Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data*

By ERIC HANUSHEK
U. S. Air Force Academy

Recent attention to education in the United States represents the merger of concern about efficiency of the educational system and concern about the distribution of educational services, particularly along racial and ethnic lines. However, there is very little guidance on how to satisfy any efficiency or distributional goals through public policy because extremely little is known about the relationship between inputs—particularly inputs available for public policy—and outputs of the educational process. Educational research has been slow in providing definitive answers to public policy questions for several understandable reasons: the subject of the educational process is extremely complex, especially as regards the physiological and psychological aspects; any theoretical development of a learning theory amenable to analysis for policy purposes is absent; and the required data traditionally have not been collected. Previous analyses have yielded some suggestive beginnings, and have provided insights into how the analysis should proceed. This analysis represents a next step of statistical inquiry into the educational process from a public policy point of view. Three fundamental educational policy questions are addressed: (1) do teachers count? (2) are schools operated efficiently now? (3) what characteristics of teachers and classrooms are important? Past studies have given ambiguous answers to these questions, largely due to inadequate data. Specifically, no data set which supplies accurate historical information on educational inputs at an individual level has been available. This study attempts to provide more conclusive answers by remedying the most glaring data problems for a set of students (third graders) in one school district.

* Many people have offered useful suggestions along the way, but special thanks go to Franklin Fisher, John Jackson, John Kain, Herbert Kiesling, Frank Sloan and Finis Welch. This project could not have been undertaken without the help of Wilber Hawkins of the California school district in arranging for data collection and discussing the overall problems. Any remaining errors are, of course, mine. This work was supported by the RAND Corporation, RPN 7502/7510, and The Carnegie Corporation of New York.

I. *Conceptual Model and Data*

The major objective of this analysis was to estimate the relationship between variables which can be controlled by public policy and educational output. The basic conceptual model of the educational process which was used can be depicted by Equation 1.

$$(1) \quad A_{it} = f(A_{it^*}, B_i^{(t-t^*)}, P_i^{(t-t^*)}, I_i, S_i^{(t-t^*)})$$

where

A_{it} = vector of educational outputs of the i^{th} student at time t
A_{it^*} = vector of entering achievement levels at time t^*
$B_i^{(t-t^*)}$ = vector of family inputs to education of i^{th} student cumulative from time t^*

$P_i^{(t-t^*)}$ = vector of peer influences of i^{th} student cumulative from time t^*

I_i = vector of innate endowments of i^{th} student

$S_i^{(t-t^*)}$ = vector of school inputs to i^{th} student cumulative from time t^*

This model simply states that educational output (A_{it}), itself a multidimensional factor, is a function of entering achievement (A_{it^*}) and of the influences of the individual's family ($B_i^{(t-t^*)}$), of the influences of his peers ($P_i^{(t-t^*)}$), of his innate abilities (I_i) and of the cumulative school inputs ($S_i^{(t-t^*)}$) over the period being studied. This abstract model provides a framework for discussion of models of the educational process which can be tested empirically.

Specific variables corresponding to Equation 1 are derived from a combination of past work in the field, theoretical considerations, and sheer data availability. For instance, it is possible to develop many measures of the output of the educational process, such as standardized test scores, juvenile delinquincy rates, post-school income streams, occupational choice or level of education completed. Yet, the availability of data has restricted most past studies of education—and this analysis—to examining a single output. This paper concentrates entirely on an analysis of cognitive development as reflected by reading achievement in the third grade. While it is believed that these scores represent differences which are ultimately valued by society, results in terms of this single measure must be considered tentative until there is confirmation in terms of different outputs.[1]

For public policy purposes, most of the interest in the model centers upon the influence of school characteristics on achievement. However, in order to avoid biased estimates of school effects, the other input vectors must be entered into the statistical analysis. The reasoning behind the inclusion of each of the input vectors is fairly straightforward, and just a brief discussion of specific measures will be given here.[2] It would be possible to specify a model which did not include an initial achievement level. But, without this input, the data requirements are multiplied considerably since the *entire* past history of all inputs must be known.

Families obviously have considerable impact on education through physical conditions, attitude formation and direct involvement in the educational process. Since these factors tend to be highly correlated with socioeconomic status of the family, this aspect of education is proxied by father's occupation and family structure. The influence of peers is much the same as that from the family, and, thus, this aspect is proxied by aggregate measures of the socioeconomic status of individuals in a given class or school. While innate abilities are included in the conceptual model, there is no direct measure of this aspect. However, there is reason to believe that biases in the school parameters due to this missing variable are minimal. First, the model with initial achievement measures the "value added" of various inputs and biases will occur only if the missing portion of innate abilities is correlated with the rate of learning (as opposed to the level). Second, at least for whites, it is reasonable to assume that this

[1] There is scattered evidence on the valuation of achievement in W. Lee Hansen, Burton A. Weisbrod and William J. Scanlon, "Schooling and Earnings of Low Achievers," *American Economic Review*, June 1970; Burton A. Weisbrod and Peter Karpoff, "Monetary Returns to College Education, Student Ability and College Quality," *The Review of Economics and Statistics*, November 1968; and Randall D. Weiss, "The Effects of Education on the Earnings of Blacks and Whites," *Review of Economics and Statistics*, February 1970.

[2] A more complete discussion can be found in Eric Hanushek, *The Value of Teachers in Teaching*, RM-6362-CC/RC (Santa Monica: The RAND Corp., 1970).

factor is captured fairly well in the family background variables. This is the case if innate abilities tend to be hereditary and if social mobility is highly correlated with ability. Severe problems, at least in the school portion of the model, do not arise unless there is a mechanism which leads to the correlation of the "nonhereditary" part of innate abilities and specific school resources.

School influences are the focus of this study and will be discussed in more detail than the other inputs. Surprisingly little is actually known about the ways in which schools and teachers affect education, largely as a result of the past fixation on inputs to education rather than outputs. One can impute a set of hypotheses about teacher effects from the behavior of schools in setting pay schedules on teaching experience and educational levels. They must believe that increased experience and further schooling have a positive relationship to educational output.

Other hypotheses can also be found in the actions of school administrators. For example, many persons argue that some forms of student distributions in the schools and classroom (for example, ability tracking or racial and social integration) have a beneficial effect on education. These are testable hypotheses about the relationship between school inputs and achievement. (It must be borne in mind, however, that the tests are restricted to the range of experiences observed.) Further, in recent literature (particularly *Equality of Educational Opportunity*,[3] or the Coleman Report) there is a suggestion that one can measure other dimensions of teacher and school quality. These include attitudes of teachers and administrators, verbal facility (and perhaps general ability) of teachers, quality of physical plant, quality of teacher education, background of teachers, and more.

Of the several studies of the educational process which have been undertaken, two major shortcomings have persisted. First, it has not been possible to match inputs at the individual level, particularly for schools, with the other inputs and outputs of the educational process. Second, there has been a lack of historical data on inputs; most past studies have relied upon cross-sectional data containing only contemporaneous information about inputs. These data problems have introduced considerable doubt into the conclusions of past studies.[4] A primary objective of this study was to come closer to Equation 1 than had been done previously by eliminating these two sources of data error.

The basic sample of data was drawn from a large school system in California during the summer of 1969. All children in the third grade during the school year 1968–1969 were initially included in the sample. For these 2,445 students, information on family background, scores on the Stanford Achievement Tests, and names of teachers were abstracted from cumulative records. At the same time, all kindergarten through third grade teachers currently in the system were surveyed for information fairly similar to that contained in *Equality of Educational Opportunity*. Information was collected on teacher backgrounds and attitudes, and on specific aspects of schooling. An attempt was made to ascertain their use of time, that is, the division in the classroom between instructional efforts, disciplinary efforts, and administration. Also, a verbal facility test was given each teacher.[5] The sample used for this

[3] James S. Coleman, *et al.*, *Equality of Educational Opportunity* (Washington, D.C.: Government Printing Office, 1966).

[4] For example, Eric Hanushek and John Kain, "On the Value of *Equality of Educational Opportunity* as a Guide to Public Policy" in Frederick Mosteller and Daniel P. Moynihan (eds.) *On Equality of Educational Opportunity* (New York: Random House [forthcoming]).

[5] Edgar F. Borgatta and Raymond J. Corsini, *Quick*

analysis was developed by applying two criteria to this group of all third graders. First, individuals were eliminated from the sample if data were not available on both their second and third grade teachers. Second, students were eliminated if both first and third grade achievement test scores were not available. When these criteria were applied, a total of 1,061 students was left in the sample. (A separate analysis of the effects of moving appears to be called for here, but it is beyond the scope of this paper.)

For analytic purposes three different samples were analyzed. As a first step, whites and Mexican-American were separated. (The latter was the only minority group represented in this school system.) There are two reasons for this stratification: (1) the nominal values of the proxies for background inputs do not necessarily have the same meaning for the two groups, and (2) there is no reason to insist on the same model of the educational process for both groups. The ethnic samples were then divided on occupational grounds—fathers in manual or blue collar occupations and nonmanual or white collar occupations. From this, the following three samples were constructed for analysis: white, manual occupation ($n=515$); white, nonmanual occupation ($n=323$); and Mexican-American, manual occupation ($n=140$).[6]

II. *Do Teachers Count?*

Recently there has been considerable controversy among those analyzing education as to whether teachers count in the educational process. This arises from interpretation of past empirical work, namely the Coleman Report. However, since our sample experience did not include children without teachers, the only testable hypothesis is whether or not there are differences in teachers that lead to differences in achievement among students. In other words, does it matter which teacher a student has, or are all teachers perfectly substitutable?

This test is done by constructing a series of dichotomous variables, T_{ij}, for each teacher in the sample. If the j^{th} student has the i^{th} teacher, $T_{ij}=1$ for him and $T_{kj}=0$, where $K \neq i$. The complete model looks like:

$$(2) \quad A_{j3} = \sum_{i=1}^{n} t_i T_{ij} + aS_j + bA_{j2} = u_j$$

where

A_{j3} = achievement in 3rd grade of the j^{th} student
S_j = 1 if j^{th} student is female; = 0 otherwise
A_{j2} = achievement in grade 2 of j^{th} student

(Other analysis of these data indicates that sample stratification by ethnic and occupational background is an adequate way to allow for family inputs. Thus, no explicit family background measure was included here.) In this formulation it is possible to ask whether the individual classroom coefficients are significantly different from a constant, or whether there are any differences among teachers in the sample in terms of their contribution to achievement gains.

At least two students from a sample had to be in a class with a teacher before the

---

Word Test: Level 2 (New York: Harcourt, Brace and World, Inc., 1964). This test appears to be superior to the test in *Equality of Educational Opportunity* as it appears to give better discrimination among teachers. One complaint voiced about the *EEO* test is that it was too easy. The complete teacher survey can be found in Hanushek, *The Value of Teachers*.

[6] The decision to stratify will be discussed in terms of statistical tests for sample homogeneity in a later section. These samples are not exhaustive. Children with only mothers or no occupation reported for fathers were not included. For whites, these groups totaled 36 students; for Mexican-Americans, these groups plus the nonmanual occupation group totaled 47. These samples were too small to study separately, and, thus, they were ignored.

TABLE 1—*F*-STATISTICS FOR NULL HYPOTHESIS OF UNIFORM TEACHER EFFECTS

Sample	*F*	d.f.[a]	R^2
3rd Grade			
White, manual	2.03[b]	(69, 426)	.71
White, nonmanual	1.57[b]	(57, 247)	.76
Mexican-American, manual	.78[c]	(29, 78)	.68
2nd Grade			
White, manual	2.96[b]	(55, 440)	.68
White, nonmanual	2.39[b]	(48, 264)	.71
Mexican-American, manual	1.09[c]	(26, 82)	.64

[a] Degrees of freedom.
[b] Statistically significant at the .01 level.
[c] Not statistically significant at the .10 level.

student and teacher were included in the analysis. For the three samples tests for differences in the t_i's in both the third and second grade were performed. (For the second grade analysis, the dependent variable is A_{j2} and one exogenous variable is A_{j1}.) The results of these six *F* tests for equality of coefficients are depicted in Table 1. For whites, the hypothesis of no teacher differences is rejected at the 1 percent level. However, for Mexican-Americans it is not possible to reject the hypothesis of no teacher differences at the 10 percent level. In other words, the teacher appears to count for whites of all social strata but not for Mexican-Americans.

One qualification is needed before any further interpretations are made. Since these students had only one teacher during the year, it is impossible at this stage of the analysis to distinguish between the effects of particular teachers and a classroom composition effect. There is no independent observation here for a given teacher with several different classrooms. This problem will be dealt with directly in the next section.

This analysis suggests that the Mexican-Americans at this lower grade level are not getting much out of school. On the average they tend to progress at a rate of about one-half grade level per year or 50 percent of the national average for reading achievement gains, regardless of which teacher they have. It is possible that the classroom composition exactly offsets teacher differences or that teachers are matched with Mexican-American classes to equalize gains. However, this seems highly unlikely since the matching needed to achieve no teacher-classroom effects calls for putting the best teacher-classroom combination with the room of worst "gainers," etc. Moreover, since the white children are sensitive to teacher-classroom differences, as indicated by Table 1, a finding of no differences for Mexican-Americans—when in fact white differences exist—implies that teachers are distributed only in conjunction to the Mexican-Americans in the class. Yet, the proportion of Mexican-American ranges from 6 percent to 63 percent in the 30 third grade classrooms that have more than 2 Mexican-American students.

III. *Characteristics of Teachers*

The preceding section suggested that the performance of white students is dependent upon the specific teacher and classroom associated with the student. For policy purposes it would be useful to identify the characteristics that contribute to increased performance. This phase of the analysis was accomplished by introducing a variety of quantitative teacher and classroom characteristics into an overall model of student achievement. This was done for both the white manual occupation and white nonmanual occupation samples.

The first step was to estimate models which included only the "pay parameters" of teacher experience and hours of graduate education to measure schools. After allowing for differences in entering achievement, family background, and school composition, these variables for second and third grade teachers were always statistically insignificant; all *t*-statistics were less

TABLE 2—VARIABLE DEFINITIONS, MEANS AND STANDARD DEVIATIONS

Variable	White, Manual		White, Nonmanual		Definition
	Mean	Stnd Dev	Mean	Stnd Dev	
A_3	55.74	19.1	64.82	16.8	Stanford Achievement Test raw score—3rd grade
F	.50	.5			Sex: =1 for female =0 for male
R	.08	.3			Repeat grade: =1 if a grade was repeated; =0 otherwise
A_1	35.17	15.1	42.43	15.8	Stanford Achievement Test raw score—1st grade
D	17.93	18.8			% of time spent on discipline by 3rd grade teacher
T_3	66.90	15.8			*Quick Word Test* score—3rd grade teacher
Y_3	1.91	1.6	2.02	1.7	Years since most recent educational experience—3rd grade teacher
T_2	68.41	19.0			*Quick Word Test* score—2nd grade teacher
Y_2	2.64	2.6	1.88	1.7	Years since most recent educational experience—2nd grade teacher
C			.19	.4	Clerical occupation; =1 if father in clerical job; =0 otherwise
S_3			7.85	8.1	Years of experience with this socio-economic level—3rd grade teacher
S_2			7.94	6.1	Years of experience with this socio-economic level—2nd grade teacher

than 1.25. This indicates that schools are now operating inefficiently but gives minimal guidance to school administrators. The remainder of the paper attempts to identify the aspects of schools and teachers which are important in education.

The estimates for the white manual sample are displayed in Equation 3. Variable definitions, means, and standard deviations appear in Table 2.

$$(3)\quad A_{j3} = 20.8 + \underset{(2.3)}{2.81F} - \underset{(-2.8)}{6.38R} + \underset{(18.8)}{.79A_{j1}} - \underset{(-2.1)}{.07D} + \underset{(2.4)}{.09T_3} - \underset{(-1.5)}{.57Y_3} + \underset{(1.9)}{.06T_2} - \underset{(-2.9)}{.68Y_2}$$

$$R^2 = .51 \qquad SE = 13.5$$

(*t*-statistics are displayed below each coefficient; *SE* is the standard error of estimate.)

This model presents an interesting view of teachers. The teacher characteristics that appear to be important are not the characteristics that are purchased by schools. For both the second and third grade teachers, the score on the verbal facility test (T) and the recentness of education (Y) are the most important factors.[7] Additionally, there is a "quasi teacher" characteristic of percent of time spent on discipline by the third grade teacher. Each of these has important implications for school operations.

The verbal facility test (T) probably

[7] The model was not constrained to have the same characteristics for second and third grade teachers; this results from the analysis of various characteristics without constraint. If we test the joint hypotheses that all four strictly teacher characteristics together have no effect on education, we reject at the .01 level with $F_{4,506}=5.68$.

TABLE 3—SIMPLE CORRELATIONS FOR TEACHER CHARACTERISTICS—WHITE MANUAL SAMPLE

	D	T_3	Y_3	T_2	Y_2	$Exper_3$	$Units_3$	$Exper_2$	$Units_2$
D	1.00								
T_3	−.19	1.00							
Y_3	.01	.08	1.00						
T_2	.07	.19	.13	1.00					
Y_2	−.14	−.09	.11	−.19	1.00				
$Exper_3$	−.14	.37	.11	.17	.05	1.00			
$Units_3$	−.09	.01	−.14	.09	.12	.53	1.00		
$Exper_2$	.09	−.11	−.01	−.09	.20	−.18	−.11	1.00	
$Units_2$	−.02	.07	.01	−.03	−.15	−.02	.03	.43	1.00

plays two roles: first, it is a measure of communicative ability; second, as the authors of the test point out, it can be taken as a quick measure of overall intelligence or general ability. Thus, general ability seems important, regardless of formal training. There are some important policy implications surrounding the verbal test measure of teacher quality. By interchanging teachers at the top and bottom of the verbal ability scale for this system, achievement changes by .2 to .4 grade levels.[8] This is very significant given the powerful effect of a student's early education on later achievement as evidenced both by the increasing grade level disparities in the Coleman Report and by the strength of first grade achievement in Equation 3. Further, since this test has national norms, it is possible to get some idea of how the teachers being hired in this system rate alongside other college graduates. The mean score of 68 places the teachers in this sample slightly under the median for female college graduates. Thus, this system is not being successful in attracting the best people.

In addition to teacher ability as measured by the verbal facility test, the recentness of educational experiences (Y) has a significant effect on educating students. This seems to provide the rationale for encouraging or requiring teachers to return to school periodically. However, as indicated by the results of analyzing graduate units and the effects of Master's degrees, it does not really matter whether the teacher is enrolled in an advanced degree program or is taking many courses. Education of the second and third grade teacher in the last year as opposed to five years ago would be worth .2 to .3 years of reading achievement to a given third grader.

Finally, there is the measure of discipline time (D) that was labeled as a quasi-teacher characteristic. Certainly, an interaction between the classroom and the teacher is reflected in this variable. However, as expected, the more time spent on disciplinary matters, the lower the achievement level of the class. It does suggest that efforts to reduce such time could be beneficial. These efforts would include using principals or assistant principals or even teacher's helpers as disciplinarians.

It is immediately obvious that these are not the characteristics of teachers that are currently being purchased. Certainly if there is an excess supply of teachers, schools can be selective in hiring and can attempt to evaluate the general ability of teachers. However, casual observation suggests that the most selective (suburban) systems weight previous teaching experience heavily. Moreover, as suggested by the simple correlation matrix for teacher characteristics displayed in Table 3, the

[8] This is calculated by changing only the third grade teacher verbal score for the lower limit and both second and third for the upper limit. The scores are changed from 40 to 96 to represent the range found in the data. (Maximum score is 100.) The resulting achievement score is then converted to grade level equivalents.

purchased factors for the third and second grade teachers (experience and units of graduate work) are not highly correlated with the characteristics included in the model.

A model for the white nonmanual population was also estimated.[9] The results of this analysis, shown in Equation 4, provided a different set of teacher characteristics that seemed important. (Variable definitions, means, and standard deviations are displayed in Table 2.)

$$
\begin{aligned}
(4)\quad A_3 = 35.0 &+ .72A_1 - 5.1C \\
&\quad (16.0) \qquad (-3.0) \\
&- .79Y_3 + .10S_3 \\
&\quad (-1.9) \qquad (1.2) \\
&- .66Y_2 + .20S_2 \\
&\quad (-1.7) \qquad (1.8) \\
R^2 = .52 &\qquad SE = 11.8
\end{aligned}
$$

The characteristics which seem important for the white children from nonmanual occupation families are not entirely the same as for children from families with manual occupations. The recentness of education is again a significant factor, with approximately the same effect here as in the previous model. However, while teacher verbal ability does not appear to be significant, experience with this socioeconomic group assumes importance. As would be expected, the correlation between total experience and socioeconomic group experience is quite high. The simple correlation for third grade teachers is .8. Thus, the present policies of paying for experience could be reasonable in this case. Recentness of education is, however, only slightly correlated with the pay factors, as in the manual occupation sample.

In comparison to the manual occupation sample, the coefficient estimates are not as reliable in the nonmanual sample. Although there is a smaller standard error of estimate for the nonmanual model, this is coupled with a smaller variance in overall achievement; the R^2's in the two models are almost equal. The estimated coefficients in Equation 4 do indicate that teachers have less effect on these nonmanual children. Although the effects of recent education are roughly the same in the two models, the potential for change in achievement through increasing verbal facility, or general ability, is considerably greater. Not only can verbal facility be changed more rapidly—since experience usually comes by aging rather than hiring—but given percentage changes in verbal quality have a considerably larger effect on student achievement.

The previous discussions of Equations 3 and 4 must be taken within the context of the overall model. In the process of developing these models, several other hypotheses about educational inputs were tested and rejected. In particular, variables measuring school composition in terms of occupational distribution, ethnic distribution, and achievement distribution; variables measuring objective background characteristics of the teachers such as socioeconomic background, college major, and membership in professional organizations; and variables measuring subjective factors such as attitudes toward types of students were tested and found to

[9] A word on sample stratification is necessary. A formal test for equality of coefficients between the two white models was performed. The methodology of this test can be found in Franklin M. Fisher, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," *Econometrica*, March 1970. When testing the entire model and restricting the models for both samples to the form of Equation 3, the hypothesis of coefficient equality was rejected at the .025 level ($F_{9,820}=2.13$). However, since the principal interest centers upon the teacher characteristics, a test of just this subset of variables seems more appropriate. When this test is performed, the results are inconclusive since $F_{4,820}=1.57$ when the critical value for those degrees of freedom at the .10 level is 1.94. Thus, it is not possible to reject the hypothesis of homogeneity with a high level of confidence; yet, at the same time the evidence does not seem strong enough to pool the sample. Since both samples are quite large, the loss in efficiency by not pooling would not be large, and the decision was made to stratify.

have statistically insignificant effects on the students' achievement. The implication that arises from both these "nonresults" and from the models presented is that we do not have very good measures of teachers. We can identify a few objective factors that appear to affect education. Yet most of our notions about important attributes of teachers are probably too simple. Although teachers do appear important in the model, precise decision rules for hiring teachers are not readily available at this time. It appears that objective characteristics of teachers and classrooms are insufficient and that it is necessary to measure behavior in the classroom better.

IV. *Summary and Conclusions*

From this study three conclusions are apparent. It should be borne in mind that these conclusions derive from a sample of one school system; therefore, some caution should be used in generalizing to other systems. First, the present set of hiring practices leads to an inefficient allocation of resources. The analysis indicates that teaching experience and graduate education do not contribute to gains in student achievement scores. Moreover, the characteristics that do matter are not highly correlated with these factors. Yet these attributes are being purchased by the school district. Since turnover is costly, some average experience level over one year would be reasonable. However, the current average of over 11 years is certainly excessive. Second, teachers do not appear to count for Mexican-American students in the sense that different teachers and different classroom compositions do not affect the achievement outcome of the Mexican-American students. This might well be a language problem, but there is no direct measure of this. Third, the attempts to provide a set of measurable characteristics which schools could hire and control to affect achievement did not produce clear-cut answers. There is a considerable part of teaching that cannot be explained by a set of fairly standard variables measuring teachers and classrooms, particularly for whites in white collar families.

It would be imprudent to generalize too much from these findings. They refer to one school system, one measure of output, and one particular grade level in elementary schools. For this reason this study is best looked upon as being suggestive rather than definitive; as being a prototype rather than a final analysis. Yet the evidence is beginning to mount and seems to indicate fairly consistently that the past ways of operating schools leave something to be desired.