Self-Selection and Student Achievement

September 1996

HONGGAO CAO

Population Research Center and Department of Education
University of Chicago, Chicago, IL 60637

Abstract

Students in any schools are not a random collection from the population. They become schoolmates because of their parents' selections of school quality that are contingent on their genetic abilities and family background. Even specified correctly, the conventional educational production functions cannot be used to find the effects of school inputs or quality. Therefore, the weak or zero relationship between school inputs and student achievement widely documented in the literature by no means implies that school inputs or quality does not matter. Further, since students enter schools by self-selection, any observed differences in student achievement between public and private schools do not necessarily mean that these schools have any differences in the effectiveness of operation. [JEL I21]

I. INTRODUCTION

STARTING FROM the controversial Coleman Report (Coleman et al, 1966), writers on educational production functions (EPFs) have found no strong or systematic relationship between student achievement and school quality, after controlling for student own characteristics such as genetic abilities and family background. A literature review by Hanushek (1986) painted a very dismal picture for the U.S. public school system: "It appears from the aggregate data that there is at best an ambiguous relationship and at worse a negative relationship between student performance and the inputs supplied by schools." (p.10) Historical evidence seems also to support a claim that the improvement in student achievement in the U.S. has not matched the rising costs or "quality" of public education.

Two basic implications have been drawn from the findings. First, the U.S. public schools are operated in an inefficient way: School resources are not used efficiently to improve student performance. Second, school quality as measured by conventional school inputs is generally not relevant to student achievement. To substantiate the implications further, Levin (1972, 1976) hypothesizes that schools are inherently inefficient. And several studies comparing the public with private and, especially Catholic, schools (Coleman, Hoffer, and Kilgore, 1982a, 1982b; Greeley, 1982; Hoffer, Greeley, and Coleman, 1985;) not only find "Catholic-school advantages" in increasing student achievement, but associate the advantages with some variables beyond conventional school inputs.

The above findings and implications, however, have been subject to numerous objections. In the first place, spontaneous challenges came from the people in the educational circle, who had long emphasized that school quality matters. Most of these people then went on to question the relevancy of the EPFs to the effect of school quality, as many assumptions required of a standard production function do not hold water in the school context. For example, the functional form "is unknown (to both decision-makers and researchers) and must be estimated using imperfect data; some important inputs cannot be changed by the decision-maker; and any estimates of the production will be subject to considerable uncertainty." (Hanushek, 1986, p.11)

Criticisms have also come from other angles. Brown and Saks (1975) argue that the objective of schools may consist of not the student achievement level alone, as implied in most EPF studies, but also the distribution of student achievement. Based on a meta-analysis, Hedges, Laine, and Greenwald (1994) question the credibility of the negative judgements by Hanushek (1986) about the school quality effects in the EPF studies. In an effort to investigate the relationship between school quality and individuals' subsequent labor market performance, Flyer and Wachtel (1995) establish that one key variable in most EPF studies, teacher-student ratio, may be a very poor proxy of school quality. For the open enrollment policy in the U.S. public schools, parental optimizing behavior, along with the inability of public schools to make appropriate adjustments to parental behavior, would make the variable only reflecting disequilibrium conditions.

In this paper, I propose an alternative framework to examine the determination and distribution of student achievement. Instead of claiming the inherent inefficiency of school systems, or arguing about methodological shortcomings of some particular studies, I shall take for granted all relevant empirical findings. My main points are the following: Students in any schools are not a random collection from the population. They become schoolmates because of their parents' selections of school quality that are contingent on their genetic abilities and family background. Even specified correctly, the conventional educational production functions, which contain both school inputs and family background variables, cannot be used to find the effects of school quality. Therefore, the weak or zero relationship between school inputs and student achievement widely documented in the literature by no means implies that school quality does not matter. Further, since students enter schools by self-selection, the observed "Catholic-school advantages" or any other differences in student achievement between public and private schools do not necessarily mean that these schools have any differences in the effectiveness of operation. In fact, in a strong version of the self-selection model, the observed student achievement would have nothing to do with the school, even the school does contribute to the achievement production.

The idea of self-selection is not new[1]. The neglect of the idea in the studies of student achievement determination and distribution, therefore, is very surprising, especially given the well-known human capital theory and long-held belief in the effect of the school on the production of human capital. After all, if people can make optimal human capital investment by choosing quantity of education, and if school quality contributes to the production of human capital, what prevents them from choosing quantity and quality of education simultaneously?

In section II, I set up a model illustrating how self-selection happens in the context of school quality choice. In section III, I reinterpret the conventional EPF results based on the self-selection idea. These reinterpretations lead to the conclusion that the conventional EPF studies would fail to find the total or independent effect of school inputs on student achievement. Section IV discusses a more general self-selection model, which gives rise to a very surprising result about the roles of the family and the school in producing human capital in general, and improving student achievement in particular. Some applications of the model are discussed in section V.

II. SELF-SELECTION

Consider a world of full information. Three actors, schools, parents, and students, play together in this model. A representative school produces and provides for its students services called "quality" according to the following rule,

$$K = f(\bar{H}, \bar{X}, n). \tag{1}$$

Here K is an index of quality produced in the school; n is the school size, capturing the pure scale effect, and measured by the number of students in the school; $\bar{H}$ is the students' average learning ability; $\bar{X}$ is a mean vector of other inputs (averaging over the n students); $\frac{\partial f}{\partial \bar{H}} > 0$, and $\frac{\partial f}{\partial \bar{X}} > 0$. The inclusion of the students' average learning ability in the quality production reflects the peer effect, which has been emphasized in almost all educational studies of student achievement. The mean vector of other inputs includes all other variables relevant to the quality production. To simplify, we assume that $\bar{X}$ is monetarily measurable. As a result, it may be understood conveniently as expenditure per student spent in the school.

The school is a not-for-profit organization. It is operated efficiently, and charges the parents of each student exactly $\bar{X}$ as tuition or compensation for providing quality *K* unless the tuition is paid by others (the state or local government in the public school case, for instance).

A representative student is endowed with learning ability, *H*. When he studies in a school of quality *K*, his achievement index *S* will be

$$S = g(H, K), \tag{2}$$

where *g* is increasing in both *H* and *K*. Clearly, by specification (2), we assume that school inputs and, hence, quality matters in influencing student achievement.

Parents are altruists. They care about both their own consumptions and their children's performance in the school. Specifically, a representative parent, endowed with a fixed resource, *W*, maximizes his utility function

$$U = U(S, Z), \tag{3}$$

by allocating his total resource between his own consumption *Z*, and the tuition $\bar{X}$ for sending his child to the school of quality *K*.

Beside the assumptions to be made explicit as we proceed, we generally assume that a representative family consists of one parent and one child; that parents have the sole right to choose schools for their children; and that except for tuition, their access to schools of any quality is completely free. In addition, we assume that all the functions *f*, *g*, and *U* are well defined, and think of both *H* and *W* as "student characteristics".

A. The Single Student School

We start the journey by considering a very simple case: Each school has only one student. The representative parent maximizes his utility function (3) subject to the equations (1) and (2), and his budget constraint $X + Z = W$. For an interior solution, the first order condition for this problem requires that the marginal rate of substitution between his own consumption and the child's achievement be equal to the marginal productivity of X. Solving the condition, we will know how much the parent is prepared to pay for his child's education. Denote the solution as

$$X^{*} = X(H, W). \tag{4}$$

The effects of the student's characteristics on the tuition that the parent is willing to pay depend on the functional forms of f, g, and U. Presumably a richer parent tends to spend more on his child's education. Holding parental resources constant, a smarter child tends to go to a school of higher quality. However, we are not preoccupied with any particular effects of the student's characteristics here.

Substituting the X^{*} into equations (1) and (2), one gets the following achievement determination equation

$$S = g(H, K) \equiv S(H, W). \tag{5}$$

Not surprisingly, a student's achievement in the single-student school is completely determined by his own characteristics. The school seems to play no role in producing his achievement, although the underlying mechanism, equation (2), suggests that school quality matters.

B. The School as a Group of Students

We now consider a more realistic situation where a school consists of a group of students ($n > 1$). Since schools are operated efficiently, each school with quality K has an optimal size, n_k^*, satisfying the condition $\frac{\partial f}{\partial n_k^*} = 0$. To distinguish one family from another, we index a representative student and his parent with subscript j.

The parent j's willingness to pay for his child's education may be found in the way similar to equation (4),

$$X_j^* = X(H_j, W_j; \bar{H}) = X(H_j, W_j; K), \tag{6}$$

where $\bar{H}$ and K are the students' average learning ability in, and the quality of, the school that child j is attending. Holding the student's characteristics H_j and W_j constant, the parent's willingness to pay is generally a positive function of the students' average learning ability. For the child will get more positive peer effect, the higher the students' average learning ability in the school he is attending. Does this imply that the parent would choose $\bar{H}$ and, hence, school quality K, for his child as high as possible? The answer is no, because higher school quality is accessible only if he sacrifices more in his own consumption. The optimal quality that he chooses, $K_j^* = K(H_j, W_j)$, thus follows a conventional rule that equates the utility increase due to a higher quality of his child's school, and the utility decrease due to a lower own consumption.[2]

Different parents, basically, choose different K^*, depending upon their resources and their children's learning abilities. Those who end up with the same K^* may have their children enter the same school. Consequently the students in any school are not a random collection from a population characterized by the distributions of H and W. Rather, they come together following a specific rule. Their achievements are completely determined by their own characteristics, although the underlying mechanism says that school inputs and quality matters.

The student's characteristics also determine the number of schools of different quality. For quality $K = K_i$, for example, the number is equal to N_{ki}/n_{ki}^*, where N_{ki} is the number of students whose characteristics (H, W) make them enter schools of quality K_i, and n_{ki}^* is the optimal size of the schools with quality K_i. Notice, even all students in the population share the same characteristics, they may still have to be divided into different schools, as all schools have optimal sizes.

The within-school variation in achievement depends on the relationship between H and W within schools (i.e., $\frac{\partial^2 K^*}{\partial H \, \partial W}$), whereas the between-school variation primarily reflects the population distribution of the student characteristics. If all students in the population share the same characteristics, for example, both the within-school and between-school variations would be zero. If H and W is exactly linearly related, and the parental optimal choice of school quality K^* is linear both in H and W, the between-school variation would dominate the within-school variation, because the students within schools share the same characteristics, while the quality across schools differs. If the joint distribution of H and W is not very spread, but the students within schools are very diverse, the within-school variation would dominate the between-school variation, and school quality would not differ very dramatically. In any case, neither the within nor between school variation is appropriate for ascertaining the school effect.[3]

C. Public versus Private Schools

A school is called "public" if it provides a certain level of quality to any students for free. The levels of quality provided in public schools depend on two sets of interrelated factors. One is various political forces, which determine how much a local or state government pay for its public education. The other is the students' average learning ability in the public schools. If the average learning ability differs from one school to another, even identical physical inputs do not yield the same level of quality. On the other hand, since public schools adopt open enrollment policy, differences in physical inputs may not imply differences in quality across the schools. For otherwise, all parents would choose for their children the schools providing the highest quality. Indeed, if migration were free, and if the parents' resources were not sensitive to where they live, all public schools would provide the same level of quality, no matter how much local or state governments pay for the public education.

Of course, migration is not free; and parents' resources are dependent on where they live. Consequently, the realized school quality may differ from one school to another even in the same communities. But due to self-selection, any observed variation in student achievement would not be explained by school input variables, but by family and social variables, including those affecting parental migration decisions, and the dependence of the parent's resources on residence.

The differences in student achievement between public and private schools tell us little about the relative effectiveness of the two school sectors, even if the two sectors are characterized by the same sets of physical inputs. In other words, higher achievements in private schools do not necessarily imply that private school are operated more efficiently, as they may draw in more talented students, who are supposed to achieve more.

People who insist on private or Catholic school advantages have found that those advantages are attributable to some policies not present in public schools, notably the policies in curriculum assignment, advanced course work, homework, and discipline climate (Hoffer, Greeley, and Coleman, 1985). One might argue, then, if those policies turn out to be the reason for the advantages, why public schools cannot just adopt the same policies for the benefit of their own students? The answer based on the self-selection model may be the following: the observed differences in school policies between the two sectors reveal the underlying differences in student characteristics; a policy good for students in one sector may not be appropriate for students in the other.

III. EDUCATIONAL PRODUCTION FUNCTIONS

A typical educational production function may be specified as

$$S = a + X\beta + F\gamma \tag{7}$$

where S, as defined earlier, is a student's achievement index; X and F are vectors of school inputs and student characteristics; a is a scaler constant; and β and γ are coefficients to be estimated, measuring marginal productivity of X and F.

A school with input vector X has a linear cost function

$$C = X\alpha, \tag{8}$$

where α is a vector of coefficients reflecting the marginal costs of school inputs. Because of self-selection, a student with characteristics F would attend a school with certain inputs X. The relationship between F and X may be expressed as a set of "student-characteristics supply equations",

$$F = G(X). \tag{9}$$

Since the school is efficient, it selects the optimal sets of inputs to maximize the following objective:

$$M \equiv S - C\lambda = a + X(\beta - \alpha\lambda) + G(X)\gamma, \tag{10}$$

where λ is a constant indicating the value of one unit of school cost in terms of achievement. The first order conditions require

$$\beta + G'(X)\gamma = \alpha\lambda, \tag{11}$$

which yield

$$G(X)\gamma = X(\alpha\lambda - \beta) + m, \tag{12}$$

with m a constant of integration, capturing the individual-specific objective on achievement.[4]

For student j in school i and with characteristics F_{ij}, the school's equilibrium objective or performance is $M^*_{ij} = a + m_{ij}$, independent of both the school inputs and student characteristics appeared in the educational production function. However, since m_{ij} is individual-specific, the school does not achieve the same objectives for all students. The role of an efficient school is simply to satisfy the predetermined needs of various families for their children's achievements.

The overall performance of the school may be defined by the following metric,

$$M_i = E(M_{ij}^*) = a + E(m_{ij}), \tag{13}$$

where the expectation is over all the students in the school. Clearly, the conventional educational production function is not able to measure the school performance, because the later is not determined by the school inputs or student characteristics in the function. On the other hand, since the metric results from the assumption of school efficiency, it is not an appropriate measure for assessing the school effectiveness.

The student *j*'s equilibrium achievement can be expressed in two alternative ways:

$$S_{ij} = a + m_{ij} + X_i\alpha\lambda \equiv M_{ij}^* + R_i, \tag{14.1}$$

or

$$S_{ij} = a - \theta m_{ij} + F_{ij}\gamma(1+\theta), \tag{14.2}$$

where $\theta = (\alpha\lambda-\beta)'[(\alpha\lambda-\beta)(\alpha\lambda-\beta)']^{-1}\beta$; and $R_i \equiv X_i\alpha\lambda$ may be interpreted as the independent effect of the school on achievement. Notice that, instead of being β, the total effects of individual school inputs on achievement are equal to the marginal costs of the inputs. Accordingly, the total effects of student characteristics are not characterized by vector γ, but by $\gamma(\theta+1) \equiv \gamma^*$. Also, the role of the school is two-fold: beside achieving the objective expected by the family, M_{ij}^*, it can exert an independent effect on the student's achievement.

Now consider the following empirical specifications:

$$S_{ij} = b_1 + X_i\alpha^* + \mu_{1ij}, \tag{15.1}$$

$$S_{ij} = b_2 + F_{ij}\gamma^* + \mu_{2ij}, \tag{15.2}$$

$$S_{ij} = a + X_i\beta + F_{ij}\gamma + \mu_{3ij}. \tag{15.3}$$

Here, a, b_1, and b_2 are scaler constants; and μ_{kij} are i.i.d. random disturbances. All other variables and coefficients have been previously defined.

The first two regressions are simply empirical version of the equilibrium relationships (14.1) and (14.2), while the last one represents almost all *ad hoc* models of educational production function. The dominant finding based on this *ad hoc* regression is that the estimates of the coefficient vector β are either close to zero or insignificant, which often leads to the inference that school inputs do not matter, or that schools are operated inefficiently. From the analysis above, however, the effect of school inputs or quality on student achievement should not be assessed through the *ad hoc* regression. While the partial effects of the individual school inputs may be unimportant, their total effects, which determine the independent effect of the school as a whole in the achievement production, may not. Therefore, a weak relationship between student achievement and school inputs from the conventional education production function studies does not necessarily imply that school quality plays no or insignificant role in the student achievement production, nor it implies that schools are inefficient. Unfortunately, very few efforts have been made to measure the total effects of individual school inputs or the independent effect of the school as a whole as specified in regression (15.1).

To show how the self-selection idea contributes to the understanding of empirical findings, one may take advantage of the equilibrium equations derived above. Assuming the regressions (15.1) -

(15.3) are feasible, one should have the following results.[5]

Result 1. The partial effects of school inputs would understate the independent effect of the school as a whole on student achievement if $X_i\hat{\beta} < X_i\hat{\alpha}^*$; and vice versa.

Result 2. If the partial effects of school inputs understate the independent effect of the school as a whole on student achievement, the partial effects of student characteristics would overstate the non-school effect; and vice versa.

Proofs: Obvious.

Result 3. The relative importance of the school to the family in student achievement may be measured by $\rho \equiv \frac{Var(X_i\hat{\alpha}^*)}{Var(S_{ij}-X_i\hat{\alpha}^*)}$, where $\hat{\alpha}^*$ is the estimates of α^*.

Proof: According to (14.1), the residuals from regression (15.1) estimate the family-specific objectives on student achievement, while $X_i\hat{\alpha}^*$ estimate the independent effect of the school on student achievement. Notice that this result presumes no measurement errors: The disturbance term, μ_{1ij}, measures only the family-specific variation in m. When measurement errors are substantial, one might be interested in another metric, $1/\rho$, which may be interpreted as an index of the relative importance of the family to non-family factors in determining student achievement.

Result 4. The overall performance of school i as defined in equation (13) may be measured by $\hat{M}_i \equiv \sum_j [S_{ij} - X_i\hat{\alpha}^*]/n_i$, where n_i is the number of students in school i.

Proof: $E(\hat{M}_i) = E(\sum_j [S_{ij} - X_i\hat{\alpha}^*]/n_i) = b_1 = a + E(m_{ij})$. The last equality is due to the equation (14.1).

Result 5. If there are several distinct school sectors, the overall performance of sector I, which consists of N_I schools, may be measured by $\hat{M}^I \equiv \sum_{i \ni I} (\hat{M}_i / N_I)$.

Proof: Obvious.

As indicated earlier, the performance measures should not be used to judge the relative effectiveness of different schools or school sectors. For both measures are derived from the assumption that schools are operated efficiently. The differences in performance, however, can be regarded as indicative of self-selection. The findings, if any, that students in private schools outperform their counterparts in public schools do not necessarily imply that private schools are more efficient; it may be the case that the private schools are expected to achieve more by the families of their students.

IV. A More General Model

Previous criticisms of the educational production functions have questioned the validity of many required assumptions. Few of them, however, have asked the following question: Given the underlying rationale that school quality and family background, rather than individual school inputs and student characteristics, determine student achievement, what would happen if we had a way to know school quality and family background ***per se***?

The question should have been answered. For then one would have been able to distinguish the effect of school quality from the efficiency of schools in producing quality. The opposition to the empirical studies on the educational production functions would not have been so fierce as it is: Although many people never believe that school quality does not matter, they probably agree that school operations need to improve.

The benefit from using school quality and student family background ***per se*** in the educational production functions may be more substantial in the framework of self-selection. Consider the following educational technology:

$$S = \delta_1 K + \delta_2 B. \qquad (16)^{6}$$

A student's achievement is determined linearly by the quality of the school he is attending, K, and his own family background, B. The school quality is produced efficiently by using a set of inputs X according to some rule $K = K(X)$, while the family background is related to the family variables through the equation $B = B(F)$.[7] The definition of X and F has been given previously.

Students are sorted into schools, with a reduced form family-background supply equation

$$B = \pi(K). \qquad (17)$$

For student j with family background B_{ij}, the optimal school quality is the K_i that maximizes $S_{ij} = \delta_1 K_i + \delta_2 \pi_j(K_i)$. The equilibrium relationship between B_{ij} and K_i is then

$$B_{ij} = \phi_j - \frac{\delta_1}{\delta_2} K_i, \qquad (18)$$

where ϕ_j is a family-specific constant of integration. The observed achievement of the student j is $S_{ij}^* = \phi_j \delta_2$, a family-specific constant independent of not only F and X, but also K and B. Notice that even school quality matters, and even schools are operated efficiently, the observed student achievement does not have to be explicitly related to school inputs. This finding is very strong.

The result that the observed achievement is family-specific but not dependent on family background seems puzzling. But it could be the case that each family has an exogenous objective for children's education, and this objective is independent of the family background. Poor families, for example, do not necessarily have lower objectives than rich families for their children's education. As long as it is feasible, all families would choose optimal plans to achieve their objectives, no matter what background they have.[8]

The observed achievement's independence of school quality points to a unusual relationship between the school and the family in producing student achievement in particular and human capital in general: they are substitutes, and work together to achieve prescribed objectives. The fundamental role of the school, therefore, is not to increase student achievement or human capital, but to be a complement to the family in realizing a certain goal.

In the human capital literature, the family and the school have usually been treated as two different organizations in terms of their roles in human capital production. The outcome of an individual's human capital investment is often perceived as a function of his/her family background (including learning ability), and sensitive to his/her education. Based on the self-selection model sketched here, however, the perception is not totally right. If individuals are rational, and able to make their own choices, the outcomes of their human capital investments are predetermined, and should not depend on which schools they are in, and how much education they get, for the same reason that student achievement should not be expected to depend on school quality.

V. Applications

A. International Comparisons

The above analyses critically hinge on the availability of selection. They predict that, when selection is available, the conventional EPF approach may not identify the school effect, even school inputs contribute to student achievement production. Presumably, the more the parents have freedom to choose schools for their children, the closer the situation is to our model, and the more likely the conventional EPF approach fails to identify the effect of school inputs. This prediction provides a good explanation for different findings about the effects of school inputs in developed countries, particularly in the United States, from in developing countries. While the general opinion about the U.S. school systems is pessimistic, studies in many developing countries, including India (Shuluka, 1974), Uganda, Kenya, Ghana, Papua New Guinea, and Somalia (Heyneman, 1980; Heyneman and Loxley, 1983), conclude that school inputs are much more important than perceived in developed world. For example, "it has been shown in a sample of twenty-nine countries that the proportion of explained test score variance attributable to school quality is lowest in developed countries such as Australia, Japan, Sweden, and United States, but it is twice or three times as high in Brazil, Botswana, India, or Thailand." (Psacharoupolos and Woodhall, 1985, p.217)

The U.S. school systems, public and private alike, are perhaps the freest in the world. Parents can choose for their children private or religious schools if they do not want to send their kids to public schools. In order to have their children receive education best satisfying their needs or objectives, they may relocate to anywhere they want. It is not an abnormality that parents move homes from one community to another simply for the sake of their children' education.

On the other hand, parents in developing world do not have much freedom to choose schools. In some countries, educational resources are so limited that parents, if ever wanting their children to receive any education, can only send them to whatever schools practically accessible. In other countries, relocation, if not impossible, is very difficult. Students in schools are much more like a random collection from the population than in the developed world. Consequently, the conventional EPF studies in developing countries are more likely to reveal the effects of school inputs on student achievement.

B. Time-Series Evolution of Student Achievement in the U.S.

One piece of evidence supporting the claim that the U.S. public schools are operated inefficiently is: While school inputs in terms of per pupil expenditures have approximately tripled over the past several decades, student achievement has shown no improvement, sometimes even declining. The evidence is not sufficient.

If the school is simply the complement to the family in determining student achievement, the time-series variation in achievement would only represent the time-series variation in family's objectives for children's achievements, which may result from time-series change in demand for human capital. Instead of increasing achievement independently, improvement in school quality may have shifted part of responsibilities originally taken by the family in producing children's achievements. The zero effect of improved school quality on student achievement says nothing about either the efficiency of school operations or the relative importance of the school.

C. School Inputs and Individuals' Subsequent Labor Market Performance

In labor economics, studies relating school inputs to individuals' performance in labor market have been divergent. In one line, Welch (1966), Morgan and Sirageldin (1968), Johnson and Stafford (1973), Wachtel (1976), Rizzuto and Wachtel (1980), and Card and Krueger (1992) have found significant effect of school inputs on individual earnings. On the other hand, Heckman, Layne-Farrar, and Todd (1994) and Betts (1995) have found no evidence on the effect. The self-selection model may help reconcile the differences.

A typical study on the effect of school inputs on an individual's labor market performance specifies an individual's earnings in age t as a function of the individual's age, schooling years, and school input indexes. The underlying assumption for any explicit effects of school inputs is that those inputs can influence the individual's intertemporal earnings independent of his schooling years. If an individual is allowed to make choices, however, his/her schooling years and the quality of schools that he/she attended are interdependent. A model including both quantity and quality of education would therefore only produce biased estimates for the effects of educational quantity and quality.

The more general self-selection model makes the studies of school input effects even more vulnerable. If the school and the family are complements, and an individual's optimal amount of human capital is independent of his decisions on education, there would be no justification for the conventional specification using educational variables as determinants of an individual's earnings.

VI. Concluding Remarks

The question "Does school quality matter?" has been raised time and time again. Substantial efforts have been made to identify the expected effects of school inputs on student achievement, and, when no such effects are found, to explain why. This paper refutes two of the primary, and the most controversial, implications from the conventional educational production function studies: (i) Schools in the U.S. are operated inefficiently; and (ii) School quality does not matter.

The U.S. school systems probably need to improve; the relationship between student achievement and school inputs in the U.S. are indeed very weak. But due to self-selection, which is most likely in the U.S. than in any other country, the conventional EPF studies cannot be trusted to draw conclusions about the effectiveness of schools, as well as the relative importance of school quality in student achievement. More efforts are needed to address these issues.

REFERENCES

Betts, J.R. (1995) Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth. *Review of Economics and Statistics* **77**, 231-250.

Brown, B.W. and Saks, D.H. (1975) The Production and Distribution of Cognitive Skills within Schools. *Journal of Political Economy* **83**, 571-593.

Card, D. and Krueger, A.B. (1992) Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy* **100**, 1-40.

Coleman, J.S. (1990) *Equality and Achievement in Education* Westview Press.

Coleman, J.S. et al. (1966) *Equality of Educational Opportunity* Washington: Government Printing Office.

Coleman, J.S., Hoffer, T. and Kilgore, S. (1982a) Achievement and Segregation in Secondary Schools: A Further Look at Public and Private School Differences. *Sociology of Education* **55**, 162-182.

------. (1982b) *High School Achievement*. New York: Basic.

Flyer, F. and Wachtel, P. (1995) Parental Optimizing Behavior and the Measurement of School Quality. Unpublished Working Paper, New York University.

Greeley, A.M. (1982) *Catholic High Schools and Minority Students*. New Brunswick, N.J.: Transaction.

Hanushek, E.A. (1986) The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature* **24**, 1141-1177.

Heckman, J.J., Layne-Farrar, A.S. and Todd, P.E. (1995) Does Measured School Quality Really Matter? Unpublished Working Paper, University of Chicago.

Hedges, L.V., Laine, R.D. and Greenwald, R. (1994) Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. *Educational Researcher* **23**,

Heyneman, S.P. (1980) Differences between Developed and Developing Countries: Comments on Simmons and Alexander's "Determinants of School Achievement". *Economic Development and Cultural Change* **28**, 403-406.

Heyneman, S.P. and Loxley W. (1983) The Effect of Primary School Quality on Academic Achievement across 29 High- and Low-Income Countries. *American Journal of Sociology* **88**, 1162-1194.

Hoffer, T., Greeley, A.M. and Coleman, J.S. (1985) Achievement Growth in Public and Catholic Schools. *Sociology of Education* **58**, 74-97.

Johnson, G.E. and Stafford, F.P. (1973) Social Returns to Quantity and Quality of Schooling. *Journal of Human Resources* **8**, 139-155.

Levin, H.M. (1974) Measuring Efficiency in Educational Production. *Public Finance Quarterly* **2**, 3-24.

------. (1976) Concepts of Economic Efficiency and Educational Production. In *Education as an Industry* (Edited by Froomkin, J. et al.) Cambridge, Mass: Ballinger Publishing Company.

Morgan, J. and Sirageldin, I. (1968) A Note on the Quality Dimension in Education. *Journal of Political Economy* **76**, 1069-1070.

Psacharoupolos, G. and Woodhall, M. (1985) *Education for Development: An Analysis of Investment Choices* Oxford University Press, published for the World Bank.

Rizzuto, R. and Wachtel, P. (1980) Further Evidence on the Returns to School Quality." *Journal of Human Resources* **15**, 240-254.

Shuluka, S. (1974) Achievements of Indian Children in Mother Tongue (Hindi) and Science. *Comparative Education Review* **18**, 237-247.

Willis, R.J. and Rosen, S. (1979) Education and Self-Selection. *Journal of Political Economy* **85**, 507-536.

Wachtel, P. (1976) The Effects on Earnings of School and College Investment Expenditures. *Review of Economics and Statistics* **58**, 326-331.

Welch, F. (1966) Measurement of the Quality of Schooling. *American Economic Review*, Papers and Proceedings **56**, 379-392.

NOTES

1. For example, in an application of Roy model, Willis and Rosen (1979) used the idea to show why the conventional earnings-generation equation would yield biased estimates of the return to education. When questioning the validity of the aforementioned studies of "Catholic school advantages", many scholars also resorted to ability bias arising from self-selection.

2. Substituting equations (6), (1), and (2) into the parent's utility function (3), one has $U_j = U(H_j, W_j; K)$. The optimal choice of school quality, $\boldsymbol{K}_j^*$, can then be found by solving the first order condition $(\partial U_j)/(\partial K) = 0$.

3. In the "Coleman Report", the between-school variation in student achievement was found to be only a small portion of the total variation. Coleman et al used this finding to support a claim that "schools bring little influence to bear on a child's achievement that is independent of his family background and general social context." (Coleman, 1990, p.119) Even their claim was right, however, the finding does not necessarily implies that schools are not important.

4. The second derivatives of M with respect to X are equal to zero, which means that the M^* corresponding to the solution (12) could be the minimum, rather than the maximum as expected. But since the solution is the only equilibrium, and the school is efficient, one should regard the M^* as the best objective the school can achieve.

5. These results are apparently sensitive to the linear educational technology (7). If the educational technology is not linear, the relationship between the partial and total effects of school inputs may be more complicated, and so may the relative effect of the school to the family. One fundamental result, however, would prevail in the self-selection framework. That is, any educational production function that includes family and school variables simultaneously would fail to find the total effects of school inputs, as well as the independent effect of the school as a whole, on student achievement.

6. Many other well-defined technologies will not change the basic result of the analysis.

7. For simplicity, we neutralize the impact of the students' average learning ability on the production of school quality, and regard a student's family background as including components of his learning ability.

8. When analyzing the "Catholic-school advantages" in improving student achievement, Hoffer, Greeley, and Coleman (1985) find the following: "It is the most disadvantaged Catholic-school students who are most likely to profit from attending Catholic secondary schools--i.e., those from low socioeconomic backgrounds, members of the black or Hispanic communities, those with low sophomore test scores, and those who start off their high school careers with disciplinary problems." This could be evidence on the zero relationship between students' achievement objectives and their family background.