interactions remain pretty much as we found them before. Experience and Test interact positively, indicating either interaction between Experienced and high Test teachers, or that those teachers with both qualities are super teachers. Given Test Score, the teacher who Prefers Whites is associated with higher Verbal Score whites, though considering that the mean score is for approximately (interpreting liberally) 60 percent white, this might indicate teachers who want the security of a dominantly white school, but don't necessarily prefer to teach white children.[38]

Some partial derivatives for these equations are given in Table IV-10, along with their values at the means. The difference from the simple linear coefficients is also calculated and given in the last column. These differences

[38] The question allowed for "all white," "mostly white," "half and half," etc. "Mostly white" was coded as 9, with all white as 10, mostly nonwhite as 1, all nonwhite as 0. I presume that on an individual teacher basis there are more 9 answers than 1, so that had there responses been coded as 8 and 2, the mean score (5.8) would have been lower. Thus translating 5.8 to percentages is an exceedingly loose interpretation of this variable.

TABLE IV-10

Partial Derivatives for Level-Dependent Equations
Eastmet City Whites

		Evaluated at Means	Standard Errors Different from Simple Linear Coefficient*
Verbal			
$\frac{\partial Y}{\partial \text{Test}}$	$= 4.970 - .414T + .125E + .506P + .061R$	.703	-1.6
$\frac{\partial Y}{\partial \text{Experience}}$	$= -2.627 + .125T$	.242	-3.3
$\frac{\partial Y}{\partial \text{Possessions}}$	$= -10.422 + .506T$	1.191	+ .8
Reading			
$\frac{\partial Y}{\partial \text{Test}}$	$= 1.564 - .196T + .151E + .221P$	.234	-1.0
$\frac{\partial Y}{\partial \text{Experience}}$	$= -3.403 + .151T$	.062	-4.4
$\frac{\partial Y}{\partial \text{Possessions}}$	$= -4.958 + .284P + .221T$	2.099	+4.0

*Signs indicate: + higher than simple linear estimate
- lower than simple linear estimate

are large, and consistently increase the estimated effect of background, decrease the estimated effect of school resources, as compared with the simple linear equation. The Teacher Test derivatives are negative at 30 questions, holding the other variables constant at their means. In fact, at these mean levels, the Teacher Test derivative turns negative at 24.6 questions for Verbal Score, and 24.1 questions for Reading Score. Can it be that in a school with mean values of Experience, Race Preference and Possessions, adding a teacher with a high Test Score but otherwise average values will reduce the output of the school? Or does it mean, an interpretation I much prefer, that we are still far from estimating a production function?

The R^2 from these equations are not spectacularly higher than those from the simple linear equations. Each R^2 increased by less than 10 percent of its initial value. Thus the advantage to such complicated forms would seem to lie in presenting a better picture, if they do, of the way in which the inputs are related to the outputs. They do ot markedly increase our ability to explain variations in output with these inputs. As noted earlier, I find this inability to explain much more variance quite comforting: I would not like to live in a world much more determinate than this.

One last point should be discussed here. A Beta weight greater than 1 is a rare phenomenon in social data. To think that a standard deviation increase in an input could produce more than a standard deviation in an output seems extreme. One might look fo, an explanation in the relative invariance of the output measure, so that a couple of rare points create this effect. Or one might look for collinearity in the input measures, again an argument that the relationship is spurious, a result of the non-representative distribution of error.

Six variables out of 14 in the Verbal equation, and five out of 12 in the Reading equation have Beta's greater than 1. However, one need not look for spurious statistical artifacts to explain them. The problem is that with transformations and interactions the Beta weight for one coefficient simply has no meaning. The Beta's for those variables not involved in these manipulations are nicely behaved (i.e., small). There are no unique Beta's for the other variables, as explained above. Therefore I have calculated effective Beta weights, at the means, for the three variables whose derivatives appear in Table IV-10. These appear in Table IV-11. These are approximate values, calculated from the partial derivatives.

All the extreme Beta values are explained by this simple calculation. The resultant Beta weights are far smaller than the Beta's from the simple linear equations for the school variables, and greater than the simple linear Beta's for the background variable. Thus if Beta's are used to measure relative "importance," school variables are even less important, relative to background variables, than predicted by the highly averaged simple linear equations. As always, whether this means that policies affecting the home environment are more cost-effective than policies affecting schools cannot be determined from this result. But to the extent that these equations are considered more accurate than the simple linear equations, they are also more depressing in terms of leading to effective school policies for increased cognitive skills.

Along the Great Frontier

One of the tasks promised for this paper was an attempt to locate schools

TABLE IV-11

Approximate Beta Weights for
Level-Dependent Equations

	Verbal	Reading
Teacher Test	.142	.069
Experience.	.093	.035
Possessions	.242	.624

Note: Derivatives are taken from Table IV-10, and therefore for Teacher Test refer to the interval from one-half standard deviation below the mean to one-half standard deviation above the mean.

at the production frontier, and re-esti ate production relationships with these schools. No pretense was made that this experiment would be a success, and I believe it has lived up to its expectations. Fronti . schools were defined by dividing schools into four regions by their average Test Scores: more than one standard deviation below the mean, within one standard deviation below the mean, and similarly for above. Within these categories the schools with the greatest positive residual

$$R - \hat{R},$$

from the simple linear regression were selected as Frontier schools. Eighteen schools had positive residuals, and of these, seven were chosen. Sixty-seven children were involved, and the sample was re-weighted to give each school equal weight, and each child equal weight within each school.

The resulting estimated equation appears in Table IV-12. I only used the simple linear form for this experiment. All variables except Sex had coefficients larger than their standard errors. What would one expect from Frontier observations? The basic expectation is that the school variables will have larger regression coefficients. I would have no expectation about background variables, because I am not selecting for home production, but for the most productive schools, controlling for background.

The Teacher Test coefficient is indeed larger than it was in the original equation. On the other hand, Teacher Experience and Tracking have reversed signs. Besides these two and the insignificant Sex coefficient, <u>all</u> the other coefficients have increased. There is nothing special, therefore, about the Test c fficient. The Beta weights should be ignored in a sample which

TABLE IV-12

Reading Equation for Frontier Schools

	Regression Coefficient	Beta Weight
Sex	.587 (1.54)	.038
Age 12+	-9.344 (2.50)	-.416
Possessions	2.172 (.34)	.593
Father's Education	.507 (.37)	.153
Teacher Test	1.043 (.28)	.464
Teacher Experience	-1.862 (.44)	-.560
HiSes-LoResources-MidPeer	7.167 (3.25)	.251
Tracking	1.64 (.97)	.221
Constant	-7.100	
R^2	.583	

promises nothing in terms of representativeness.

I would conclude that the experiment is totally inconclusive. The students in these schools seem more responsive to just about any change, which is an appealing characteristic which well managed schools should disclose. Few researchers will be convinced that a regression on 67 children better represents production conditions (when well-managed) in Eastment than the 974 white sixth grade children in the original city sample, or the 597 of these who remained in one school from the first grade; or that seven schools represent production characteristics better than the 36 or the original sample. I want to stress that, in the long run, these opinions about maintaining sample size will be proved wrong. Some small sub-sample of schools and children will better demonstrate the technical possibilities (within a given bureaucratic structure) than the entire sample. It seems safe to say, however, that this particular sub-sample has not done so.[39]

Lessons from these Experiments

The first thing to note is that the functional form is important in determining the effectiveness of different policies. To judge that one resource is educationally helpful on the basis of a simple linear regression might neglect important interactions which could render that resource ineffective in some situations. Though the use of non-linear transformations was instructive, probably the most important information in the preceding sections came from the

[39] As a post-script I will note that the explanation here, that these figures diverge too much from those found before, seems exceptionally weak. What needs to be done is to set out the criteria under which one would accept the sub-sample results as in fact the frontier relationships. These criteria should probably be independent of comparisons with results from the entire sample. I have not yet undertaken this task, and I consider it a formidable one.

interaction terms. Teachers interact. Within a school, one might consider the mix of teachers in making new assignments. In fact, judicial enforcement of racially balanced faculties is a step in this direction: assignment of one teacher is dependent on the characteristics of the teachers already at that school. Although I do not know the educational effect of such assignment criteria, at least they open the door for other resource-interaction considerations.

In addition, resources interact with backgrounds: Teacher assignment should consider the types of children in the school. This is quite an important result. If one were to believe, as many of the educational skeptics try to read from the EEOS study, that schools contribute nothing to cognitive skills, then one would believe that wealthy suburbanites were foolish in their high per-pupil expenditures. If, on the other hand, the marginal productivity of a dollar expended in a high class neighborhood, in terms of producing cognitive skills, is greater than in a lower class neighborhood, then it appears that these people are particularly sensible. That is what this interaction term implies. It follows everyday observation: that schools in lower class neighborhoods expend a great deal of effort in behavioral outputs--under the general heading of "discipline." It follows from a radical analysis which sees schools as places of socialization first, cognitive achievement last: upper class children are already socialized. This interaction defines the most important challenge to the educational establishment today: from "cost-benefit" and other de-humanized (but "rational") approaches, an educational production function with a home-school interaction term like this one will dictate putting money where the pupils are most prepared to use it, where they by and large already have a great deal. This would not be true only if prices worked the other way--if resources were more cheaply supplied where they are more scarce. But everything we know about prices indicates the opposite, that resources (such as teacher talent) prefer to go where there is already a fund of that resource, and where the students are already prepared to take advantage of it.[40] The challenge, then, is to use a human calculus in allocating resources, not a monetary calculus: resources should go where they are scarce, even though their marginal effectiveness will be small there.

The practical effect of this interaction is to help explain muckraking studies which show that funds spent on compensatory education are not very productive. Most of the argument against such views has been--rightly--that these funds have been so badly administered that they cannot be said to have gone for real compensatory education. Furthermore, they have been diluted, spread over too many children. The home-school interaction term indicates that a small amount of compensatory program actually reaching a low-income child can be expected to have little effect. That same amount reaching a high-income child could have a larger effect. Consider the obvious impact of financing a science fair in an otherwise well-off school district, compared with the increment to science knowledge from the same amount of funds spent in science education in a low-income school. The pre-conditions for effectively absorbing such funds are so different, that of course the measured benefit from the former project will exceed that from the latter. Yet different children are

[40] For evidence on the prices of teachers with different amounts of Test Score, holding constant other factors, see Levin (13). For a model which determines that teachers of higher quality supply themselves to suburban schools, see Owen (19). Martin Katzman, in a personal communication, reports results of a regression of percent teachers with Masters (or Masters and Bachelors, in another regression) against Masters salary and social class. "Percent Masters was highly responsive (negatively) to social class, but unresponsive to salary."

involved in these two measures of benefit. They are not strictly comparable. A "human calculus" would weight higher the small gain of deprived children, than the large gain of the already privileged children.

On the other hand, the final equation did produce results which showed the schools even less effective in producing the cognitive skills measured than we would believe from the simple linear equation. The answer may lie in non-school programs,, such as day-care centers for young children, extensions of the Police Athletic Legaue, special after-school programs in conjunction with museums, summer camp programs, etc. The fact that the coefficient for Possessions seems powerful leaves open the question of what Possessions really is. The facile argument that Possessions represents money, so that cash grants to parents will produce reading achievement, I find unsupportable. One would have to posit a production relationship between money and cognitive skills. This could run: money releases the parent's time to be spent with the children, and this time is the actual producer of these skills. Yet the time of a now poor parent may not be so productive. We certainly don't know that it is from regression coefficients from parents who have achieved some higher income status.

The unfortunate truth is that until the actual productive variables are identified, we cannot well estimate their costs. And without both production and cost estimates, we cannot know what policy might be most cost-effective. Add to this the fact that we do not know what form a production relationship might take, and the discovery in the preceding sections that the form makes a great deal of difference in terms of estimated effects, then the lesson of this paper is clear: much more work, theoretical, experimental, and statistical, needs to be done before we will be able to rationalize the production of cognitive skills.

V

A BRIEF CONCLUSION

This paper began with a discussion of the concept of "production functions," and a description of how a production function can direct decision making toward economic efficiency. This, it was noted, required that the function was derived from technically efficient (well managed) production units.[1] Where profit making is the force behind the production, one can assume some degree of technical efficiency. This does not apply to public education.[2] The concept that one should try to isolate efficient schools--those operating on their production frontier--was outlined, and even tested in Part IV of the paper.

The output focus of the school was taken to be another problem. A single output measure, or even an index, has very limited use in actual production estimation of such conglomerates as the public education system. This is especially true if there is a non-linear expansion path as resources increase, or if schools with different kinds of children (by social class, urbanness, ethnicity, etc.) aim at different outputs. This argument appeared in Part III.

Even with schools with a common output goal, and a suitable measure of that goal; even if the firms were observed to be technically efficient, other problems occur. Previous attempts to statistically relate cognitive output to

[1]Alternatively, it has been implied, inefficient production can be assumed to be the rule, and the cost of efficient management should be part of the economic decision.

[2]It does raise the thought that private education is the place to look for production function estimates. The problem here, of course, is that too few observations of poor background children in resource-rich private schools could be found to disentangle these two influences.

school inputs have not separated out the children who were not in that school in previous years, hence not actually associated with those inputs. If lower class children are more likely to change schools, then once again it may be extremely difficult to make any estimates about input-output relationships which affect them. Also, the range of administrative jurisdictions in the data sample may affect the results. Thus it was argued that within one city, teachers with seniority can associate themselves with higher output children, presenting an upward bias to the "production" estimate unless made from a suitably identified and estimated simultaneous model. On the other hand, multi-city, and especially city and suburb samples allow the same kinds of associations with other teacher (and principal) qualities. A teacher with experience can move well within a city. A teacher with high IQ, social ease, or whatever else is desired by suburban education establishments may move more easily between cities. These issues were discussed in the beginning of Part IV.

Even if all these problems are considered solved, the mathematical form of the production function requires consideration. A few basic theoretical issues were discussed in Part II. The focus there was on basic characteristics implicit in functional forms, some of which are independent of the data. The elasticity of substitution between factors, and the elasticity of output with respect to a factor--two obviously important production considerations--were discussed. The major point made was that though many considerations of form could be obviated by careful data transformations (though this is certainly not true of all of these considerations), the researcher like myself has very little guidance from educational theory as to how to proceed in either direction.

To demonstrate the importance of this latter problem, there being little I could do about the others, Part IV investigated several kinds of estimations with one sample of data. Weighted regressions were used to give 30 production units (schools) equal weight, but to correct for variations in the qualities of the raw materials (pupils) as they entered the process. It had been pointed out in Part I that in fact this cannot be done with present data, and that the background variables therefore account for initial quality, continued background influence during the production process, allocation of the highly averaged school resources within the production unit, and variable response to school resources. Estimates were given for a simple linear form, for the simple linear plus interactions, the simple linear with quadratic terms, a logarithmic transformation, and the simple linear with both interactions and quadratic terms. Estimates of the partial derivatives of this last form, with respect to two school variables and a background variable (evaluated at mean values) were significantly different from those of the simple linear form.

In addition, by presenting the equations for two different outputs, I have tried to caution the reader against drawing conclusions from equations from one output only. Despite high correlations between these two outputs, the equations were different, including different sets of both school and background variables. Some major conclusions--such as the relative impact of background and school variables--would be the same from these two particular outputs. But since education is admittedly a multi-faceted product, one should be forewarned about accepting or rejecting notions based on just one of those facets.

I have tried not to weary the reader with argument about the triviality of the notion that background exceeds school inputs in explaining output variance.

It should be clear that such a finding is not surprising, not depressing, and in short not important. The ultimate efficacy of any program will be determined on its cost-effectiveness, that is, its ability to deliver output for dollars. For this calculation, estimates of the increment to output from increasing inputs are necessary, and estimates of associated variance, of no import. What was depressing was to find that the presumably better specified equation reduced the estimates of the response of output to school inputs from the simple linear estimates. Thus accurate estimates of cost-effective policies rely heavily on accurate production function estimation. I have tried to stress how far we are from that goal.

Last of all, I questioned the calculus by which even cost-effective decisions are made. It is on this point that I will end this paper. If "benefit" is calculated independent of who receives it, and if the national objective is to "maximize benefit," then policies which favor the already favored could easily be recommended. This has long been a problem in U. S. Office of Education planning. For example, if the returns to a program are calculated as the increment to lifetime earnings of the individuals receiving the benefits, discounted to the time of the program, then aid to graduate students will almost surely dominate aid to young children. Just the fact that the young children receive no "benefit" as calculated for many years lowers the value of that benefit according to these rules.

Since no such accounting mechanism would ever justify aid to young children, the Office has long compartmentalized its objectives. It considers aid to young children a desideratum, and evaluates those programs against each other. What I am suggesting is that even <u>within</u> a category such as "young children," benefits should not be equal weighted. This is, in fact, the philosophy behind Title I of the 1965 Elementary and Secondary Education Act. Funds from this Act supposedly go to educationally deprived children in poverty settings. No comparisons are made with the benefits which would accrue to educationally favored children if they had access to these funds. That is well and good.

The same compartmentalization should be continued in the actual administration of these funds. If the interactions are as I have found them, then nothing short of massive and well managed aid will show any success. If the expansion path of outputs is as I have suggested in Part III, then most of this money is going into schools which are not attempting to produce the outputs by which Title I is being evaluated. This dilemma could lead, at the moment, only to a longer discussion than I care to be involved in at this time. Some policy options are obvious: This money (or other money to the same children) should be directed <u>outside of schools</u>, in afternoon and summer programs not under the charge of the school bureaucracy. Research and training are needed to supply methods and teachers specifically geared for a particular, deprived, child population. Rewards need to be output oriented, without being so specific that other outputs are grossly sacrificed, or that only the output <u>measure</u>, not the output itself, is raised.

It is not my place, at least not here, to discuss these options further.

---

[3] Other outputs may well be sacrificed to achieve one--that is the nature of the production frontier. But some tradeoffs are unacceptable. The output measure might be a vocabulary, spelling, etc., test. The output, however, is vocabulary, spelling, etc. Training for the test is hardly worth rewarding.

A truly heroic effort at determining an educational production function would carefully consider all the times and places in which a person acquires the skills which we wish to consider as outputs. Then--and only then--we could design programs which could be effective (and better yet, cost-effective) in producing those skills. To limit the investigation to currently administered public schools is perhaps the basic flaw, superceding all the others mentioned above, in the unrewarding efforts to estimate educational production functions.

BIBLIOGRAPHY

1. Bowles, Samuel. "Towards an Educational Production Function," paper delivered at the NBER Conference on Research in Income and Wealth, November, 1968.
2. ________. Educational Production Function, Report to the U. S. Office of Education, February, 1969.
3. ________, and Henry M. Levin. "The Determinants of Scholastic Achievement--An Appraisal of Some Recent Evidence," The Journal of Human Resources, Vol. 3, No. 1, Winter, 1968.
4. Bruner, Jerome. Toward a Theory of Instruction, Belknap Press of Harvard University, 1966.
5. Burkhead, Jesse, with Thomas G. Fox and John W. Holland. Input and Output in Large-City High Schools, Syracuse University Press, 1967.
6. Cain, Glen G., and Harold W. Watts. "Problems in Making Policy Inferences from the Coleman Report," American Sociological Review, 35, April, 1970; pp. 228-242.
7. Coleman, James S., Ernest Q. Campbell, and others. Equality of Educational Opportunity, U. S. Office of Education, OE-38001, 1966.
8. Guthrie, James W., George B. Kleindorfer, Henry M. Levin, and Robert T. Stout. Schools and Inequality, The Urban Coalition, 1969.
9. Hayes, Donald P., and Judith Grether. "The School Year and Vacation: When Do Students Learn?", presented at the Eastern Sociological Association Convention, April, 1969.
10. Kain, John F., and Eric Hanushek. On the Value of Equality of Educational Opportunity as a Guide to Public Policy. Program on Regional and Urban Economics, Discussion Paper No. 36, Harvard University, May, 1968.

11. Katzman, Martin T. Distribution and Production in a Big City Elementary School District, Ph.D. Dissertation, Yale University, 1967.

12. Kershaw, Joseph A., and Roland N. McKean. Systems Analysis and Education, RAND Corporation, Research Memorandum RM-2473-FF, 1959.

13. Levin, Henry M. "A New Model of School Effectiveness," in Do Teachers Make a Difference?, U. S. Office of Education, Bureau of Educational Personnel Development, 1970; pp. 55-78.

14. ________. Recruiting Teachers for Large-City Schools, Charles E. Merrill, 1970.

15. Michelson, Stephan. "Equal Protection and School Resources," Inequality in Education, Number 2, December, 1969.

16. ________. "The Association of Teacher Resourceness with Children's Characteristics," in Do Teachers Make a Difference?, U. S. Office of Education, Department of Educational Personnel Development, 1970, pp. 120-168.

17. Mood, Alexander M., "Do Teachers Make a Difference?" in Do Teachers Make a Difference?, U. S. Office of Education, Bureau of Educational Personnel Development, 1970; pp. 1-24.

18. Owen, John. "The Distribution of Educational Resources in Large American Cities," unpublished, Johns Hopkins University Center for the Study of the Social Organization of Schools, 1970.

19. ________. "The Allocation of Teaching Resources in American Public School Systems," unpublished, Johns Hopkins University Center for the Study of the Social Organization of Schools, 1970.