DOCUMENT RESUME

ED 089 413 EA 006 000

AUTHOR	McAfee, Jackson K.
TITLE	A Mathematical Model for Allocation of School Resources to Optimize a Selected Output.
PUB DATE	Apr 74
NOTE	36p.; Paper presented at American Educational Research Association Annual Meeting (59th, Chicago, Illinois, April 15-19, 1974)
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	Decision Making; *Educational Economics; Educational Planning; *Educational Programs; Educational Research; Efficiency; Expenditure Per Student; *Mathematical Models; *Productivity; Program Budgeting; *Program Costs; *Resource Allocations; Speeches

ABSTRACT

The methodology of costing an education program by identifying the resources it utilizes places all costs within the framework of staff, equipment, materials, facilities, and services. This paper suggests that this methodology is much stronger than the more traditional budgetary and cost per pupil approach. The techniques of data collection are presented. Also discussed are the use of mailout questionnaires and site visit interviews to complement each other, the use of standardized and actual costs, the problems of defining and delineating a program within a larger overall program, and the problems of estimating the research, development, and annual operating cost associated with each measure element. (Author)

ED 089413

A MATHEMATICAL MODEL FOR ALLOCATION OF SCHOOL RESOURCES TO OPTIMIZE A SELECTED OUTPUT

by

Jackson K. McAfee

Presented at the Annual Conference of the
American Educational Research Association,
Chicago, Illinois, April 15-19, 1974.

EA 006 000

Educational decision-makers are faced with the task of allocating school resources in such a way as to maximize student outcomes. The problem of how best to proceed is complicated by the question of priorities among the outcomes to be attained; and because of the multiplicity of student outcomes, it is not likely that all outcomes can be maximized at once. Also, aside from questions of technical efficiency, decision-makers are constrained in their choice of plans by a limited budget and by social, political and legal forces. Thus, the process of maximizing outcomes is irrevocably placed in the context of planning the best available use of scarce resources. The desire for rational decision-making entails that decision-makers have some empirical notions of how the various inputs into the educational process relate to student outputs. The prices of school inputs must be known if the question of allocative efficiency is to be considered.

Given this background, the rationale for the present study was a response to the need for studies that demonstrate the utility and limitations of mathematical models as aids in improving educational planning and decision-making. The problem of the study was to develop a mathematical model to facilitate the decision-making process in selected areas of educational activity by optimizing the allocation of scarce resources and to empirically illustrate its application. In development of the model attention was given to:

1. Determining the educational production function the describes the input-output relationship between selected variables.

2. Determining the optimal combination of inputs to maximize output subject to certain costs and other constraints.

3. Determining the optimal combination of inputs to maximize output, subject to certain constraints, given increments in the budget constraint.

To accomplish the above, first required that the process of education be conceptualized. Mathematical models do not evolve in the absence of a conceptual framework. Basically, the process of education was viewed as a function of the various inputs into the process. These inputs include student, community and school factors. The first two are regarded as essential to the conceptualization of the process, as these factors necessarily interact with, indeed may be causal determinants of, the various school factors. Given this interaction among the various input factors together with presumed causal directions, the process of education results in certain outcomes. Among these are cognitive achievement, affective growth and physical development.

With this conceptual framework, the translation to a mathematical model occurred. Concurrent with this translation was the desire to allocate school resources in such a way that would maximize certain student outcomes. Thus a mathematical model was first required to estimate the educational production function and the budget equation. Given the production function as an objective function and the budget equation as one of the constraints, an optimization strategy was employed. Other constraints can be based on empirical evidence or the subjective desires of decision-makers. Also, the constraints

should include minimum standards to be achieved by other outputs and bounds on the observed variation in the independent variables.

METHODS

In the empirical application, data were utilized to illustrate the model. To estimate the production function and the budget equation, cross-sectional data were used in a survey-type design. The data were collected in a study of school productivity under the direction of the National Educational Finance Project (NEFP).[1] The data consisted of aggregate measures on input factors in 181 local school districts in a given state. Twenty-four variables were measured providing information on (1) student and community inputs, (2) median reading achievement at the sixth-grade level, and (3) selected school resources.

Description of the Variables

Variables were selected from the NEFP data that were descriptive of both in-school and out-of-school factors of each school district. Although there have been many variables identified in related research the variables selected for this study were variables used previously by researchers to examine correlates to school performance or represented variables that were selected by the Research Staff of the National Educational Finance Project as potential predictors of school productivity.

Data for the following variables were compiled from available records at the state department of education for a given state, except for variables x_1 through x_3:[2]

Income per-pupil (x_1)--Data for x_1 were taken directly from Personal Income by School Districts in the United States.

Income under $3,000 ($x_2$)--The percentage of gross income less than $3,000 was computed by totaling the number of tax returns per district and dividing this total into the number of returns reporting gross income less than $3,000.

Income over $10,000 ($x_3$)--The percentage of gross incomes over $10,000 was computed by totaling the number of tax returns per district and dividing this total into the number of returns reporting gross incomes over $10,000.

ESEA Title I Pupils (x_4)--x_4 was computed by forming a ratio of numbers of pupils eligible for Title I programs to pupils in ADM. ADM was used as the denominator rather than ADA because the attendance habits of the two groups (ESEA Title I Pupils and total pupils) would not necessarily be the same.

Minority enrollment (x_5)--The district percentages of pupils enrolled, during 1968-69, that were nonwhite, Spanish speaking, Oriental or American Indian were obtained directly from the state department of education.

Attendance (x_6)--x_6 was calculated by forming a ratio of ADA to ADM. The ADA and ADM were 1968-69 school year figures.

Future training (x_7)--The percentage of graduates receiving post high school education was computed by forming a ratio of the the number of 1969 graduates entering future training to total 1969 graduates.

Size of school district (x_8)--ADM for the 1968-69 school year

was used as the indicator of school district size.

Percentage enrolled (x_9)--The percentage of children age 5-17 enrolled in putblic school was calculated from information contained in the fall 1968 school census report filed by each district with the state department of education.

Transportation cost (x_{10})--Transportation cost per pupil was computed by dividing the 1968-69 school year cost for transportation by the number of pupils in ADA.

Local fiscal effort (x_{11})--Local fiscal effort was determined by forming a ratio of local revenue per pupil in ADA to the adjusted gross income per pupil in ADA. Local revenue per ADA was computed by dividing total local revenue by the number of pupils in ADA during the same year.

Expenses of instruction (x_{12})--X_{12} was computed by taking the percentage of total current expense disbursed for instruction during the 1968-69 school year.

Longevity experience(x_{13})--X_{13} was calculated by forming a ratio of teachers with 20 or more years of experience to the total number of teachers for the 1968-69 school year.

Teacher preparation (x_{14})--X_{14} was computed by forming a ratio of teachers with less than four years training to the total number of teachers for the 1968-69 school year.

Teacher experience (x_{15})--X_{15} was calculated by forming a ratio of teachers with less than five years experience to the total number of teachers for the 1968-69 school year.

Advanced preparation (x_{16})--X_{16} was computed by forming a ratio

of teachers with either an advanced degree or 30 hours of professional training beyond their Bachelor's degree to the total number of teachers for the 1968-69 school year.

Median teacher salary (x_{17})--Variable x_{17} was the 1968-69 median teacher salary for each school district.

Average class size (x_{18})--Average class size for the 1968-69 school year was determined by dividing the number of district pupils in ADA by the number of classroom teachers in the district.

Pupil-support personnel ratio (x_{19})--The pupil-support personnel ratio was calculated by forming a ratio of district pupils in ADA during the 1968-69 school year to the number of certified non-teaching personnel employed in the district for the same year.

Expenses for transportation (x_{20})--X_{20} was computed by taking the percentage of total current expense disbursed for transportation during the 1968-69 school year.

Median reading achievement (x_{21})--X_{21} was the median score for the school district during 1969 on a standardized reading achievement test for sixth-grade pupils developed by the state department of education.

Average daily attendance (x_{22})--X_{22} was the average daily attendance for the school district for the 1968-69 school year.

Total current expenditure (x_{23})--X_{23} was the total current expenditure for the school district for the 1968-69 school year.

Instructional expenditures per-pupil (x_{24})--X_{24} was computed by forming the ratio of expenses of instruction to ADA (x_{22})

The only output included in the present list of variables was

the sixth-grade median reading achievement score for the district (x_{21}). While there is some concern over whether or not cognitive skills are the most useful outcomes of schooling, knowing something meaningful about success in the area of cognitive scores is far superior to knowing nothing at all; and this knowledge should allow some general inferences that would be most helpful in pointing general policy directions.

Repeatedly, the necessity for knowing the ranges of each of the variables in the observed data has been indicated. This was essential since predictions cannot be extrapolated beyond the range of observed variation. At the same time, the means and standard deviations of each variable are listed to provide additional information regarding the variability of the data. This information is shown in Table 1. It was believed that these data provided the widest possible variation in existing data for a given state.

The Data Analysis Plan

The first step in the data analysis involved estimating the production function and the budget equation using multiple regression. A program from Biomedical Computer Programs (BMD) was used.[3] The particular program is identified as BMD02R - Stepwise Regression. The program computes a sequence of multiple regression equations in a stepwise manner. At each step one variable is added to the regression equation. The variable added is the one which has the highest partial correlation with the dependent variable partialed on the variables which have already been added. Equivalently, it is the variable which, if it were added, would have the highest F value. In addition,

variables can be forced into the regression equation. Regression equations with or without a regression intercept may be selected. This program proved highly satisfactory for several reasons. First, of the variables containing redundant information (highly related), stepwise regression selected only the ones that were optimally related to the dependent variable, omitting the remaining variables from the equation. Secondly, since most school variables were wanted in the equation, these variables were forced. Third, the stepwise procedure allowed the contribution to the multiple R of each variable as it entered the equation to be viewed (i.e., what has been gained by allowing this variable to come into the equation).

The process of estimating these two equations was exploratory in nature. Thus, as a first step all linear terms were considered, some being forced into the equation. Secondly, all quadratic and cubic terms were tested. Then selected interaction terms were tested. It was not possible to examine all interaction terms simultaneously. In a sample of 181, the maximum number of terms that the regression equation could have was 180. Yet with 20 independent variables there were 190 possible interactions. Thus analysis was limited to those interactions that were logically viewed as having potential impact on school achievement (community-school and school-school interactions).

To determine whether a variable that was tested made a significant contribution involved making a choice as to the significance level to be employed. For those terms that were free to come into the equation (i.e., non-forced), it was felt that they should enter only if a noteworthy contribution was being made. The investigator observed

that by setting the F-to-enter at 6.80 (${}_{.99}F_{1,160} = 6.80$), variables coming into the equation made an addition to the multiple R of generally 1 or more percent. It was felt that in latter steps of the regression analysis, any variable that could make a contribution of 1 or more percent should enter the equation.

Finally, the analysis ended with nonlinear functions consisting of certain forced linear terms and any other terms which made a significant contribution to the multiple R (linear, higher degree or interaction terms).

Given the production function and the budget equation, the second major stage of the data analysis was solving a mathematical programming problem in which the production function was the objective function and the constraints included the budget equation and bounds on the observed variation in the independent variables. A program for solving nonlinear programming problems comes from the SHARE Program Library.[4] The particular program is SDA 3189-SUMT. The purpose of this program is to solve nonlinear mathematical programming problems where the objective function and constraints may be nonlinear. The program uses SUMT (Sequential Unconstrained Minimization Technique) to solve the mathematical programming problem.[5] Users must supply a subroutine to read in the problem data and three subroutines to evaluate the problem functions, and their first and second partial derivatives. If the objective function is not concave, then it is necessary to start the algorithm at various points in the feasible domain of solutions.

RESULTS

Estimating the Production Function

The school inputs used in the production function described only a limited area of educational activity. In particular, certain teachers characteristics were considered: experience, training and level of salary. Also, classroom size, pupil-support personnel ratio, and the size of the district were used. Table 2 gives the results of the analysis for estimating the production function.

Given the various inputs (i.e., student,community and school) into the educational process, the following non-school inputs were found to make significant contributions in explaining the variation in reading achievement at the sixth grade level:

1. Percentage of gross incomes over $10,000 in the school district.
2. Attendance: ratio of ADA to ADM.
3. Percentage of graduates receiving post high school education or training.

Also, the following interactions of non-school inputs with school inputs were found to be significant:

1. Interaction of the percentage of ESEA Title I students with the pupil-support personnel ratio.
2. Interaction of the percentage of minority enrollment with average class size.

School inputs were forced into the regression equation since estimates of the effect of each were needed. The above led to the estimation of a production function (reading achievement-dependent variable) with an $R^2 = 0.79$.

The regression coefficient for the percentage of teachers with less than four years' training (x_{14}) was as expected ($b= -1.511$). However, this percentage interacted with median salary (x_{17}). The interaction coefficient stated that increases in median salary for a given percentage of x_{14} were associated with increases in achievement. In fact, for x_{17} = $10,000, an increase of 1 percent in x_{14} was associated with an increase of 0.099 units in achievement. As in all such cases where trouble is suspected, recourse was made to a frequency distribution. Table 3 reveals the results of such analysis. What was observed was that as the salary level increased the range of x_{14} became more and more restricted. Thus when

$100 \leq x_{17} \leq 110$,

then $0 \leq x_{14} \leq 12.5$.

If the interaction is to be meaningful when $x_{17} = 105$, then x_{14} must be in the interval from 0 to 12.5. This fact became more crucial when the stage of programming analysis was considered.

Estimating the Budget Equation

Basically, the same procedure was followed for estimating both the budget equation and the production function. The total current expenditure for instruction per-pupil (x_{24}) was the dependent variable. A per-pupil cost variable was preferred since using total cost leads to scaling difficulties. Since most of the school variables used in the study were considered instructional expenditures it was felt that per-pupil instructional expenditures was a better choice for a dependent variable than per-pupil total current expenditures (x_{23}). The results are given in Table 4.

School variables were forced into the equation so that the prices of each could be estimated. Only one non-school variable made a significant contribution, namely the percentage of minority enrollment. As the percentage of minority enrollment increased, per-pupil cost increased quadratically rather than linearly. Two school variables had significant square terms, namely percentage of teachers with advanced preparation and median teacher salary. The budget equation was estimated with an $R^2 = 0.94$.

In summary, it appears that the budget equation estimated the prices of the school inputs reasonably well. Furthermore, the coefficients gave the price of each variable while other variables were held constant. This was a necessary requirement for the programming analysis.

Formulation of the Mathematical Programming Problem

In solving a mathematical programming problem the educational production function becomes the objective function and the constraints include the budget equation as well as bounds on the observed variation of the independent variables. Model specification of the production function requires community and student inputs as well as school inputs. However, community and student inputs cannot be manipulated by school authorities. These inputs cannot be considered as variables in the programming analysis. Rather, the objective is to maximize the production function for a given set of community and student inputs. More specifically, given the production function:

$A_i = g(C_i, S_i, I_i, P_i),$

where A_i =vector of educational outcomes of the i^{th} student,

C_i = vector of community inputs relevant to the i^{th} student,

S_i = vector of school inputs relevant to the i^{th} student,

P_i = vector of peer influences,

I_i = vector of initial endowments of the i^{th} individual,

then for a given set of community, peer-group, and individual inputs, the above can be transformed to:

$$A_i = g(k, S_i)$$

where k is a constant. Thus for a given set of school inputs, the proglem is to

$$\text{Maximize } A_i = g(k, S_i)$$

subject to $\sum f_i(k,S_i) \leq B_i$ where the f_i are constraints.

Since the variables were macroscopic (i.e., level of data treatment was the district), individual inputs were deleted from the model. Interest centered on the analysis of educational policy for the population as a whole. Thus, for community and student inputs the mean values of the relevant variables were used. Statistically, this had an enormous benefit, since in the production function the mean level of achievement for the complete population could be predicted perfectly (i.e., no residual error occurs).

Thus, the school inputs (x_{12} - x_{19}) should become the variables in the programming analysis, all other variables assuming constant values equal to the population means. Further reflection revealed, however, that variable x_{12} (percent of total current expenditure for instruction) was not manipulable in this analysis. All other school variables (x_{13} - x_{19}) related to instructional resources and the

dependent variable in the budget equation was total instructional expenditure per-pupil (x_{24}). Thus, ways of reallocating resources in the instructional category with a given budget for instructional expenses were being considered. This does not change the percentage of total current expenditure going to instruction, since the instructional budget remains the same. Thus in the analysis x_{12} remained constant.

Another difficulty centered on the interaction term, $x_{14}x_{17}$ (PREP·MEDSAL). Preliminary data analysis revealed that the maximum observed value for x_{14} (32.1) should be used with x_{17} set at 122.5. Nevertheless, an examination of the frequency distribution did not support this conclusion since the range of x_{14} became more and more restricted as x_{17} increased. This problem was discussed earlier. The problem was resolved by making x_{14} constant.

School district size, as determined by average daily attendance (ADA - x_{22}), was potentially manipulable to the extent that consolidation of school districts can and should occur. Earlier no optimal scale of economy with respect to district size and per-pupil cost was revealed. It was observed that the larger the district the lower the per-pupil cost. However, achievement was negatively related to district size ($b = -0.004$). It is possible, therefore, that there would be an optimum size for the school district for maximizing achievement, constrained by cost. Preliminary data analysis did not suggest any solution within the feasible domain. Thus x_{22} can be set at various levels and the effects of different district sizes on achievement and the distribution of resources for a given per-pupil budget can be observed.

The foregoing suggests that analyses can be conducted at various levels of x_{14} (PREP $\leq$ 4) and x_{22} (ADA). In addition to the constraints previously described it was also observed that the sum of variables x_{13} (EXP $\geq$ 20) and x_{15} (EXP $<$ 5) must always be less than 100 and that the same held true for x_{14} (PREP $<$ 4) and x_{16}(ADPREP). With the above in mind, then, the mathematical programming problem to be solved reduced to the following:

Let $x_{21} = \phi(x) + 32.744 - 1.511x_{14} - 0.004\,x_{22}$.

Maximize

$$\phi(x) = -\,0.042x_{13} - 0.001x_{15} + 0.120x_{16} + 0.002x_{17} + 0.016x_{14}x_{17} + 0.179x_{18} + 0.002x_{19}$$

subject to:

I. (1) $-0.364x_{13} + 0.409x_{15} - 8.817x_{16} - 14.256x_{17} - 6.662x_{18} - 0.032x_{19} + 0.091x_{16}^{2} + 0.090x_{17}^{2} \leq -862.286 + 0.674x_{14} + 0.358x_{22}$

(Budget Equation)

II. (2) $x_{13} + x_{15} \leq 100$

(3) $x_{14} + x_{16} \leq 100$

III. (4) $1.7 \leq x_{13} \leq 52.6$

(5) $14.7 \leq x_{15} \leq 60.6$

(6) $22.2 \leq x_{16} \leq 84.3$

(7) $69.5 \leq x_{17} \leq 137.5$

(8) $12.3 \leq x_{18} \leq 22.2$

(9) $44.9 \leq x_{19} \leq 898.3$

<u>Scaled Variables</u>

x_{17}: 1 unit = \$100

x_{22}: 1 unit = 1,000 pupils

<u>Constants</u>

x_{14}, x_{22}

Since $x_{21} = \phi(x)$ + constant, it follows that in maximizing $\phi(x)$, x_{21} is also being maximized. For that reason programming algorithms never have a constant term in the objective function, since it is not needed and its omission simplifies the programming process. Thus a simple transformation of variables can always eliminate the constant term in the objective function. Of importance was whether or not the objective function was concave. Given that x_{14} = constant, $\phi(x)$ was a linear function. Linear functions are always concave. Thus $\phi(x)$ was a concave function. This meant that any local optimum determined by the algorithm would also be a global optimum.

Results of the Programming Analysis

In order to expedite the comparisons that were to be made with each "run" of the analysis, Table 5 lists the variables manipulated in the programming analysis along with the mean values observed in the population. Obviously, after each optimization the results can be compared against the mean values.

For the first "run" x_{14} (PREP 4) and x_{22} (ADA) were set at their mean values. The results were

x_{21}	x_{13}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
48.0	1.7 LB	14.7 LB	73.5 OPT	9950 OPT	22.2 UB	898.3 UB

LB - Lower bound of the variable

UB - Upper bound of the variable

OPT - Optimal value of the variable (LB $\leq$ OPT $\leq$ UB).

The analysis suggested that an optimal allocation of resources among

the given variables would predict an increase in achievement (x_{21}) of 4.9 units. This represented an increase of 0.71 standard deviations. Experience did not seem to count (x_{13}, x_{15}), nor did class size (x_{18}) or the pupil-support ratio (x_{19}). Advanced training (x_{16}) was deemed most important of the variables affecting achievement; salary (x_{17}) counted also. It should be noted at this point that in a programming analysis, variables are assumed to be independent of each other, that is, a change in one should not automatically cause change in others. Thus, given the existing salary increments for advanced training, an increase in x_{16} (ADPREP) was likely to lead to an increase in x_{17} (MEDSAL). The analysis could not reveal what that change would be. However, increases in x_{17} can be viewed as that accruing beyond that due to x_{16}. To express it another way, the above suggests that the median salary could still be increased by 0.69 S.D. (perhaps by raising the base salary) after increases in median salary due to increases in advanced preparation had been attributed.

Some may suggest that by allowing the pupil-support ratio (x_{19}) to go to its maximum value, other outcomes (most likely affective ones) would be affected. This could be an instance of increasing one output at the expense of another. With the model, though, minimum standards to be achieved by other outputs can be set. Thus a pupil-support ratio of 136.5 (population mean) may be necessary to maintain minimum standards on other outputs. Consider the second "run," then, where x_{19} = 136.5.

x_{21}	x_{13}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
45.7	1.7	14.7	70.1	9670	22.2	136.5
	LB	LB	OPT	OPT	UB	FIXED

By requiring a maximum pupil-support ratio of 136.5, achievement increased by 0.38 S.D. as against 0.71 S.D. previously. More resources were expended to maintain this level for x_{19}, resources that take away from advanced training and salary with consequent effects on reading achievement.

Likewise, one could argue that increasing classroom size to its maximum value would negatively affect other outcomes. Consider then if an average class size of 18.5 (population mean) is maintained.

x_{21}	x_{13}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
44.1	1.7 LB	14.7 LB	65.8 OPT	9320 OPT	18.5 FIXED	136.5 FIXED

Achievement increased by 0.15 S.D. By "trading off" only on the resources expended for experience, less is available to increase x_{16} and x_{17}. Actually, there is evidence to suggest that decision-makers, given the choice of reducing class size or increasing salaries, have chosen to increase salaries.[6] Thus in future "runs" class size assumed its maximum value.

The runs considered so far have had x_{14} (PREP $\angle$ 4) equal to its mean value. Consider what happened when $x_{14} = 0$:

x_{21}	x_{13}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
46.1	1.7 LB	14.7 LB	75.4 OPT	7970 OPT	22.2 UB	136.5 FIXED

Comparing this with a previous "run" a greater predicted increase in achievement (0.44 S.D.) was found as expected, that is, decreases in x_{14} (PREP $\angle$ 4) should be associated with increases in achievement ($b = -1.511$). Looking at x_{17} (MEDSAL), However, a decrease from the mean value of 9,150 was noted. This was due to the interaction term

$x_{14}x_{17}$ which was discussed earlier. The effect that salary had on achievement interacted with x_{14}. The greater the percentage of teachers with less than four years training, the more important the level of salary became. However, when $x_{14} = 0$, the b weight for salary was only 0.002, which was small compared to advanced training (b = 0.120). No wonder then that analysis suggested that more money be spent on x_{16} (ADPREP). One interpretation of the above could be that by reducing the base salary by a given amount, this money could then be expended for teachers with advanced training. As to what the overall median salary would be as a result is not clear. If decision-makers were reluctant to decrease the base salary, then the model allows a lower bound to be set on the salary level and the problem can be resolved.

The effect that district size had on achievement and the distribution of resources was considered. Throughout, it was assumed that $x_{14} = 5.87$ (population mean). Table 6 gives the results of analysis at various levels of x_{22} (ADA). The table shows that by increasing the size of the district, the per-pupil cost was lowered and hence within a given budget more money could be expended on x_{16} and x_{17} with the consequent effects on achievement. Johns and Morphet concluded that the optimal size of a school district should be approximately 50,000 students.[7] In going from the population mean of 9,180 to 50,000, with the same per-pupil budget, the percentage of teachers with advanced training (x_{16}) could be increased by 2.1 percent and the median salary (x_{17}) could be increased \$170 above that due to increases in x_{16}.

To complete the final stage of the data analysis the per-pupil budget was incremented by fixed amounts. It was assumed that x_{22} = 9,180 and two levels of x_{14}, namely x_{14} = 0 and x_{14} = 5.87 were considered. Table 7 gives the results of the analysis. The interpretation of these results presented some problems. First, we note that increasing the per-pupil budget by $100 had a greater effect on achievement when x_{14} = 5.87 than when x_{14} = 0 (x_{21} = 48.1 as against x_{21} = 47.2). That this was so is obscured by the following. At some point in increasing the budget $100, for the case where x_{14} = 0, x_{16} reaches its maximum value. From that point on, only salary (x_{17}) can make a contribution, but its effect was far less than that of x_{16}. That may be why x_{21} was greater when x_{14} = 0 than when x_{14} = 5.87 and the budget increment was $0, but was reversed when incremented $100. This would suggest that the lower half of the table is more easily interpreted than the upper half. At any rate, when x_{14} = 5.87, a budget increment of $100 raised achievement from 45.7 to 48.1 (increase of 0.35 S.D.). In incrementing $200, achievement increased to 49.7. The rate of increase was less for the second $100. This is so, because x_{16} reached its maximum value sometime during the second $100 increase. With only salary left to make a contribution, the rate of increase became far less.

This suggests a practical problem in incrementing the budget. When highly contributing variables reach their maximum values what would happen if extrapolation were made beyond the observed range of values? Given that x_{16} (ADPREP) makes a major contribution, it could be that extending x_{16} beyond its observed maximum (84.3) would continue to result in similar contributions. Of course, such analysis would

have to be regarded as highly speculative.

Perhaps the budget increments that were chosen were too large. It may be unlikely that a school district would increase its per-pupil budget by $100 in a given year. Given a more moderate increase, where chief variables do not attain maximum values, could generate a new population of values. Those districts that were previously near the observed maximum would use the budget increment to increase values of x_{16} beyond the previous maximum value. With a new population of values, another production function analysis would establish parameters based on the new population. Essentially, then the process of estimating a production function can be conceived as iterative in nature. Each new year provides additional information by which the analysis can be revised and extended.

Discussion

Given that the primary focus of the study was to develop a mathematical model to aid decision-makers in educational planning, the most relevant concern is whether or not the model would actually work under real conditions. If the results suggested in the analysis for the 181 school districts involved were actually implemented, would changes in achievement be within a reasonable range of that which was predicted? Of course, the answer to that question is presently unanswerable. The only way the adequacy of the model can be known is to actually implement the policies suggested by the analysis. However, a one-shot case approach could not constitute grounds for adequate verification. To prove the adequacy of the model would require that it be tested for a sufficiently large number of cases and then consider the ratio of

successes to failures, such ratio being the ultimate criterion of effectiveness. Thus, if model forecasts were shown to be reasonably correct 95 percent of the time, then it could be stated that the model has been able to describe and predict reality very well, and could be expected to do so in the near future.

The primary limitation of the present study was in the research design. The major difficulty in a survey study is that causal significance can never be attached to the conclusions. Thus in the present study differences in achievement could be estimated among schools having different percentages of teachers with advanced training. What could not be known was whether actually increasing the percentage for given schools would produce the same differences. To establish such cause and effect relationships generally requires an experimental, longitudinal study design. To find out what happens to a system when you interfere with it, you have to interfere with it, not just passively observe it.

In the present study it was recognized that omission of relevant variables can bias the estimates of the regression coefficients. Given the nature of a survey study, it can not be known with certainty if all influences acting on the dependent variables through the independent variables have been brought into the analysis. Nevertheless, merely pointing to this weakness does not discredit the study. One needs to show that a bias exists and that it matters. In the context of the conceptual framework it is important to recognize the nature and extent of controls that were used in the empirical application. For instance, an attempt was made to control for student background through the use of SES variables (percent of incomes above $10,000);

for community influences through the use of such variables as attendance and future training; and for student peer-group effects through percentage of minority enrollment and ESEA Title I pupils. No doubt, other influences were omitted from the model, but to what extent the coefficients were biased is not clear. It could be that given the present controls, remaining sources of bias may only slightly effect the coefficients.

Given this state of affairs, the chief contribution of the present study can be viewed as analytical. Analytical studies of this type are those that typically search for the most important relationships among the inputs of the educational system and its outputs, the principal purpose being to locate possibilities for improvement that appear worth exploring. The findings of such analyses are necessarily probabilistic. Even so, statements about "What would happen if . . .?" are likely to lead to far wiser decisions about educational policies than the kinds of uninformed hunches on which educational decision-makers often rely. Theoretically, there is every reason to suppose that the type of analysis displayed in the present study should be highly useful in helping educators obtain some idea of how best to deploy available funds, facilities, and personnel so as to maximize the educational outcomes students will attain, or to arrive at informed judgments about what trade-offs might be made among several kinds of inputs.

One of the shortcomings observed in the present analysis was found when the budget constraint was incremented. It was found that very quickly one reached the bounds of observed variation in relevant variables, analysis being inhibited beyond that point. At the same time, it is recognized that commitment to the modeling process in education does not mean a one-shot survey but a periodic reassessment.

Thus the production function should be estimated annually. To do so leads to two distinct advantages. First, with each new year to be considered, variation in the ranges of variables previously observed is expected to occur. Also, if results of previous analyses have been implemented, then wider ranges are to be expected in the important variables. If such a variable has previously been deployed to its maximum value, then a new population presents a new range of variation and analysis proceeds.

More importantly, with periodic assessment, one can proceed in a manner analogous to Bayesian statistical procedures. This means that prior information based on previous years' analyses can be used to improve and perfect current estimation of the regression coefficients of the various input factors. Given that cyclical variations may occur from year to year, the process is deemed important for determining the average effects of variables over a period of years.

It is hardly to be questioned that studies confined to the practical applications of models should be energetically pursued to provide, eventually, a firmer basis than now exists for dealing with broad questions of educational policy. Mathematical models, such as was developed in this study, allow decision-makers to form expectations of future consequences, these expectations being based on known empirical relationships and the decision-makers' judgments.

All of the foregoing does not mean that in applied work the researcher even pretends to be able to find the best of all possible decisions. The data are too incomplete and sometimes inaccurate, the tools of analysis are often too blunt, and the researchers knowledge of

the educational process is too limited for him to be able to come up with anything more than approximations to the ideal of the true optimum. Nevertheless, an analysis which is specifically designed to look for optimal decisions, crude and approximative though it may be, is very likely to do much better than the workable but relatively arbitrary rules of thumb which play so prominent a part in educational practice.

FOOTNOTES

1. See Scott N. Rose, A Study to Identify Variables to Predict Local School District Productivity in Two States (Ed.D. dissertation, University of Florida, 1972).

2. Data for variables x_1 - x_3 were taken from Personal Income by School Districts in the United States, Dewey H. Stollar and Gerald Boardman (Gainesville, Florida: National Educational Finance Project, 1971).

3. Biomedical Computer Programs, ed. by W. J. Dixon (Berkely: University of California Press, 1971).

4. The SHARE Program Library consists of a central file and documentation system available at most IBM installations.

5. For a discussion of the algorithm, see Fiacco, Anthony V. and Garth P. McCormick, "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming", Management Science, X (July, 1964).

6. See Herbert J. Kiesling, The Relationship of School Inputs to Public School Performance in New York State (Santa Monica, California: The Rand Corporation, October, 1969), P-4211, p.24.

7. Edgar L. Morphet et. al., Educational Organization and Administration (Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1967), p.274

TABLE 1

BASIC DATA DESCRIPTION

No.	Variable Title	Mean	Standard Deviation	Maximum	Minimum
1	Income Per-Pupil	11,456.35	8,217.34	65,797.00	893.00
2	Incomes under $3,000	29.27	5.03	43.20	20.40
3	Incomes over $10,000	22.53	12.74	56.70	3.60
4	ESEA Title I Pupils	4.15	5.20	38.50	0.0
5	Minority Enrollment	7.85	15.56	95.60	0.0
6	Attendance	93.90	1.55	96.60	87.60
7	Future Training	64.53	15.38	98.00	17.60
8	Size of School District	10,084.36	79,023.75		
9	Percentage Enrolled	86.47	10.41	99.00	47.00
10	Transportation Cost	65.48	26.50	144.22	7.70
11	Local Fiscal Effort	5.59	3.94	32.10	0.30
12	Expenses for Instruction	64.42	3.45	79.40	55.20

13	Longevity Experience	17.57	8.92	52.60	1.70
14	Teacher Preparation	5.87	5.46	32.10	0.0
15	Teacher Experience	37.38	9.57	60.60	14.70
16	Advanced Preparation	52.49	12.05	84.30	22.20
17	Median Teacher Salary	9,149.64	1,157.65	13,750.00	6,950.00
18	Average Class Size	18.50	1.66	22.20	12.30
19	Pupil-Support Personnel Ratio	136.55	82.20	898.30	44.90
20	Expenses for Transportation	5.64	2.33	11.70	0.60
21	Median Reading Achievement	43.07	6.86	55.48	28.73
22	Average Daily Attendance	9,180.50	70,852.56		316.00
23	Total Current Expenditure	11,490,572.00			
24	Instructional Expenditures Per-Pupil	764.51	139.32	1,287.90	554.13

TABLE 2

REGRESSION COEFFICIENTS FOR THE PRODUCTION FUNCTION
(DEPENDENT VARIABLE: READING ACHIEVEMENT--X_{21})

No.	Variable Title	Coefficient[a]	F-to-Enter[b]	F-to-Remove[c]	
Linear Terms					
3	Incomes over $10,000	0.150	33.95	13.26	(2)[d]
6	Attendance	-52.807	8.91	14.35	(2)
7	Future Training	0.093	10.35	14.35	(2)
12	Expenses for Instruction	4.292	14.26	6.87	(3)
13	Longevity Experience	-0.042	3.13	1.04	(3)
14	Teacher Preparation	-1.511	5.09	11.12	(3)
15	Teacher Experience	-0.001	2.80	0.00+	(3)
16	Advanced Preparation	0.120	40.23	10.53	(3)
17	Median Teacher Salary	0.002[e]	3.19	0.00+	(3)
18	Average Class Size	0.220	3.39	1.75	(3)
19	Pupil-Support Ratio	0.011	3.41	9.63	(3)
22	Average Daily Attendance	-0.004[e]	5.51	0.71	(3)
Quadratic Terms					
6	X_6^2	0.291	133.39	15.16	(2)
12	X_{12}^2	-0.035	7.51	7.51	(2)
Interaction Terms					
	X_4X_{19}[f]	-0.002	19.34	10.55	(2)
	X_5X_{18}[g]	-0.005	9.42	16.35	(2)
	$X_{14}X_{17}$	0.016	11.88	8.80	(2)
Constant Term					
		2282.167			

$R^2 = 0.79$

[a] The regression coefficient is the b weight (unstandardized partial regression coefficient)
[b] F test of significance of a single variable in a stepwise regression at the step of entry into the equation
[c] F test of significance of a single variable in a stepwise regression after the final step of the regression
[d] 2 = free variable; 3 = forced variable
[e] Variables X_{17}, X_{22} were scaled so that truncation errors would not occur. X_{17}: 1 unit = $100; X_{22}: 1 unit = 1000 pupils
[f] X_4: ESEA Title I Pupils
[g] X_5: Minority Enrollment

TABLE 3

AN EXAMINATION OF THE INTERACTION BETWEEN THE PERCENTAGE OF TEACHERS WITH LESS THAN FOUR YEARS TRAINING (X_{14}) AND MEDIAN TEACHING SALARY (X_{17})

Interval No.[a]	X_{17}[b]	X_{14}[c]	No. of Cases
1	[69.5, 80)[d]	[2 , 32.1]	22
2	[80 , 90)	[0 , 25]	73
3	[90 , 100)	[0 , 23.5]	48
4	[100 , 110)	[0 , 12.5]	25
5	[110 , 120)	[0 , 3]	10
6	[120 , 130)	[0.9, 1]	2

[a]Given an interval for X_{17}, such as [69.5, 80), the interval of corresponding values for X_{14} was determined. This means that as X_{17} assumed values in the interval [69.5, 80), the range of X_{14} was restricted to the interval [2, 32.1].

[b]X_{17} was scaled. 1 unit = $100

[c]X_{14} represented a percentage

[d][,) means that the interval includes the number on the left, but not on the right.

TABLE 4

REGRESSION COEFFICIENTS FOR THE BUDGET EQUATION
(DEPENDENT VARIABLE: INSTRUCTIONAL
EXPENDITURES PER PUPIL--X_{24})

No.	Variable Title	Coefficient[a]	F-to-Enter[b]	F-to-Remove[c]	
Linear Terms					
10	Transportation Cost	7.352	13.86	277.51	(2)[d]
12	Expenses for Instruction	7.730	41.30	51.33	(3)
13	Longevity Experience	-0.364	2.19	0.72	(3)
14	Teacher Preparation	-0.674	27.70	0.78	(3)
15	Teacher Experience	0.409	11.06	0.72	(3)
16	Advanced Preparation	-8.817	6.22	17.51	(3)
17	Median Teacher Salary	-14.256[e]	197.31	15.30	(3)
18	Average Class Size	-6.662	99.21	8.54	(3)
19	Pupil-Support Ratio	-0.032	12.97	0.78	(3)
20	Expenses for Transportation	-85.886	136.10	235.80	(2)
22	Average Daily Attendance	-0.358[e]	12.77	36.82	(3)
Quadratic Terms					
5	X_5^2 [f]	0.021	13.26	51.88	(2)
16	X_{16}^2	0.091	21.20	20.77	(2)
17	X_{17}^2	0.090	22.06	22.06	(2)
Constant Term					
		1125.488			
$R^2 = 0.94$					

[a] The regression coefficient is the b weight (unstandardized partial regression coefficient)

[b] F test of significance of a single variable in a stepwise regression at the step of entry into the equation

[c] F test of significance of a single variable in a stepwise regression after the final step of the regression

[d] 2 = free variable; 3 = forced variable

[e] Variables X_{17}, X_{22} were scaled so that truncation errors would not occur. X_{17}: 1 unit = \$100; X_{22}: 1 unit = 1000 pupils

[f] X_5: Minority Enrollment

TABLE 5

VARIABLES MANIPULATED IN THE MATHEMATICAL PROGRAMMING ANALYSIS

No.	Variable Title	Mean	Standard Deviation
13	Longevity Experience	17.57	8.92
14	Teacher Preparation[a]	5.87	5.46
15	Teacher Experience	37.38	9.57
16	Advanced Preparation	52.49	12.05
17	Median Teacher Salary	9,149.64	1,157.65
18	Average Class Size	18.50	1.66
19	Pupil-Support Ratio	136.55	82.20
21	Median Reading Achievement[b]	43.07	6.86
22	Average Daily Attendance[a]	9,180.50	70,852.56
24	Instructional Expenditures Per-Pupil[c]	764.51	139.32

[a]For each programming problem, X_{14} and X_{22} assumed constant values

[b]X_{21} was the variable maximized

[c]X_{24} was the budget constraint

TABLE 6

EFFECT OF DISTRICT SIZE ON ACHIEVEMENT AND THE DISTRIBUTION OF SCHOOL RESOURCES (X_{14} = 5.87)

x_{22}	x_{21}[b]	x_{13}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
1,000	45.7	1.7	14.7	69.6	9,630	22.2	136.5
5,000	45.7	1.7	14.7	69.8	9,630	22.2	136.5
9,180[a]	45.7	1.7	14.7	70.1	9,670	22.2	136.5
20,000	45.8	1.7	14.7	70.6	9,710	22.2	136.5
30,000	45.9	1.7	14.7	71.2	9,760	22.2	136.5
50,000	46.0	1.7	14.7	72.2	9,840	22.2	136.5
70,000	46.1	1.7	14.7	73.2	9,920	22.2	136.5
100,000	46.3	1.7	14.7	74.6	10,030	22.2	136.5

[a] 9,180 was the mean value observed in the population

[b] 43.1 was the mean value observed in the population

TABLE 7

THE EFFECT OF INCREMENTING THE BUDGET ON ACHIEVEMENT (X_{22} = 9,180)

Increments in Per-Pupil Budget	X_{21}	X_{13}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}
X_{14} = 0.0							
$0	46.1	1.7	14.7	75.4	7,970	22.2	136.5
		LB[a]	LB	OPT	OPT	UB	FIXED[b]
$100	47.2	1.7	14.7	84.3	10,250	22.2	136.5
		LB	LB	UB	OPT	UB	FIXED
$200	47.2	1.7	14.7	84.3	11,990	22.2	136.5
		LB	LB	UB	OPT	UB	FIXED
$-50	44.4	1.7	14.6	61.7	7,940	22.2	136.5
		LB	LB	OPT	OPT	UB	FIXED
X_{14} = 5.87							
$0	45.7	1.7	14.7	70.1	9,670	22.2	136.5
		LB	LB	OPT	OPT	UB	FIXED
$100	48.1	1.7	14.7	82.1	10,640	22.2	136.5
		LB	LB	OPT	OPT	UB	FIXED
$200	49.7	1.7	14.7	84.3	12,040	22.2	136.5
		LB	LB	UB	OPT	UB	FIXED
$-50	43.8	1.7	14.7	60.0	8,850	22.2	136.5
		LB	LB	OPT	OPT	UB	FIXED

[a] LB - Lower bound of the variable
UB - Upper bound of the variable
OPT - Optimal value of the variable (LB < OPT < UB)

[b] X_{19} was manipulated, but the upper bound was fixed at 136.5

SELECTED BIBLIOGRAPHY

Blalock, Hubert M. Causal Inferences in Non-Experimental Research. Chapel Hill: University of North Carolina Press, 1961.

Bohrnstedt, George W. "Observations on the Measurement of Change," in Sociological Methodology. Edited by Edgar F. Borgatta. San Francisco: Jossey-Bass, Inc., 1969.

Cohen, Jacob. "Multiple Regression as a General Data-Analytic System," Psychological Bulletin, LXX (1968).

Cohn, Elchanan. "Economies of Scale in Iowa High School Operation," Journal of Human Resources, III (Fall, 1968).

Levin, Henry M. "A Cost-Effectiveness Analysis of Teacher Selection," Journal of Human Resources, V (Winter, 1970).

________. "The Effect of Different Levels of Expenditure on Educational Output," in Economic Factors Affecting the Financing of Education. Edited by Roe L. Johns et al., Gainesville, Florida: National Educational Finance Project, Vol. II, 1970.

McNamara, James F. "Mathematical Programming Models in Educational Planning," Review of Educational Research, XLI (1971).

Riew, John. "Economies of Scale in High School Operation," Review of Economics and Statistics, XLVIII (1966).

Thomas, J. Alan. The Productive School: A Systems Analysis Approach to Educational Administration. New York: Wiley & Sons, Inc., 1971.

Wegner, P. "Relationship Between Multivariate Statistics and Mathematical Programming," Applied Statistics, XII (November, 1963).