Educational Considerations

Volume 35 | Number 2 Article 2

4-1-2008

Optimizing Educational Resources: A Paradigm for the Pursuit of Educational Productivity

James L. Phelps

Recommended Citation

Phelps, James L. (2008) "Optimizing Educational Resources: A Paradigm for the Pursuit of Educational Productivity," *Educational Considerations*: Vol. 35: No. 2. https://doi.org/10.4148/0146-9282.1186

Optimizing Educational Resources: A Paradigm for the Pursuit of Educational Productivity

James L. Phelps

> The advantage and perhaps the major motivation for using "seat-of-the-pants" decision making is that it obscures the assumptions made in arriving at a decision. If no one knows the assumptions upon which you based your decisions, then even though they may be uneasy with the decision they will have a difficult time criticizing your assumptions or decisions. (Schrage, p. 305)[1]

The never-ending organizational challenge is to allocate available resources to best achieve its goals. Out of this fundamental question several models have evolved. One is a conceptual model—a way to think about how organizations operate. A second is a statistical model estimating the magnitude of relationships among goals and elements of the organization. This article presents a third model, an optimization model building upon the other two in order to analyze various policy options by simulating "what if" situations arising in organizations. These three models are complementary rather than competing.

Optimization Modeling

What Is a Model?

Over time, scientific endeavors have increasingly relied on models combining fact (observations), theory (assumptions), laws (usually mathematical), and methodology (procedures) into a system describing phenomena behavior. Models evolve as anomalies, are identified in older models, and are replaced with different facts, theories, laws, and methodologies describing the behavior of the phenomenon in question more comprehensively and with greater precision. Only by discarding previous beliefs and replacing them with a different set is the newer model accepted.[2]

There are mathematical models designed to represent the elements within the structure of an organization and to describe their relationships with the organization's goals. These mathematical models use equations representing the presumed "reality" to solve "what if" questions by changing the model parameters.[3] In this case, the organization under consideration is a school.

Why Build a Model?

According to Williams, the value in model building is threefold.[4] First, building a model often reveals structures, elements, and relationships usually taken for granted until the underlying assumptions are stated and tested. Once the original ideas are stated and tested, they usually give way to more sophisticated and accurate representations of the actual situation. Second, once the model is constructed, analyzing it mathematically suggests courses of actions not readily apparent. In essence, the model challenges conventional thinking. Third, experimentation is possible within a model that is not practical in actual situations. Through experimentation more potentially successful options may be identified. Unlike "seat-of-the pants" decisions, models can be tested.

Fundamental Assumptions

To start, there are five fundamental assumptions regarding desirable school outcomes: (1) Student outcomes as measured by achievement tests are appropriate measures of school performance; (2) Other student outcomes, such as school retention, graduation, and employment rates are also appropriate measures of school performance; (3) Because many of the measures of student performance are highly associated with the school's community socioeconomic status (SES), it must be taken into consideration; (4) Because all schools will not have the same success in achieving student outcomes due to differences in organizational effectiveness, school effectiveness should also be taken into consideration; and (5) When considering alternative policies to achieve the desired outcomes, cost-effectiveness is a critical component.

Next are five fundamental assumptions regarding modeling school organizations: (1) Based on the properties of the normal curve, achievement tests are stochastic in nature, and the model must be consistent with these stochastic properties;[5] (2) Because achievement tests have a definite upper limit rarely, if ever, achieved by all students within a school, "perfection" is not obtainable, and therefore there is a point after which additional resources will produce diminishing returns; (3) Schools pursue multiple outcomes simultaneously; (4) Schools are complex organizations balancing multiple elements and processes to achieve their multiple goals; and (5) Because there will be a unique solution for each modeled school based on the initial conditions of the organization, there will not be a single policy to achieve the desired results applicable to all schools.

Conditions to Achieve Optimization

Mathematical programming (sometimes called "linear programming") is merely a method of solving simultaneous equations. The solution could represent the optimal use of resources to produce the optimal level of outcome. The basic structure of a mathematical programming problem is illustrated by this example:

$$
\begin{array}{ll}
\text{Maximize:} & 3X + 2Y \\
\text{Subject to:} & X + Y < 4 \\
 & 2X + Y < 5 \\
 & -X + 4Y > 2 \\
\text{Constraints:} & X \geq 0 \\
 & Y \geq 0
\end{array}
$$

Establishing equations accurately representing the organization to be modeled is the key to mathematical programming. These equations must meet certain conditions in order to be solved. The four basic conditions listed below are developed throughout the paper:

(1) There must be a single expression, the "objective function" to be maximized, minimized, or set to a specific value representing the underlying purpose of the model.

(2) There must be simultaneous equations accurately representing the structure and elements of the organization and their relationships

James L. Phelps is a former Education Assistant to the Governor of Michigan and Deputy Superintendent in the Michigan Department of Education.

Published by New Prairie Press, 2017

to the organization's desired outcomes for which there are solutions or boundaries.

(3) The boundaries may be of various types:

- Intersection of lines (lines with positive and negative slopes)
- Maximum or minimum points of nonlinear functions (curves with a change in the sign of the slope)
- Diminishing returns (curves with a changing slope approaching asymptotic).

(4) There are usually constraints or a series of expressions setting limits on any or all of the variables. Cost is a frequent constraint.

Why Is Education Different?

Much of the mathematical modeling has been developed in areas such as business where the outcomes are in discrete and limitless increments, and the relationships are frequently linear. For example, if the purpose of the organization is to produce and sell widgets, it is straightforward to calculate how many machines and how much material is needed and what staffing levels are required to operate and maintain the equipment. The associated cost with these elements can also be determined. With this information, different combinations can be explored to determine the best—the most economical—way to proceed. There is no limit as to the number of widgets that can be produced although there may be a limit to the number that can be sold.

In contrast, there are areas, such as education, where outcomes are stochastic—measured by normally distributed achievement tests—and the relationships among organizational variables and outcomes are less straightforward. The results from a change in the organization's activity can only be estimated based on probabilities and within a margin of error rather than with great certainty. Also, there are definite limits. If the average score on a standardized achievement test were 100, there is no way to modify the school organization at any cost to double the score, to 200, if a perfect score was 150. Indeed, while it is possible to make a plethora of widgets virtually identical, it is virtually impossible to make the achievement of a plethora of students identical.

Given the difference between nonstochastic manufacturing products and stochastic education outcomes, the model presented here is designed to address the fundamental question raised previously: How can schools allocate available resources to best achieve student performance goals?

The Production Function and Regression Analysis

Conceptual Elements of Production

A helpful model for thinking about organizations is the production function.[6] Conceptually, the production function is divided into three main parts: (1) the outcome to be achieved; (2) the input required; and (3) the process used to convert the input into the outcome. It is represented by the following equation:

$$\text{Outcome} = \text{Input} + \text{Process}$$

In most cases, each of the parts is comprised of many variables.

As the equation requires, the level of outcome increases if either the input or process variables increase, but the "trick" is to determine which input or process variables to increase and by how much. In modeling, if the levels of inputs and process variables and their relationships to the outcome are known, the level of outcome can be predicted. This knowledge provides insights on how a change the input and process levels will alter the level of the outcome. When deciding the variables to include and the mathematics to estimate the relationships and to calculate the predicted outcome, the basic operational assumptions of the organization, either implicit or explicit, are incorporated into the model.

The production function may be optimized via mathematical programming when the input and process variables and their relationships to the outcomes are known. When the relationships are unknown, they are usually estimated though the statistical model of regression. However, regression analysis does not directly provide answers to optimization questions.

Estimating Relationships Via Regression

The basic regression model estimating the relationships (weightings) is straightforward:

$$\text{Outcome} = X_1{*}I_1 + X_2{*}I_2 +..X_n{*}I_n + Y_1{*}P_1 + Y_2{*}P_2 +..Y_n{*}P_n + \text{Unknown} + \text{Error}$$

The X's and Y's represent the estimated weightings measuring the relationship between the outcome and the input and process variables. The I's represent the variables defining the inputs. P's represent the variables defining the processes. "Unknown" represents the important variables in the production function for which data are unavailable. "Error" represents the portion of the equation that cannot be explained because of measurement error.

In order to get meaningful results, the distributions of the outcome, input, and process should be normal or near normal with a substantial degree of variation. Variation is required to accurately place each observation. In education, student achievements test are designed based on these characteristics and, therefore, are stochastic. (See footnote 5.)

Interpreting Regression Results

The most common conclusion of a regression analysis is the statistical significance of the weighting; if it is significant, then it is thought appropriate to increase the level of the input or process variable. However, the level of significance does not help determine how much to increase the variable.

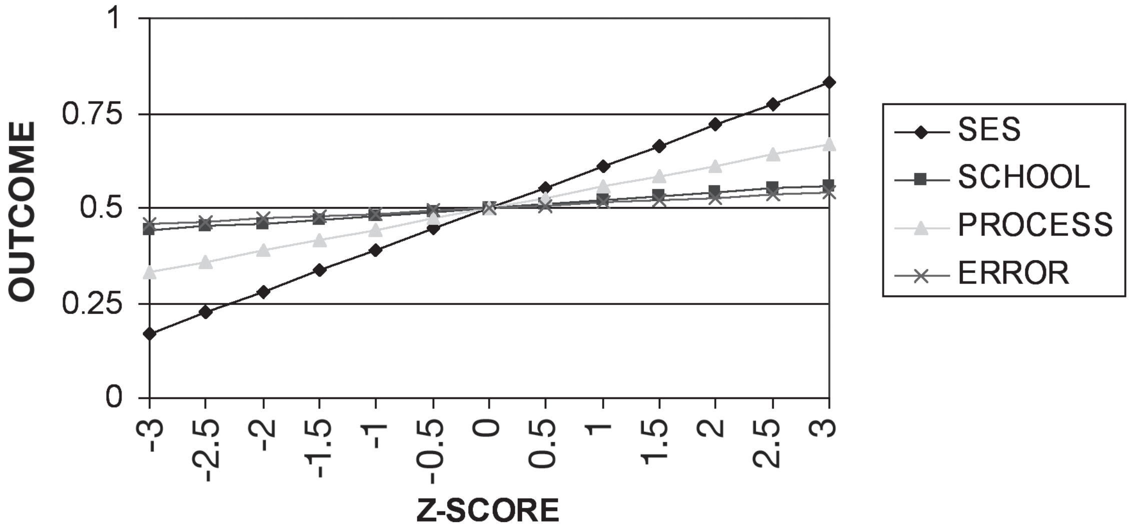
The weighting measuring the relationship between the outcome and the independent variable(s) is interpreted as slope; that is, the unit-change in the level of the outcome for each unit-change in the input or process variable. Slope is also the mechanism for predicting the most likely value of an outcome from the known value of an input or process variable. The slope does provide some greater help in determining which variable to increase because it only makes sense to increase the variable(s) with the highest slope—"the biggest bang."

Many of the following illustrations have been taken from a previous study by the author where the production function was divided into the community input of socioeconomic status (SES) and the school inputs of staffing quantity, staffing quality, and other financial resources.[7] There were no direct data representing the process, which is usually the case. The process component was defined as the effectiveness of the school organization to produce scores higher than what was predicted from knowing the other inputs—the residual. The slopes of the categories of the study are depicted in Figure 1.

Because each of the variables has a unique descriptive statistic, it is difficult to compare their influence on achievement without first converting all outcomes and variables to standard scores (Z-scores). The slope is then the standard regression coefficient. This is the

http://newprairiepress.org/edconsiderations/vol35/iss2/2
DOI: 10.4148/0146-9282.1186

Figure 1
Regression Slope



convention in the remainder of this paper. Most frequently the graphic representations of the outcome and variables is based on "Cartesian" geometry with the navigation point being the origin (X and Y = 0) with the outcome(s) on the Y-axis and the variable(s) on the X-axis. Because the mean value of an independent variable predicts the mean value of the outcome (dependent variable), charting mean against mean as the navigation point will be used. (A standard score or Z-score of zero is the mean.) The outcome in this illustration is measured in percentiles for reasons to be given later.

With this interpretation of slope, there comes a predicament: Why increase any but the variable with the highest slope if the other variables will make less of a difference in increasing the level of the outcome? This contradicts one of the basic assumptions of the production function: It takes a combination of variables combined in a balanced way to improve outcomes rather than just one or two variables in high concentration. This predicament will be addressed later.

There is another aspect to the regression analysis--predicting the outcome level based on the values of the input and process variables. By substituting the actual values back into the regression equation with the estimated weightings, a predicted level results. The difference between the actual outcome level and the predicted outcome level is the residual, or, an unfortunate name, "error."

Residual as Effectiveness

The notion of the residual being all error is misleading. An important variable may not have been included in the original equation, and, if it were, the error term would be reduced. Therefore, part of the error term is usually due to a misspecification of the equation, but what if the residuals were compared over several periods of time and there was a tendency for the residuals of each observation to have the same sign and magnitude? In this case, it would be fair to assume the pattern of the residual actually measures something real but unobserved. Because organizations utilize their resources to different degrees of effectiveness, a logical conclusion would be for any consistent pattern of the residuals over time to be associated with an unobserved effectiveness factor.[8]

Limitations of Regression to Optimize

While of great value in estimating the magnitude of relationships, the statistical model of regression does not directly address the fundamental question of how to best allocate resources among the input and process variables.

The basic assumption of the regression model is that of linearity of the weightings; as each unit of the independent variable is increased, there will be a constant increase in the outcome. To have a "perfect" outcome, e.g., all students with a perfect score, it is mathematically possible by increasing any one of the model inputs sufficiently to obtain a predicted perfect score. In practice, this situation does not occur. Indeed, some students achieve perfect scores within the existing resources, but there is a distribution of scores for all the students with the average score well below perfect. In order to achieve a perfect score for an individual school, the variation among students would have to be reduced to zero as well as an improvement of all scores below perfect. Perhaps this could be achieved by eliminating some students from the population or "dumbing-down" the test, but these efforts would negate the basic purpose of assessing student progress. At the heart of the stochastic assumption is the recognition of the existence of individual differences over which the school has only partial control.

While it is possible to introduce some degree of nonlinearity into variables, e.g., introducing an additional term calculated by squaring the variable value, these results are seldom significant. Even if significant, there is seldom a change in the sign of the slope—a maxima or minima point—and thus, predicted "perfection" is still possible.[9]

Thus, if all the variables are linear (or at least always with a positive slope), what is the optimum allocation of resources? Initially this question may be addressed by standardizing the weightings, converting all variables to standard scores so they are comparable. After the weightings are standardized there is the question of cost. This can be addressed by comparing the standardized weightings per dollar.

After these procedures are completed, there is still no answer to the fundamental question. Because only one variable will have the best cost per unit improvement of the outcome defined as cost-effectiveness, mathematical logic still leads to placing all the resources in a single variable. While logical mathematically, it is not logical operationally. Organizations operate effectively because of blending many variables to achieve the best outcome, not by selecting just one "basket for all the eggs." In addition, most organizations have the mission of achieving multiple outcomes, but regression, with just a single equation, addresses only one. While various outcomes could be combined to form a single outcome, much of the valuable information unique to each outcome would be lost.

In summary, the regression statistical model as an optimization tool is deficient in four respects: (1) It does not directly model the relationships among multiple outcomes and the organizational inputs and processes; (2) It assumes linearity in the weightings, precluding a systematic balancing of the various variables to achieve the best possible outcomes; (3) With linearity, outcome "perfection" can be achieved given sufficient resources and investment in only one variable; and (4) There is no provision within the model for addressing cost-effectiveness.

Using Regression to Seed an Optimization Model

Based on everyday experience, the assumptions represented by the statistical model of regression are not consistent with school organizational reality. One would be hard-pressed to identify a school organization operating under the assumptions of the regression model, but is it possible to take the analytical results from regression and insert them into a mathematical programming model more consistent with reality?

Estimates from Regression Into Mathematical Programming

Regression, with a single outcome, is not designed to optimize. This can be easily addressed by formulating individual equations for each of the outcomes, establishing a set of simultaneous equations, an essential characteristic of mathematical programming. The explicit goal is to achieve the highest possible level for the sum of the multiple outcomes. (A mathematical transformation can be made to accommodate something like a dropout rate where it is desirable to have the rate low.) If some outcomes were thought to be more important than others, a weighting system among the outcomes could be included. Addressing the second and third deficiencies mentioned above is more involved.

Transforming Relationships to Achieve Diminishing Returns

Conceptually, there are three general ways to describe the relationship between inputs and outcomes, sometimes called "returns to scale": (1) Increasing returns to scale or the inverse, decreasing returns to scale; (2) Maxima or the inverse minima; and (3) Constant returns to scale. (See Figure 2.) Note that one curve is increasing for the first half and decreasing the second. The slope determines the type of relationship based on whether the slope is increasing or decreasing, whether there is a point where the slope is zero, or whether the slope is constant. The return is measured in percentiles.

In order to solve simultaneous equations, as mentioned previously, there must be either intersection of lines; maxima or minima points of curves; or curves representing diminishing returns. Assuming positive linearity of each regression weighting, there can be no intersection of lines or maxima and minima points, therefore no solution to the equations. The most likely alternative to solving the simultaneous equations is to form nonlinear functions indicative of diminishing returns.

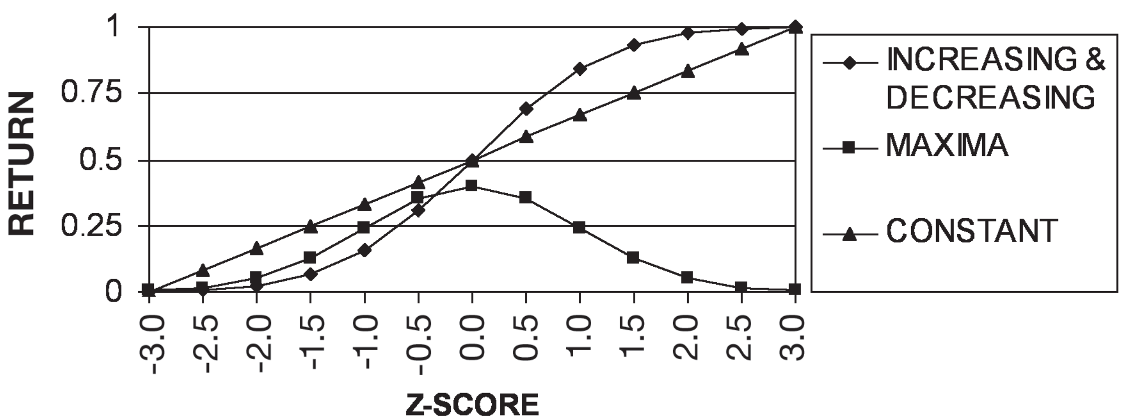
Diminishing Returns Function Within Regression Analysis

At this point, there is an essential digression to demonstrate mathematically the existence of a nonlinear function indicative of diminishing returns based on regression analysis.

Students in beginning statistics courses are taught several descriptive statistics, but they most likely do not fully appreciate their full beauty and power. Usually, an early step is to construct a histogram underlying the distribution of a bell-shaped curve. Students are then asked to calculate the mean and standard deviation. After calculating the mean, the deviations from the mean are calculated, these deviations are squared, and then they are summed. The result is called the sum of the squares and commonly noted as "SS." The sum of the squares is then divided by the number of observations (N) to produce the mean of the squares (MS). This is also called the variance as symbolized by $\sigma 2$. When the square root of the variance is taken, the result is called the standard deviation or σ. The variance is some notion of area, but area of what? The standard deviation is some notion of length, but length of what?

The primary purpose of regression analysis is to make predictions regarding the level of the dependent variable (outcome) based on the values of the independent variables (inputs). The basic idea is to plot

Figure 2
Returns to Scale



DOI: 10.4148/0146-9282.1186

the dependent variable on the Y-axis and the independent variable on the X-axis to determine if these points tend to fall on a line. While this can be inspected visually, it can be measured with great accuracy mathematically. The line is considered the "best fit" when the distance from the observation point to the regression line is squared, summed for all observations, and minimized. This method is called the "least-squared" solution. The line is represented algebraically as the slope of a line. It is presented in two forms, one using the original values, i.e., the regression coefficient, and another using standard scores—the standard regression coefficient. When the variables are measured in standard scores (Z-scores) and the slope is measured in terms of the standard regression coefficient (r), the value of the outcome can be predicted from the value of the independent variable with the equation: $Z_{(y)} = r\ Z_{(x)}$.

However, the regression analysis provides another estimate, the amount of variance explained by each of the variables. Regression programs calculate the sum of the squared deviations for the independent variable(s) and well as for the residual, what is not accounted for by the independent variable(s). These sums of the squared deviations are converted to percentages of the total and called the coefficient of multiple determination, or R^2. It is a measure of the "goodness" or "strength" of the prediction of the variable(s), with the higher value indicating a greater strength. When the R^2 is 100%, there is "total strength," and when the R^2 is 0%, there is "no strength." When the percentage of what can be explained or attributed is added to the percentage of what cannot be explained or attributed, the sum is 100%.[10] Can the R^2 be related to the probability curve?

What Is the Probability Curve?

The idea of the probability curve is rather straightforward. If one tossed a number of coins a number of times and calculated the number of times each combination of heads and tails occurred, the result would form a histogram high in the middle and low at the edges. (The probability of each combination can be calculated via a binomial expansion and represented by the coefficients depicted in Pascal's triangle.) The probability curve is merely the probability histogram as the number of observations approaches infinity and converted to a continuous bell-shaped curve. It answers questions regarding the probability that any event will occur. Of course, there are limits or boundaries to probability. No event can occur more than 100% or less than 0% of the time.

The continuous bell-shaped curve is represented by the expression, $e^{-Z^2/2\sigma^2}$. The denominator of the exponent contains the variance (σ^2) from the descriptive statistics. The area under the probability curve, when normalized, is by definition 1 because the chances of something happening cannot be greater than 100 percent; so there must also be a denominator added to the expression representing the area of the curve. When the denominator equals the area of the numerator, the result is 1. The area of the probability curve is $\sigma\sqrt{2\pi}$, so the complete expression for the normalized curve is $(1/\sigma\sqrt{2\pi})\ (e^{-Z^2/2\sigma^2})$. The standard deviation (σ) appears in the calculation of the area. The variance and standard deviation are parameters of the probability curve.

Reformulating the Regression Results Into the Normal Curve

From regression, the explained variance by the independent variable plus the unexplained variance equals 1, as represented by the following equation:

$$R^2 + K^2 = 1$$

R^2 is the explained variance, and K^2 is the unexplained variance. If additional variables are added to the equation, the proportional relationship is maintained as represented in the equation:

$$R^2_1 + R^2_2 + K^2 = 1$$

Therefore, each term in the equation explains a proportion or percentage of the total variance. Variance is a measure of area based on the principle of squared deviations.

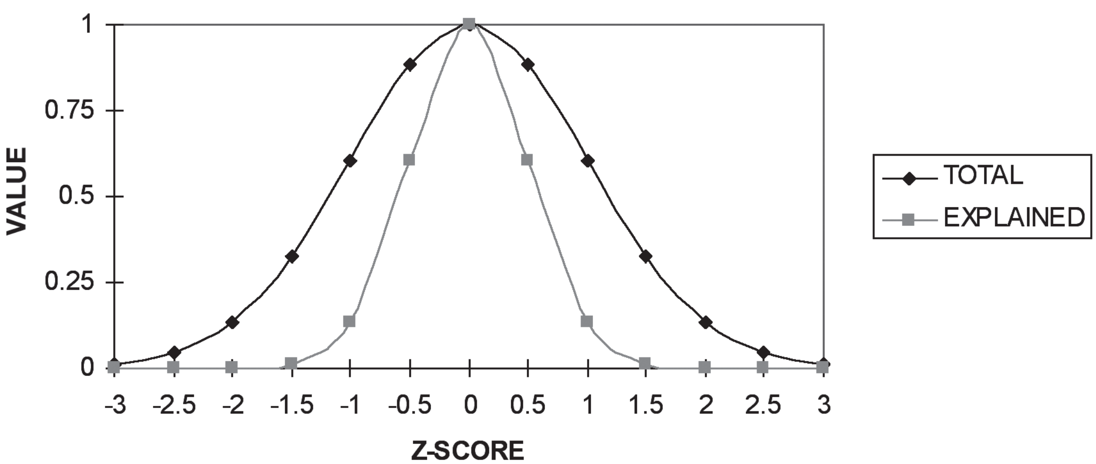
For the ease of notation, I will call the area of the probability function $f(z)$, where the measurement of the X-axis is in terms of Z-scores, or standard scores, and the area of the probability curve is normalized ($f(z) = 1$), and is represented by the following equations:

$$(R^2_1 + R^2_2 + K^2)\ f(z) = 1$$

or

$$R^2_1\ f(z) + R^2_2\ f(z) + K^2\ f(z) = 1$$

Figure 3
Comparison of Variance



Published by New Prairie Press, 2017

Given a specific observation as measured by a Z-score, the relative position of that observation can be easily calculated and reported as the percentile ranking. Therefore, the predicted placement, measured as a percentile ranking $Y(p)$, for a specific observation across all terms is calculated by substituting the appropriate Z-score for each term, with K^2 representing the margin of error, as follows:

$$Y(p) = R^2_1\ f(z_1) + R^2_2\ f(z_2) + K^2\ f(z_3)$$

or

$$Y(p) = R^2_1\ f(z_1) + R^2_2\ f(z_2) +/- 1/2\ K^2\ f(z_3) \text{ Q.E.D.}$$

In other words, the reformulated equation is a regression equation measured in terms of the proportion of area under the normalized curve or percentile and the predicted outcome value can be calculated for any combination of Z-scores. This representation of the R^2 is easily demonstrated graphically for it now relates to the proportion of area under the normal curve. (See Figure 3.)

Interpretation of the Normal Curve

While there is a maximum point at a Z-score of zero (the mean), the slope then turns negative, signifying declining returns rather than the more plausible diminishing returns. There is no evidence or theory suggesting that benefits would or should start decreasing when resources move past the mean. Is there another way in which to view theses curves that is more consistent with evidence and theory?

To review, the area under probability curve ($\sigma\sqrt{2\pi}$) is determined by the width parameter σ (standard deviation). The probability curve is represented by the expression $e^{-Z^2/2\sigma^2}$. The Z symbol Z represents the standard score or Z-score, and when Z equals zero, the function equals one. (See Figure 3.) As one might expect, the calculations of area of this expression are messy, to say the least. Instead, a single ideal normalized curve is established: area = 1 when $\sigma = 1/\sqrt{2\pi}$. The calculations of area are made on the ideal curve and given either in a table in a statistics book or as a part of a computer program. Hence, the cumulative area under the normal curve can be calculated for any given Z-score.[11] The formal name of the resulting S-shaped curve is the standard normal cumulative distribution, or cumulative area curve for short. Given this metric, it is possible to determine easily the percent of observations above and below a given score—the percentile.

This cumulative area curve represents the concept of diminishing returns because the benefits gradually reduce as the variable increases but never reaches a maximum point. (See Figure 4, marked "Area.") This representation appears to match the evidence and theory of the correlates of student performance. One could argue that having more textbooks in the classroom would be positively related to student outcomes, but only up to a certain point. After each student has one textbook, what would be gained by having more? Even in the case of class-size, it would seem illogical to argue that more than one teacher per student at any one given time would lead to higher achievement than having just one. A case can be made in virtually all circumstances that there is a point where additional resources would reap little or no benefit. Optimization will help determine where these points lie.

Importantly, the cumulative area curve can be used for solving simultaneous equations. Even more importantly, the shape of the cumulative area curve is determined by the R^2 value from regression analysis. The probability and cumulative area curves are related through the mathematics of calculus. The cumulative area curve is the integral of the probability curve and the probability curve is the derivative of the cumulative area curve. This means the probability curve is the slope of the cumulative area curve at the same Z-score. At a Z-score of zero, the value of the probability curve is one, so the slope of the cumulative area curve is also one. When area curve is adjusted for the R^2 value, the slope of the curve at a Z-score of zero is the R^2 value. Through the application of mathematics, the estimates from regression analysis can be transformed into a function suitable for solving simultaneous equations.

By way of illustration, if there were a single independent variable in the equation and the R^2 was 1.00, there would be a perfect relationship between the independent variable and the outcome. The key is that the distribution of the independent variable is measured in terms of standard scores, or Z-scores while the outcome or dependent variable is measured in terms of the proportion of variance explained—the cumulative area under a probability curve, or percentiles. For every standardized-unit increase (Z-score) in the independent variable, there is a corresponding increase in the outcome. In graphic terms, the distribution of the independent variable moving from the lowest to the highest corresponds with the cumulative area under the curve of the outcome from lowest to highest. In other words, the distributions of the outcome and independent variable would be identical but measured in different terms, and, thus, the independent variable explains all the variance of the dependent variable. (See Figure 5.)

If the R^2 were zero (.00), there would be no relationship between the independent variable and the outcome. There would be no width to the outcome variable distribution and no width to the cumulative outcome distribution. In essence, every value of the independent variable would make the same predicted value for the outcome—the mean value. Instead of a spread of the cumulative distribution, there would be a single horizontal line at the mean (50th percentile). Thus,

Figure 4
Normal Curve

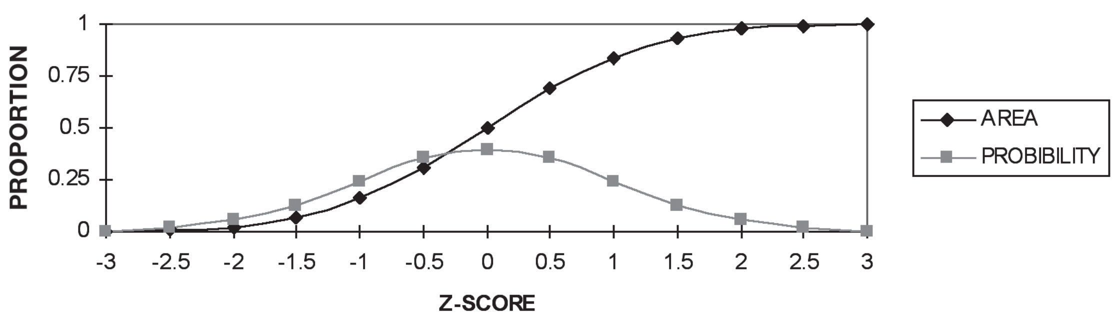
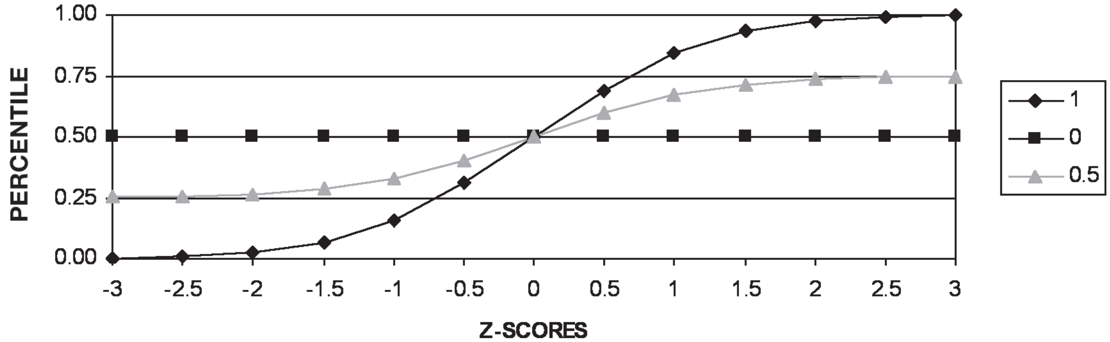


Figure 5
Area Under Curve



the independent variable would explain none of variation in the dependent variable, and the slope of the area curve would be zero. (See Figure 5.)

If the R^2 were .50, there would be a strong relationship between the independent variable and the outcome. The mean value of the variable would still predict the mean value of the outcome, but what about the other values? Because the area of the independent variable would be half of the outcome, half of that area (or one-quarter) would be above the mean and half would be below. When graphed, the S-shaped cumulative curve will be asymptotic to lines representing .75 and .25 of the area. These parameters conveniently represent percentiles. (See Figure 5.)

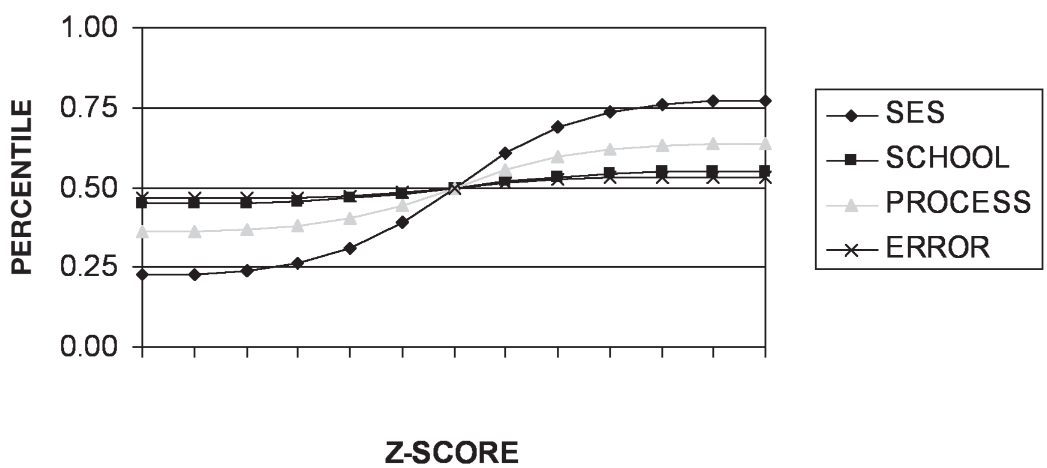
The R^2 terms can be calculated using the respective regression coefficient (r) and the standard regression coefficient (βr). In one sense, this calculation is more precise because it can be negative if r is negative, indicating an inverse relationship between the outcome and the independent variable. On the other hand, a negative R^2 term will not satisfy the summation to 1.0 and is changed to a positive (absolute value) for that purpose in statistical programs. This anomaly should be considered when determining the value of R^2 in a model. A negative coefficient makes the same contribution to the explanation of an outcome as does a positive value, so if there is an inverse relationship between the independent and dependent variables, the sign of the R^2 value should be set to negative in the simultaneous equations.[12]

In summary, the relationship between the distribution of a probability curve and the cumulative area curve is a straightforward transformation suitable for solving simultaneous equations. Conceptually, it is merely converting the outcomes to percentiles and the independent variables to standard scores.

The S-shaped curves are all asymptotic to the lowest and highest values as determined by the R^2, thus solving the boundary dilemma of achieving perfect scores by allocating an infinite amount of resources. While an increase of resources may improve the outcome level, it is both conceptually and mathematically impossible in this interpretation to achieve perfection because the asymptotic curve will never reach the maximum. This situation is consistent with the basic assumption of school performance. When applied to actual estimates of the production function, the respective relationships are depicted in Figure 6.[13]

With this transformation, the mechanics of optimization are rather straightforward even though the preparation of the data is somewhat tedious. The multiple R^2 weightings are inserted into a set of simultaneous equations based on the cumulative area function. Then, the principles of mathematical programming are applied to solve for the optimal levels of variables that will produce the highest level of summed outcomes. Importantly, the simultaneous equations model also requires the inclusion of constraints consistent with organization practice, the most notable being cost. Other upper and lower limits can

Figure 6
Relationships



be included as organizational practice requires. It should be emphasized that this solution is not for the weightings as they were estimated via regression analysis in the form of the R^2. Rather, the solution is for the values of the independent variables that will predict the best result—the highest predicted level of outcomes summed across the several equations.

The shift from the standard regression model to an optimization model may be more difficult psychologically than mathematically. Because of common use, most people are more comfortable with regression, but the critical difference is in the acceptance of the deficiencies listed above and their practical consequences. It is much easier to believe in continuous improvement for increased resources than it is to believe in diminishing returns—a point where an increase in resources produces little, if any, improvement. However, can the simultaneous equations with the transformations actually be solved and will the solution provide insights into the fundamental question—what is the best allocation of resources to achieve the optimal outcomes?

The Optimization Model

The optimization model takes a form common in mathematical programming, with the following elements: Objective function as the sum of the outcomes; equations defining the relationships between multiple variables and the outcomes; equations calculating the cost; and constraints limiting the upper and lower bounds of the variables

There is no method to predict future outcomes with complete accuracy. There are changes in the organization plus there is a certain degree of measurement uncertainty. As a result, the estimated outcomes are stochastic and based on predictions. Therefore, there must be two sets of simultaneous equations defining the outcomes, somewhat like a "before" and "after." Before and after are not different time periods; rather, they are the predicted results before and after the optimization. Before estimates the actual predicted target utilizing the existing variable values, and after estimates the optimized predicted target utilizing the optimized values.

The basic structure of the equations is similar in form to regression equations:

$\text{Outcome}_a = W_{a1}*V_1 + W_{a2}*V_2 +\ldots W_{an}*V_n + \text{Residual}$

$\text{Outcome}_b = W_{b1}*V_1 + W_{b2}*V_2 +\ldots W_{bn}*V_n + \text{Residual}$

$\text{Outcome}_n = W_{n1}*V_1 + W_{n2}*V_2 +\ldots W_{nn}*V_n + \text{Residual}$

W's are weightings, potentially different in each equation while V's are variables, the same in each equation. For each set of equations, the outcomes are summed to produce a target:

Actual Predicted Target (Before) = Set One (Outcome_a + Outcome_b + ... Outcome_n)

Optimized Predicted Target (After) = Set Two (Outcome_a + Outcome_b + ... Outcome_n)

The objective function, the value to be maximized, is the gain in the predicted outcomes achieved by changing the resource allocation pattern:

Objective Function = Optimized Predicted Target (After) – Actual Predicted Target (Before)

The constraints control the total cost as well as minima and maxima for each of the variables:

Total Cost = $V1*\$_1 + V_2*\$_2 + \ldots V_n*\$n$

Total Cost = specified value

$Vn \geq$ = specified value

$Vn \leq$ = specified value

The weightings (W's) are the R^2 for the respective variables (V's). The R^2 are estimated via regression analysis, but, as noted earlier, it can be negative. The respective variables (V's) are the actual observation values for calculating the actual predicted target in the first equation, and the optimized values for calculating the optimized predicted target in the second equation. The total cost of each set of equations is calculated by multiplying the value of the variable (V) by the average cost of the variable ($).

There must be a cost constraint; organizational resources are always limited. The values of $V_1*\$$, $V_2*\$\ldots$ $V_n*\$$ when added must be equal to or less than a specified amount, the total cost of the resources available to the organization. The purpose of optimization is to maximize the sum of optimized predicted outcomes while staying within the cost boundary. A "cost" expression is inherent to optimization but missing in regression.

What makes this model unique is the function used to represent the relationship between the independent variables and the outcomes. Because achievement outcomes are measured in stochastic terms—normal distributions—the relationships are measured in the same way. Rather than defining the relationship between the independent and dependent variables as the linear slope, the relationship is defined in terms of a type of standard scores. Because the area under the normal curve can be represented in terms of percentiles, the unique function is the integral of the normal curve—an S-shaped curve—adjusted by the degree of relationship, the R^2. The higher the R^2, the more vertically expanded the S-shaped curve, and vise versa. The integral of the normal curve is asymptotic at high and low points, so it is impossible to reach the absolute maximum or minimum points. While the slope at the mid-point (Z-score of zero or the 50th percentile) is the R^2 value, the slope gradually diminishes as it progresses upward and is symmetrical downward. (See Figure 4.) The basic idea is to increase the allocation level in favor of the variable when the slope is the greatest and decrease the allocation level in disfavor of the variable when the slope is the least. This decisionmaking rule is the essence of diminishing returns.

Data Requirements

Most state departments of education have data on the most frequently considered variables, such as the numbers of staff, salaries, qualifications, etc. The model can be specified for either school districts or school buildings. There is the obvious relationship between the sophistication of the data and the model; that is, the more sophisticated the data, the more sophisticated the model will be. With advancements in technology, the data for the model are easily obtainable through information systems.

The following data are required for the model: (1) Population data on the outcomes and variables to calculate means and standard deviations; (2) observation data for the outcome and variables, including actual levels; (3) cost data for the variables of the observation; and (4) estimates of the relationships between the outcomes and the various variables in terms of the R^2.

The model can be established based on two types of scenarios: Improvement based on redistribution of existing resources when the constraint of total cost is set at the existing level (an increment of

DOI: 10.4148/0146-9282.1186

Table 1
Summary Table of Data for Elmstown School and State

		Staffing					
		Students	Classroom Teachers	Support Staff	Teacher Aides	Adminis-trators	Sum
State	n (total)	100,000	4,000	1,000	750	650	6,400
	Per thousand		40.00	10.00	7.50	6.50	64.00
	Std Devition		5.00	2.00	2.00	2.00	11.00
School	n (total)	1,000	40	10	8	7	64
	Per thousand		40.00	10.00	7.50	6.50	64.00
	Z-Score		0.00	0.00	0.00	0.00	
	Percentile		0.50	0.50	0.50	0.50	
		Salaries					
			Classroom Teachers	**Support Staff**	**Teacher Aides**	**Adminis-trators**	**Sum**
School	Mean		$50,000	$55,000	$25,000	$75,000	
	Total		$2,000,000	$550,000	$187,000	$487,500	$3,225,000
		Student Achievement and Socioeconomic Status					
		Math3	**Math5**	**Read3**	**Read5**	**SES**	
State	Mean	1,400	1,200	1,400	1,300	50	
	Std Deviation	300	200	350	250	10	
School	Z-Score	0.00	0.00	0.00	0.00	0.00	
	Percentile	0.50	0.50	0.50	0.50	0.50	
		Effectiveness					
		Math3	**Math5**	**Read3**	**Read5**		
State	Mean	0.00	0.00	0.00	0.00		
	Std Deviation	0.06	0.06	0.06	0.06		
School	Mean	0.00	0.00	0.00	0.00		
	Z-Score	0.00	0.00	0.00	0.00		
	Percentile	0.50	0.50	0.50	0.50		

zero); or improvement based on a cost increment when the constraint of total cost plus an increment is set.

An Optimization Example

The optimization model is illustrated here using fictitious data from a state and a school building—Elmstown. The purpose of the optimization is to improve the predicted achievement outcome levels by changing the staffing levels in the categories of classroom teachers, support staff, teacher aides, and administrators. For the state data, converting each variable into "staff per one thousand students" normalizes the raw numbers. The means and standard deviations are required in order to calculate Z-scores and percentiles. Also, the mean and standard deviation are required for each of the outcome variables, in this case mathematics and reading at the third and fifth grades, in order to calculate Z-scores and percentiles. The same statistics are required for SES and effectiveness variables for each of the outcomes.

At the school building level, data are required for the number of staff in each category as well as the average salary for each staffing category. With this data, the salary total is calculated (number of staff times the average salary summed across categories). Using the state data, Z-scores and percentiles are calculated for the achievement variables. These data are seeded into an Excel spreadsheet to carry out the optimization. In order to focus on the school input variables, SES and effectiveness variables are set to the mean, or 50th percentile. In an actual example, these data will assure the analysis optimizes the school variables without the influence of the other factors. Table 1 illustrates the state and school data.

Setting Model Parameters

In order to carry out the optimization, two sets of parameters must be added. These estimates do not have to be exact, but do have to fall within a reasonable range. According to Schrage, "The first law of modeling is don't waste time accurately estimating a parameter if a modest error in the parameter has little effect on the recommended decision."[14] The first set of parameters includes the estimates of the relationships between the staffing categories and the multiple outcome variables as measured in terms of the R^2, the proportion of variance explained by each of the staffing variables. The researcher selects these estimates based on ranges produced by regression analysis of the population. There is, however, a mathematical limit to these estimates: The sum may not exceed 1.00. The second set of parameters contains the minimum and maximum levels for each of the staffing categories. These constraints address other practical considerations required by the organization and are selected by the researcher. There also must be a cost constraint, the total amount available to spend.

Calculations in the Equations

The model contains two sets of equations predicting the outcomes before and after the optimization. The before scenario is based on the actual organization values—the predicted target—and the after is based on the optimized values--the optimized target. The calculation for each of the terms (variable times weighting) is particularly noteworthy. The calculation is based on the notion that the best predictor of an outcome is the mean (Z-score = 0, or 50th percentile) when no other information is available. So when some information is available, the calculation is measured by how much the estimate varies above or below the 50th percentile. The calculation for each term is as follows:

$$\text{Term} = R^2 * (\text{Percentile} - .5)$$

The predicted outcome is the sum of the terms plus the 50th percentile. The calculation answers the question: How many percentiles above or below the 50th percentile will the prediction be? The calculation is as follows:

$$\text{Outcome} = \Sigma\ \text{Terms} + .5$$

The optimization process selects new values for each of the staffing categories producing the optimal gain above the predicted target, also known as (the objective function or "gain in target," given several constraints. In this illustration, the major constraint is the total cost of staffing, which must be the same for the before equations and the after, or optimized, equations. Of course, the conditions of maximums and minimums for the respective variables in both equations must be honored. In essence, this scenario is to redistribute the existing financial resources across the staffing categories. If the total cost of the optimized equations were set higher than the before cost, the scenario would be incremental in nature. In Excel, the solver identifies the objective function as the "target cell" and optimum values as "by changing cells." The constraints are identified in under the heading, "subject to the constraints."

Because the optimization is conducted here on a single observation—here a school building—the solution is unique to this building. The regression model implies the same outcome increase for the same change in variable level for every observation regardless of starting point. In contrast, the optimization depends on the unique starting points of each observation, so the amount of increase is always unique.

In order to make the results of the two predicted outcome values as close as possible to the actual outcome values, it is critical to include SES as a variable in the model. In virtually all studies, SES is the highest predictor of student achievement.[15] A measure of school effectiveness is also included to make the predictions as robust as possible.

Return to the Production Function

Earlier, the notion of the production function was introduced. The original conceptualization was:

$$\text{Outcome} = \text{Input} + \text{Process}$$

At this point in the discussion, it has more practical implications. Through the refinement process, the function has become more sophisticated. First, the input has been divided into two categories, the school inputs and the community input of SES. Second, the process element has taken on the character of the effectiveness variable represented by the regression residual. The residual of a regression equation is comprised of an unobserved variable, a variable not in the equation, and error due to the inaccuracies in measurement. Assuming the residual is an unobserved variable of effectiveness, it can be separated from the error by averaging the residual over time. The average is the effectiveness portion, and the difference between the average and the residual is the error. The production function evolves into the form:

$$\text{Outcome} = \text{SES} + \text{Effectiveness} + \text{School Inputs} + \text{Error}$$

For the sake of illustration, assume the SES and Error terms are identical over two periods of time. The function express in terms of change (Δ) is then:

$$\Delta\ \text{Outcome} = \Delta\ \text{School Inputs} + \Delta\ \text{Effectiveness}$$

Consider the following scenario. What if the school input weightings in the optimization are inflated or raised higher than what might be considered reasonable? The predicted optimized target will then increase, but what if the actual outcome level does not increase at the same pace? The equation demands balancing, so effectiveness declines. Simply stated, within the rigors of the mathematical model, any overstatement of school inputs will be offset by an decrease in the level of school effectiveness. Hence, attempts to "game the system" by inflating inputs will have the consequence of being labeled less effective.

Table 2
Range of Weightings

Subject	SES	Effectiveness	Error
Math3	0.532	0.381	0.087
Math5	0.635	0.297	0.068
Reading3	0.712	0.223	0.065
Reading5	0.706	0.226	0.068
Mean	0.646	0.282	0.072

http://newprairiepress.org/edconsiderations/vol35/iss2/2
DOI: 10.4148/0146-9282.1186

Table 3
Optimization Results

Table 3.1 Original Values and Optimal Values										
	SES	Effectiveness	Classroom Teachers	Support Staff	Teacher Aides	Adminis-trators	Total Cost	Incre-ment		
Original Values	50.00	0.00	40.00	10.00	7.50	6.50				
Percentile	0.50	0.50	0.50	0.50	0.50	0.50				
Cost			2,000,000	550,000	187,500	487,500	3,225,000	100,000		
Optimized Values	n/a	n/a	44.42	5.00	10.63	7.51	3,325,000*			
Change			4.42	-5.00	3.13	1.01				
Z-Score			0.88	-2.50	1.57	0.51				
Percentile			0.81	0.01	0.94	0.69				
Cost			2,220,846	275,000	265,806	563,348	3,325,000*			
Minimum			35	5	5	3				
Maximum			50	15	15	10				
*Must be equal										

Table 3.2 R-Square with Goal										
	SES	Effectiveness	Classroom Teachers	Support Staff	Teacher Aides	Adminis-trators	All School	Total	Error	Total
Math3	0.600	0.2500	0.030	0.020	0.010	0.020	0.080	0.930	0.070	1.000
Math5	0.600	0.2500	0.035	0.020	0.010	0.020	0.085	0.935	0.065	1.000
Reading3	0.650	0.2000	0.035	0.020	0.020	0.010	0.085	0.935	0.065	1.000
Reading5	0.650	0.2000	0.030	0.020	0.020	0.010	0.080	0.930	0.070	1.000
Average	0.617	0.233	0.033	0.020	0.013	0.017	0.083	0.933	0.067	1.000

Table 3.3 Predicted Target										
	SES	Effectiveness	Contribution				School	Pre-dicted	Actual	Effi-ciency
Math3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	50.00%	50.00%	0.00%
Math5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	50.00%	50.00%	0.00%
Reading3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	50.00%	50.00%	0.00%
Reading5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	50.00%	50.00%	0.00%
Sum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
Average	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	50.00%	50.00%	0.00%

Table 3.4 Optimized Target										
	SES	Effectiveness	Contribution				School	Pre-dicted		
Math3	0.0000	0.0000	0.0093	-0.0099	0.0044	0.0039	0.0078	50.78%		
Math5	0.0000	0.0000	0.0109	-0.0099	0.0044	0.0039	0.0093	50.93%		
Reading3	0.0000	0.0000	0.0109	-0.0099	0.0088	0.0019	0.0118	51.18%		
Reading5	0.0000	0.0000	0.0093	-0.0099	0.0088	0.0019	0.0102	51.02%		
Sum	0.0000	0.0000	0.0405	-0.0395	0.0265	0.0116	0.0391			
Average	0.0000	0.0000	0.0101	-0.0099	0.0066	0.0029	0.0098	50.98%		
Gain in Target	0.0000		0.0405	-0.0395	0.0265	0.0116	0.0391	0.98%		

Ranges of Relationship Weightings

There is no fixed set of weightings measuring the relationship between the outcome and the model variables. Every study will produce different estimates. Nevertheless, most studies fall within some consistent range. The author has not completed a thorough study to document these ranges, but based on data from one state, these ranges, measured in terms R^2 or percentile points, seem to be justified. (See Table 2.) In this state, the influence of SES tends to be about 10 points higher for reading than for mathematics while the influence of effectiveness tends to be about 10 points higher for mathematics than for reading.[16] Each investigator will have to determine a range based on what data are available for the population under study. The consequence of overestimating has already been addressed.

After the data have been entered into the spreadsheet model and the optimization conducted, the results can be presented in a format illustrated by Table 3.

Summary of Results and Analysis

All the school variables in this illustration were set to the mean to more easily focus on the features of the optimization. Therefore, the predicted target and actual outcome levels were all at the 50th percentile. In a real situation, these variables will reflect the actual status of the school. When the optimization is applied, the optimized values are indeed changed in that there is an increase in the more cost-effective variables and a decrease in the less cost-effective variables. The total cost of the pre-optimization and post-optimization is equal, thus an incremental scenario. There is an incremental value that could be set to zero by the researcher for a redistribution scenario. The constraints have been met in that the support category is at the minimum. The gain in the predicted gain in target is an average of .98 percentiles.

The optimization also produces some analytical information of potential usefulness. The contribution of each of the variables for each of the outcomes is provided indicating the respective cost-effectiveness. The average contribution for each of the variables is also provided. The contributions of the school variables are provided separately. There is a check of the R^2 sum to assure that it is not greater than 1. The sum of the R^2 of the school input terms is provided to assure it falls within a reasonable range. A measure of efficiency is given, calculated as the difference between the optimized predicted outcome and the actual outcome level. It could be considered error or doing better (or worse) than predicted. As this example demonstrates, there is a mathematical solution to the stochastic simultaneous equations model. Only by building and interrogating a "live-data" model with all of the policy relevant variables will it be known if there is a practical policy solution.

Observations Regarding the Optimization Model

Modeling through Estimates

There will never be enough comprehensive and accurate data. Realistically, data can be used to make estimates of relationships between outcomes and input variables; however, these estimates will always vary over time and populations. Importantly, this optimization model is most effective when realistic ranges of the relationships are examined. Because the cost of a variable is known with great accuracy, it is logical that there is an implied relationship between the cost and the cost-effectiveness of the variables. That is to say, if variable A is three times as costly as variable B, then variable 1 must be three times as effective for the two variables to be equally cost-effective

Setting the relationship variables first produces the predicted target level. Importantly, the higher the relationship, the higher the predicted target values. This is not a "freebie," in that the actual relationship values are, by definition, set so half of the observations will do better than predicted and half will not. This difference is in small part due to error in data measurement, but mostly the difference is due to the inescapable fact that some organizations are more effective in turning resources into outcomes. Therefore, if the relationship variables are set too high, indicating that more resources will produce higher predicted outcomes, it will also tend to increase the gap between the predicted outcome and the actual outcome, indicating a higher degree of ineffectiveness. Increasing the relationship coefficients will have the effect of indicating higher potential achievement scores for greater resources, but it will also render the school less effective when the actual results are measured and the school fails to meet the prediction. In essence, the greatest value is achieved when the parameters are set realistically rather than quixotically.

Inevitable Conclusions

As outlined above, there are some inevitable conclusions associated with the optimization model as compared with the regression model. First, because of the inherent nonlinear structure of the optimization model, it is impossible to achieve entirely the desired goal unless the goal has been completely achieved by other similar organizations. That is to say, it is impossible to set values predicting a perfect outcome score unless it has been actually achieved by other organizations, and the Z-score for that organization can be identified. In terms of student achievement testing, it is highly unlikely any organization records perfect scores for all students.

Second, there is an inherent point of diminishing returns due to the nonlinear stochastic function. At a certain point, any given variable will have reached its potential, and investments in other variables will indicate better results. As a general rule, if an organization is among the highest on a given variable when compared to other organizations, an increase the variable will indicate little increase of outcome in the model. On the other hand, an increase in a variable for which the organization is low as compared to others will indicate a larger increase of outcome. Of course, the variables must be compared based on the cost-adjusted value.

Third, as suggested by point two, the solution to the model will be different for each organization, because the starting point is unique to each organization. Theoretically, if all organizations were moved to the high end (for example, the third standard deviation above the mean) for all variables, the predicted results for all organizations based on the allocation of resources would be similar. Any differences in predictions would be based on variables not included in the resource allocation category such as socioeconomic status or effectiveness. In other words, achievement equity is not possible solely through resource allocation. For complete outcome equity, resources, SES, and effectiveness must all be equal.

The optimization model has two basic strategies: (1) Invest in high cost-benefit variables where the organization level is low compared to other organizations; and (2) Do not invest in low cost-benefit variables where the organization level is high compared to other organizations.

http://newprairiepress.org/edconsiderations/vol35/iss2/2
DOI: 10.4148/0146-9282.1186

Ranges of Input and Process Categories

In supplying the estimate of weighting in the equations, these conditions must be recognized. First, there is a maximum of an R^2 of 1.00. Second, if the estimated weightings are larger than the actual weightings, the effectiveness ratings of the observations will be reduced; that is, the actual performance on outcome will be less than the predicted outcome level. In theory, the weightings will be close to correct when the effectiveness of all observations is normally distributed with a mean of zero. Over the last several decades, educational research has identified several categories thought to be associated with student learning outcomes. The community and school inputs are: SES; staffing quantity (ratio of various staff classifications to students); (3) staff quality (qualifications, experience, etc); and (4) materials and supplies. Less attention has been paid to the process categories of instruction, including time, curriculum, out-of-school influences; and effectiveness. A comprehensive model could include all these independent variables as long as there are data defining the variables and statistics estimating their relationships with outcomes. While outcomes are usually defined by student achievement measures, other desirable outcomes such as dropout rates and college-bound rates could be included in the model as long as the data for the variables and estimates of the relationships are available. Because there tends to be a high degree of correlations among school variables, adding variables to the model does not always have the effect of increasing the predicted levels of the outcomes. Instead, adding variables merely redistributes the influences. Also, because of the correlation between some school variables and SES, it is appropriate to test the model within reasonable ranges.

Testing the Model

There are some elements of school operations for which there are no estimates of the relationship with outcomes. Probably the best example is that of the school year. Mostly because of state laws, virtually all schools are in operation for the same amount of time. Because there is little variation, there can be no estimated relationship in a regression analysis. But there are options within the optimization model. First, the cost of an extension of the school year can be calculated. Second, the cost can be compared with the cost of other options where the relationship with outcome is estimated. With this information, a calculation can be made as to the relationship level of extending the school year to make an equal contribution as the other option. In a more ideal situation, a national or state research initiative could be conducted by first applying the optimization model and then applying an experiment—in this illustration, a longer school year—to determine if the estimates in the model are realized. Surely this is a more practical method than instituting a statewide policy without any experiment evidence.

Sensitivity Analysis

There is a notion of opportunity cost developed by accountants. Simply, it is how much profit can be gained by increasing production by a given amount. In the optimization illustration, a marginal cost-benefit is provided for each element within the model indicating how much would be gained in student outcome by a certain investment. Obviously, it would be appropriate to invest in the element with the highest cost-benefit. However, the cost-benefit will not be the same for each school because each school has a unique starting point.

Summary, Research, and Policy Issues

The model used for investigating school resource allocation questions has a definite influence on the policy conclusions reached. At the beginning of this article, three potential benefits of building a model were identified. First, building a model often reveals structures, elements, and relationships usually taken for granted until the underlying assumptions are stated and tested. Once the original ideas are stated and tested, they usually give way to more sophisticated and accurate representations of the actual situation. Second, once the model is constructed, analyzing it mathematically suggests different courses of actions not readily apparent. In essence, the model becomes a challenge to conventional thinking. Third, experimentation that is not practical in actual situations is possible within a model. Through experimentation, more potentially successful options may be identified. In essence, models can be tested, unlike "seat-of-the pants" decisions. Now it is time to assess if any of these potential benefits have been realized through the process of building an optimization model.

Underlying Assumptions of the Optimization Model

In building this optimization model, the structures and relationships of other models were analyzed and their underlying assumptions challenged. The optimization model makes different assumptions and, most importantly, the model defines the relationships between outcomes and inputs differently.

The fundamental assumption regarding education is that it is stochastic in nature because the goals of education are mostly measured by student achievement tests having theoretical and practical upward limits. The critical step in actually building the optimization model was identifying the mathematical function fitting the stochastic nature of education to a diminishing returns curve rather than a constant returns line. Considerable attention was paid to the mathematical evidence demonstrating the existence of a diminishing returns curve derived from a transformation of the regression analysis. Using the principles of mathematical programming, it was possible to: (1) Incorporate these diminishing returns curves into multiple regression equations representing the simultaneous educational goals; (2) Incorporate additional equations reflecting the constraints on the organization, most importantly, cost; and (3) Develop the methodology for finding feasible solutions to this optimization model. The optimization model is more sophisticated than other models because these concepts are incorporated; and because they are incorporated, the optimization model more accurately represents the actual situation.

Observations Regarding the Optimization Model

The generalized results of the optimization model suggest different courses of action challenging conventional thinking in several ways. First, there is a unique resource allocation strategy for every school, depending on its starting conditions, rather than a common strategy applying to all schools as is the case with other models. Second, while additional resources can make some difference, merely adding educational resources will never completely overcome the influence of SES or the shortcomings in organizational effectiveness. This distinguishes the optimization model from those that resources can overcome all other shortcomings. Third, in some cases, more is better, but in other cases more (e.g., money) produces little or no increased in benefits. In other models, more is always better. Unquestionably, these findings are in direct contrast to the conventional and somewhat "seat-of-the-pants" thinking prevalent in education today.

Identifying and Testing Potentially More Successful Options

There are many "ifs" in model building. In this case, there is the question of whether the stochastic model presented here has greater logical and mathematical merit than other models. Next is the question of the accuracy of magnitude of the relationships presented. Are the estimates of the influence of SES, school effectiveness, and school inputs reasonable? Assuming the responses to these questions are in the affirmative, then there is the inescapable question: Why focus so much attention on the allocation of school resources when the largest impact on student achievement will come through improving school effectiveness and addressing the issues associated with community SES?

SES poses its own set of problems. First, SES is not a changeable "thing," at least changed in a way that relates to student achievement. SES is a concept, and researchers employ proxies to measure the concept. The measure usually includes, for example, income, education levels, and verbal aptitude of the mother. No one seriously proposes policy changes in these variables in order to improve student achievement. More likely, the concept of SES represents a set of behaviors associated with families and communities where students test favorably. Is it the amount of time devoted to reading or homework, or the amount of time not devoted to television? Is it the amount of time parents spend talking with their children about school or the amount of time a family engages in serious discussion about the importance of an education? We do not know. It does seem potentially rewarding, however, to find out more about these behaviors and then devise programs for schools, communities, religious organizations, and social service agencies to become more engaged in an way that is likely to bring more success.

Education is not well-suited for testing the optimization model—or any model—through experimentation. State laws, professional attitudes and traditions, and public opinion make it all but impossible to adopt the conclusions of the optimization model into practice. Some expectations of change have been placed on charter schools, but the evidence is not hopeful. Perhaps the critical question is whether using a different model—an optimization model—can have an impact on lawmakers' actions, professional attitudes, and public opinion?

The Optimization Model as a Paradigm

This article was heavily influenced by Kuhn's ideas and, especially, his thoughts regarding a "paradigm shift" in *The Structure of Scientific Revolutions*.[13] The optimization model in the context of a paradigm has a larger purpose: To put all the individual pieces of an educational organization into a single, comprehensive, and logical framework, much like particle physics and the "Standard Model." With such a framework in place, it is possible to make more sophisticated inquiries and predictions. The results then become the empirical basis for policy decisions. The driving force for a new model was the anomaly presented by regression analysis; that is, regression could not accommodate all the elements and outcomes of the organization simultaneously, and it could not comprehensively respond to the best use of resources questions.

The intent of the optimization model as a paradigm is to demonstrate its greater robustness compared to its competitors in that it substantially adds scope and precision to the "what if" questions. In addition, the model establishes a framework for future research. First, it builds upon the idea of the production function by adding the element of effectiveness with a theoretical basis and a practical method for its measurement. Second, it incorporates a reformulation of the regression statistics into a type of glue serving to hold the multiple outcomes together with the multiple elements in a comprehensive and mathematically logical way. Finally, it incorporates a mathematical programming methodology for modeling the intricacies of the educational organization.

What is missing? There seem to be at least three major pieces missing for a concerted research strategy: (1) A conceptual structure guiding research efforts; (2) a set of reliable and replicated measurements of the structure elements and their relationship with outcomes; and (3) methods to address technical shortcomings.

Other sciences have conceptual structures guiding research efforts. While there are many illustrations, the periodic table from chemistry serves as an instructive analogy. The periodic table identifies the basic chemical elements by their measurable characteristics. Based on these characteristics, research is directed toward understanding how they interact with one another in more complex situations. What if there were a comparable conceptual structure for educational organizations? What if there were a consensus regarding the structure and elements of the educational organization along the lines presented herein? It would encourage the direct comparison of research results—a type of unification. Like chemistry, additional elements could be included as their unique characteristics and contributions are identified and measured. With a consensus of the structure and elements of an organization, research would focus on what is in common among organizations so the anomalies could be identified and addressed.

What if there were a comprehensive set of measurements estimating the characteristics of these elements and their relationship to outcomes? While they would not be exact, as they are in chemistry, they would fall within ranges, and these ranges would be valuable in seeding the optimization model. While they will undoubtedly be difference estimates, there is no reason to believe the underlying effect of staffing quality or staffing quantity would be different due to the school district or state of residence of a student. Most likely, it is the unique combination of factors making the difference. Therefore, the key is to identify those underlying factors, their magnitudes, and their relationships.

What if there was a concerted effort to address some of the technical shortcomings of this and other models—the multicollinearity among variables, for example? For example, it may be possible to incorporate the multicollinearity into the optimization model by adding defining equations.

Walberg worked on developing a comprehensive framework for the analysis of productivity starting in 1975.[14] (While he developed a method of measuring relationships between outcomes and school variables—effect size—he neither proposed an economic adjustment nor an optimization method.) Levin addressed the important relationship of cost-effectiveness with educational policy,[15] and Monk described the pro's and con's of the production function.[16] The optimization model builds on Walberg's plea for a comprehensive framework, Levin's push for cost-effectiveness, and Monk's call for greater sophistication in the production function.

With these caveats in mind, the ultimate value of this model is its potential for becoming a paradigm for the continued pursuit of educational productivity.

http://newprairiepress.org/edconsiderations/vol35/iss2/2
DOI: 10.4148/0146-9282.1186

Endnotes

[1] Linus Schrage, *Lindo: An optimization Modeling System* (San Francisco: Scientific Press, 1991).

[2] Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1970).

[3] H. P. Williams, *Model Building in Mathematical Programming* (New York: Wiley & Sons, 1985).

[4] Ibid., 3.

[5] There are also criterion referenced assessments with the purpose of identifying those students who have achieved a minimum academic standard. The distribution of these assessments is not normal, but more of a "J-shape" with a tail at the lower end and a large grouping at the high end. However, when the assessment scores are grouped by building, they tend to take on the shape of the normal curve.

[6] For a thoughtful and comprehensive discussion of the production function, see David H. Monk, (New York: McGraw-Hill: *Educational Finance: An Economic Approach*, 1990).

[7] James L. Phelps and Michael F. Addonizio, "How Much Do Schools and Districts Matter? A Production Function Approach to School Accountability," *Educational Considerations* 33 (Spring 2006): 51-62. .

[8] Ibid.

[9] For those who counter with the Glass and Smith meta-analysis on class-size, see Appendix A.

[10] J. P. Gilford, *Fundamental Statistics in Psychology and Education* (New York: McGraw-Hill, 1965), 399.

[11] In Excel, the function is NORMDIST.

[12] Gilford, 394-400.

[13] It maybe helpful to compare Figure 6 with Figure 1.

[14] Schrage, 8.

[15] Combining several variables into an SES index would make the model structure more straightforward.

[16] Note that the school variables were highly correlated with the SES variable.

[17] Kuhn, *The Structure of Scientific Revolutions.*

[18] Herbert J. Walberg, "Improving the Productivity of America's Schools," *Educational Leadership* 41 (May 1984):19-27.

[19] Henry M. Levin, "Cost-Effectiveness and Educational Policy," *Educational Evaluation and Policy Analysis* 10 (Spring (1988): 51-69.

[20] Monk, *Educational Finance.*

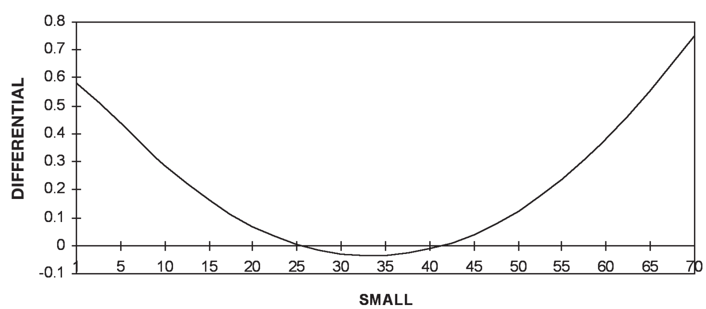
Appendix A
Observations Regarding Meta-Analysis of Class-Size

For those who might cite the class-size meta-analysis by Glass and Smith as an example of increased returns to scale rather than diminishing returns, they may wish to consider the following.[1] First, the equation Glass and Smith used to plot the frequently cited curve included a squared term, indicating the plot is a parabola. When fully plotted across the entire class-size range in the data, the achievement prediction for a class-size of 60 was the same as for a class-size of 10, with the minimum being a class-size of about 32. Because the data included substantial observations of class-size above 40, the full curve should be considered when drawing conclusions rather than just the "attractive" side of the curve. Second, because the report included the data, a re-analysis is possible. When this author conducted a re-analysis, no relationship was found between class-size and achievement levels when the range was restricted to class-sizes between 10 and 60. Third, the class-size scale is not equal interval; therefore it would take four times as many teachers to reach a class-size of 10 starting at 40 as it would to reach 20.

When looking at the entire curve, three first-impression questions come to mind: (1) Can it be that a class-size of 65 will produce the same results as a class-size of 1? (2) What will be the results if there were more teachers than students in the class- would achievement continue to improve? (3) At what class-size does the left-hand side of the curve level off or is perfect achievement attainable? (See Figure A.)

[1] Gene V. Glass, and Mary Less Smith, *Meta-analysis of Research on the Relationship of Class-size and Achievement* (Portland, OR: Far West Laboratory of Educational Research and Development, 1978).

Figure A
Range of Weightings



Appendix B
What Makes Education Stochastic?

After describing much of the details of the stochastic model, it may be useful to revisit the reasons why education evaluation is stochastic. Student achievement tests are based on the properties of the normal or probability curve and administered to students usually during the same grade in school producing another normal-like distribution. This is unlike most outcome measures in other organizations. Therefore, the relationship between student achievement and independent variables should also be based on these same properties. What are these properties?

First and most importantly, the normal curve is bounded. While the curve actually extends from minus infinity to plus infinity, both arms are asymptotic to the abscissa; that is, while the extreme values may closely approach the boundary, they never do. If in a mathematical model the boundaries could be reached, there would be the "out of bounds" paradox. In the case of education, it would mean all students can be above average, and under some circumstances all students can be perfect. Because this is not the case in practice or in theory, modeling education with stochastic functions more appropriately resembles reality. Second, the relationship among normally distributed variables is nonlinear, a critical condition for solving simultaneous equations. Third, when the predicted results are presented in terms of percentiles, one may answer the question: What are the chances the result will be achieved when the conditions of the model have been met? As the following illustration will show, the changes are limited largely because of the SES element and, to a lesser degree, school effectiveness. In contrast, the regression model implies a 100% chance of achieving perfection given enough resources, regardless of SES or effectiveness.

Because of the stochastic nature of student achievement testing, there is a fundamental difference in how schools are judged compared to most other organizations. All widget-making companies are thought to be successful as long as they stay in business; there is no stochastic judging scheme. While there have been other attempts to judge the performance of schools--for example through accreditation—with the current emphasis on standardized testing, schools have been relegated to a unique fate prescribed by the normal curve.

DOI: 10.4148/0146-9282.1186