An Approach to Evaluating Relative Effectiveness in Non-Profit Institutions

By

Yaw M. Mensah*
Rutgers Business School ,
Rutgers University,
New Brunswick, NJ 08854,

mensah@rbsmail.rutgers.edu

Kevin C. Lam
Faculty of Business Administration,
Chinese University of Hong Kong
Hong Kong

kevinl@msmail.cuhk.edu.hk

and

Robert Werner
Rutgers Business School,
Rutgers University,
New Brunswick, NJ 08854
,
werner@rbsmail.rutgers.edu

April, 2005

* Corresponding author:

An Approach to Evaluating Relative Effectiveness in Non-Profit Institutions

Abstract

We present, in this study, a method for comparing the relative effectiveness of different non-profit institutions with similar objectives. In addition, we show how this measure of relative effectiveness is related theoretically to their relative efficiency. Finally, drawing on developments in data envelopment analysis, we illustrate the new methodology using data from 107 institutions of higher education.

Keywords: Effectiveness; Efficiency; Non-profit institutions; Performance evaluation

1. Introduction

The evaluation of organizational effectiveness has a long history that cuts across the disciplines of economics, operations research, personnel and organization management, and strategic management, as well as accounting. Economists, operations researchers and accountants have tended to focus on organizational efficiency, while personnel management, organizational and strategic management researchers have tended to specialize in the measurement of organizational effectiveness.

While most scholars of organizations will agree that organizational performance encompasses both effectiveness and efficiency, the tendency for the different disciplines to focus on different aspects of performance is understandable for a reason that becomes apparent once the terms are defined formally. The commonly accepted definition of *effectiveness* is that it measures the degree of success with which organizational goals are achieved (see, for example, Fink, Jenks and Willets 1983). Typically, *efficiency* is defined as "the attainment of a desired outcome or goal with a minimum of effort, cost or waste" (Reed 1991). Thus, while efficiency involves the quantitative ratio of inputs to outputs, effectiveness involves the softer concept of organizational goals versus organizational achievements. The disciplines where optimization concepts are emphasized can thus easily deal with the problems of efficiency measurement, but cannot easily handle the evaluation of effectiveness, particularly in areas like the non-profit sector where output/outcome measures are often not quantifiable. The management disciplines have filled this void with softer constructs which, unfortunately, do not lend themselves easily to quantification. Thus, the exact theoretical linkage between effectiveness and efficiency is lacking in the literature.

The purpose of this paper is to advance a quantitative measure of relative effectiveness of different organizations with similar goals, and to show how effectiveness is linked theoretically to efficiency. We illustrate the concepts presented by employing the data for 107 private institutions to show the relationship between the concepts.

In the sections that follow, we first review the literature on the measurement of effectiveness and efficiency in the different literatures in Section 2. In Section 3, we present our conceptual model that shows how relative effectiveness of non-profit institutions can be measured quantitatively. This construct is then related to relative efficiency In section 4, we provide empirical estimates of the relative effectiveness and efficiency of a sample of private IHEs to illustrate the feasibility of the approach developed in the paper. Our conclusions are presented in the final section 5.

2. Effectiveness and Efficiency Measurement

We present in this section an overview of the literature on the measurement of organizational effectiveness and efficiency, and identify the areas where this study can make a contribution. Section 2.1 reviews the literature on effectiveness, and Section 2.2 on efficiency.

2.1 Effectiveness

It is a mild understatement to state that few topics that are of such fundamental importance and have attracted such attention from both academics and professionals are still subject to as much confusion as the concept of effectiveness. It is perhaps indicative of the confusion surrounding the topic that, in a much-cited synthesis of extant work on the subject, which led to yet another measurement approach, the competing values approach, Quinn and Rohrbaugh (1983) state in their conclusions:

> Organizational effectiveness is not a concept. It is a socially constructed, abstract notion carried about in the heads of organizational theorists and researchers. (page 374).

The unsatisfactory state of affairs led many researchers in the past to call for a moratorium on organizational effectiveness studies (Goodman, Atkin and Schoorman 1983; Hannan and Freeman 1977). As pointed out by Cameron (1986), however, the evaluation of organizational effectiveness is far too important to everyday usage for the failure of academic researchers to agree on measurement criteria to result in its abandonment. Furthermore, the problem of lack of appropriate criteria is acute only in the nonprofit sectors of the economy, since in the for-profit economic sectors, many observers will agree that profitability and customer satisfaction constitute acceptable dimensions on which effectiveness can be measured (Joys 2001).

A brief summary of the most important operational constructs developed to measure organizational effectiveness are the following four major approaches: (1) the goal model of effectiveness (Campbell 1977; Scott 1977); (2) the system resource approach (Yutchman and Seashore 1967; Pfeffer and Salancik 1978); (3) the multiple constituencies approach (Connolly, Conlon and Deutsch 1980; Zammuto 1982); and the spatial model/competing values framework (Quinn and Rohrbaugh 1983; Lewin and Minton 1986).[1]

The "goal model of effectiveness" approach adopts the traditional view that organizational effectiveness should be evaluated with reference to the declared objectives of the organization. The "system resource" approach assesses effectiveness by the extent of resources the organization is able to marshal. The "multiple constituencies" approach recognizes that an organization has multiple constituencies who are likely to use different criterion to assess its effectiveness. Finally, the "spatial/competing values" approach is a synthesis of four different approaches to effectiveness measures: the human-relations model, the open-system model, the internal-process model, and the rational-goal model. Out of this synthesis emerges the view that these diverse criteria relating to productivity/efficiency, stability/control, resource acquisition/external support and the value of human resources be combined. This requires that values be articulated, weights placed on the values, and a formula be developed for combining the weighted scores on each criterion.

One difficulty with the existing models of efficiency is that the concept *of relative effectiveness* is left fuzzy and ill-defined. Quite clearly, in making judgements of institutions, individuals do comparative evaluations of different institutions. None of the proposed criteria for evaluating organizational effectiveness, however, provides the basis for making comparative judgements of relative effectiveness. If an organization's effectiveness is judged solely by reference to its declared objectives, the perceptions of its multiple stakeholders, the resources it is able to acquire, or a complex combination of the above, comparisons across different organizations with even the same objectives become problematic. For an organization can choose objectives that fall short of (or far exceed) what is attainable given its potential, and stakeholder perceptions of its performance may not reflect a comparative knowledge of what is attainable with the resources provided.

These deficiencies in the existing state call for a new approach, one that attempts to introduce more rigorous comparative methodology into the assessment of relative effectiveness. To do so, a review of the development of the concept of efficiency as a measurement construct is first necessary. This review is provided in the next section.

2.2 Efficiency

Essentially, efficiency is a question of relating actual inputs used to the actual outputs/outcomes generated. The conceptual basis of the measurement of efficiency may be said to have originated with the development of "activity analysis" by Koopmans (1951), and its empirical operationalization by Farrell (1957). The specific identification of possible inefficiencies in production originated with Debreu (1951). The linear programming reformulation of activity analysis was demonstrated by Charnes and Cooper (1957, 1961), and the final evolution to data envelopment analysis (DEA) by Charnes, Cooper and Rhodes (1978, 1981).

With the hundreds of articles that have been devoted to DEA and its various extensions, it is fair to state that the measurement of relative organizational efficiency is no longer a conceptual problem. What remains to be resolved is exactly how efficiency is related to effectiveness.

A start in this direction has been made by Reed (1991) when, in the context of diversification, he argued that bimodality in measures of diversification is due to organizations trading off efficiency against effectiveness. He argued that both efficiency and effectiveness in diversification are subject to the laws of diminishing returns approached from different directions, and thus require joint optimization. Unfortunately, Reed's model does not provide any basis for resolving the difficulties in the measurement of effectiveness discussed earlier. Nevertheless, his model provides one way to visualize how effectiveness and efficiency are related. In the model presented below, we provide an alternative way in which efficiency is embedded in the concept of effectiveness.

3. A Model for Measuring Relative Effectiveness

In non-profit environments, three key characteristics frequently play a dominant role in organizational performance. The first characteristic relates to the organization's resource-acquisition ability, the second to the attainability of the organization's goals, and the third to organizational efficiency. Regardless of whether the organization charges for its services or not, because of the usual requirement of long-run financial breakeven operations, the quantity and quality of the organization's output/outcomes depend vitally on the organization's ability to raise revenue or otherwise acquire resources that it can then convert into inputs.

At the same time, the open-ended or closed-ended nature of the organization's goals is an important determinant of the effort required of the organization. If the goals are closed-ended, their achievement caps the effort required, and the issue of relative organizational effectiveness is then moot. If the goals are open-ended, then it reduces to one of output/outcome maximization subject to the resources available to achieve those goals. This is a much richer situation, and the discussion below assumes such to be the case.

3.1 Measuring Relative Effectiveness

The fundamental shift in thinking required to develop the notion of relative organizational effectiveness is the idea that, similar to efficiency, relative effectiveness requires a comparison of resources to outcomes. There are, however, two key differences. The first difference is that, unlike the case of efficiency, for the measurement of relative effectiveness, the resources under consideration are not the actual resources used by the organization, but all potentially usable resources. By "potentially usable resources" (hereinafter called PURs), we mean potential assets (including intangible ones) that, by creative thinking, could be converted to inputs that are utilizable by the organization.

The second key difference relates to achievable outputs/outcomes (for relative effectiveness) versus actual outputs/outcomes for relative efficiency. By "achievable outputs/outcomes" we mean the outputs/outcomes that an effective and efficient organization can generate if all the potentially usable resources were deployed and utilized in pursuit of the objectives of the organization. The rationale for substituting “achievable output/outcome” for the actual targeted objectives of the organization is that the measurement of relative effectiveness should be independent of any systematic biases in judgments that organization may make in selecting their targeted levels of output. If this were not the case, an organization with vast resource endowments that (by deliberate choice or failure to gauge its potential) chooses low targeted outcomes would be rated effective merely, because it achieved those objectives. At the same time, a more ambitious organization with higher targets relative to its resource base would be rated “ineffective” because it failed to achieve those lofty goals.

****
Insert Exhibit I here
****

A schematic representation of the proposed model of relative effectiveness and its relationship to efficiency is presented in Exhibit I. As shown in Exhibit I, the PURs (Z_i) have to be transformed into inputs that are used by the organization. In the nonprofit situation, revenue-generation is a primary consideration, so the level of revenues and other resources that can be raised from the PURs is the primary issue. The revenues and other resources are inputs that are used by the organization to generate outputs/outcomes. The rate at which the revenues, once raised, are converted to current outputs determines the organization’s efficiency rating.

One important factor here relates to the relationship between revenues and costs. Although, in the long term, revenues raised must equal cost expenditures for nonprofit institutions, in the short term, this equality need not hold. This is because, to survive for the long term, a nonprofit institution must build financial reserves (i.e., a ‘rainy day’ fund) to offset temporary imbalances between revenues raised and current expenditures needed to sustain operations at desired levels. Thus, while the PURs must be related to revenues and other input resources raised (denoted X^R_A), the determination of organizational efficiency involves relating costs and other inputs (denoted X^C_A) to the current level of outputs (denoted Y^C_A).

Differences between X^R_A and X^C_A lead to the build-up or draw-down of reserves (denoted ΔR_A). Actual output (Y_A) then consists of Y^C_A plus ΔR_A.

The organization's own goals or objectives can be used to identify the appropriate cohort or peer group. Once this peer group is identified, the organization's relative effectiveness (compared to the peer group) can be evaluated by the degree to which its actual output compares to the maximum output (Y_{MAX}) that could be generated given the potentially usable resources of the organization. Y_{MAX} is again defined to the the level of maximum current output (Y^C_{MAX}) plus maximum buildup (or minimum drawn-down) of financial reserves (ΔR_{MAX})

To determine the PUR domain, a researcher has to identify all resources that, conceptually, could be converted into usable inputs (given sufficient creativity). Some guidance on this can be provided by a study of the activities of the peer group in obtaining inputs (a kind of benchmarking). An effective non-profit organization (with open-ended goals) is one that is able to convert as much PURs as possible into usable inputs in order to bring its actual output level to the maximum output level. This implies that organizations are seeking to maximize output levels, subject to PUR constraints.

In summary, the concept of relative effectiveness implies output/outcome maximization subject to PUR constraints. In situations where the organization has specifically limited output objectives, and existing resources are sufficient to achieve those objectives (regardless of efficiency), relative effectiveness is not a meaningful concept. In such situations, only relative efficiency is pertinent. In the more general situation where an organization is faced with unlimited demand for its services/products but has limited PUR, its relative effectiveness can be evaluated by the actual output it is able to generate relative to its maximum achievable output. Since maximum achievable output is defined relative to PUR, effective organizations (in resource-constrained environments) are those that are able to transform as much of these PURs into usable inputs in order to push its output/outcomes to the limit.

Since the achievement of high relative effectiveness (in resource-constrained environments) involves the conversion of untapped resources to usable inputs in order to increase the output capability of the organization, returns to scale and scope are particularly important. Naturally, one would expect that the resources that are most easily exploited are the ones that are initially tapped. Thus, one would expect the usual increasing returns to scale and scope in the early ranges of the input/output transformation process. As these resources are more fully exploited, the constant-returns-to-scale ranges are reached. Attempts to further exploit the unutilized PUR then gets an organization into the decreasing-returns-to-scale and -scope regions. This performance degradation, once certain levels of output are achieved, is to be expected, as less desirable resources are forced into use. A graphical representation of the expected relationships is depicted in Exhibit II.

****
Insert Exhibit II here
****

In Exhibit II, the three curvilinear lines shown are VRS(variable returns to scale) production functions for Most Efficient (ME), Average Efficiency (AE), and Least Efficient (LE) organizations. The straight line from the origin is the CRS (constant returns to scale) production function for the ME organization. Points ME, AE and LE denote the maximum outputs that can be generated by the ME, AE and LE entities, respectively, when the potentially usable resources are fully deployed. Only the most efficient organizations can achieve the Y_{MAX} output level.

The relationships between relative ineffectiveness, relative inefficacy, and relative inefficiency are apparent from Exhibit II. Relative ineffectiveness is measured by the distance between Y_{MAX} and Y_A^{AE} assuming a firm operates along the production function represented by AE. The distance from Y_A^{AE} to Y_A^{ME} measures the traditional output-oriented inefficiency. Relative inefficacy for an organization operating along the production function AE is therefore measured by the distance from Y_A^{ME} to $Y_{MAX.}$. X_A represents the actual set of inputs used, and the differences in output levels (Y_A^{ME}, Y_A^{AE}, and Y_A^{LE}) reflect the differences in organizational efficiency.

Thus, if an organization operates along the production function ME, there will be no relative inefficiency, and relative ineffectiveness is then equal to relative inefficacy. Therefore, an organization operating along production function ME that is also able to equate X_A to X_{MAX} will attain zero inefficiency, zero inefficacy, and thus zero ineffectiveness. Conversely, a firm operating along production function LE will have a higher level of relative inefficiency, and consequently higher ineffectiveness, although the level of inefficacy will not be affected as long as there is no shift in $X_{A.}$ It is apparent that the more an organization is able to tap its potentially utilizable resources, the more relatively effective it is, with efficiency held constant.

In conclusion, to obtain a perfect relative effectiveness score of unity, an organization has to fully exploit all potentially usable resources inputs *and* also be efficient. An efficient organization does not have to be perfectly effective, and a relatively effective organization is not necessarily the most efficient. A perfect score on the relative effectiveness scale (in a sufficiently large sample), however, requires that the organization also be most efficient (within the context of a variable-returns-to-scale world).

3.2 Operational Assessment of Relative Effectiveness

Based on the discussion above, the following six steps are required to evaluate the relative effectiveness (RE) of a non-profit organization against its peers:

STEP 1- Identification of PURs: A necessary first step in assessing the RE of an nonprofit organization is to validate the PURs that are initially hypothesized. This will generally require multi-equation regression (or canonical correlation analyses) in which the set of hypothesized PURs (Z) are related to the set of input resources actually used by the organizations ($X^R{}_A$) to generate their outputs (Y_A) (jncluding the buildup of financial reserves as needed).

STEP 2 - PUR Efficacy Conversion Rating: The efficacy with which the organization has been able to convert the set of statistically significant PURs into usable input resources must be assessed in the next step. Because we theorize the existence of variable returns to scale, the BCC (Banker, Charnes and Cooper 1984) version of the Data Envelopment Analysis model is suggested as being most appropriate. The BCC DEA model must be implemented, however, with due consideration for non-zero slacks. That is, Pareto-

Koopmans' efficiency rather than the Farrell efficiency measure is called for in this situation.(see Cooper, Seiford and Tone [2000, pp. 45] for a discussion of the differences between the two implementations of DEA efficiency). In this DEA model, the Zs are related to X^R_A. Since Pareto-Koopmans' Efficiency is being implemented, the usual dichotomy between input and output orientation in DEA analysis is immaterial. We denote the PUR efficacy rating of organization i as γ_i.[Z,X]

STEP 3 – Derivation of X_{MAX}: The PUR efficacy rating provides the basis to project the actual input resources currently used by the organization (X_A) into the maximum set of inputs that the organization could theoretically generate given its set of PURs (X_{MAX}). Given the PUR efficacy rating (γ_i), that the relationship can be presented as:

$$X_{MAX,i} = \gamma_i\,[Z,X_{A,i}] \qquad (1)$$

Step 4 – Efficiency Rating: In this step, the usual DEA efficiency measuring the rate at which the organizations converted their inputs (X_A) into outputs (Y_A) is derived. Since variable returns to scale is assumed, the analysis uses the BCC-DEA model and Pareto-Koopmans' Efficiency concept[2]. We denote the efficiency rating of organization i as Φ_i[X]

Thus,

$$Y_{A,i} = \Phi_i\,[X_{A,i}] \qquad (2)$$

Step 5 – Derivation of Y_{MAX}: In this fifth step, the X_{MAX} determined in Step 3 along with the input-output efficiency frontier defined in Step 4 are used to derive the set of maximum outputs that could have been generated by each organization from its X_{MAX}. Again, we assume the possible applicability of variable returns to scale, so the BCC DEA model and Pareto-Koopmans' efficiency are applied. The solution to this problem is:[3]

$$Y_{MAX,i} = \Phi_i\,[X_{MAX,i}] \qquad (3)$$

$$= \Phi_i\,[\gamma_i\,[Z,X_{A,i}]] \qquad (4)$$

Step 6 – Derivation of Relative Effectiveness (RE): The final step is the computation of RE and related concepts. RE is the projection of Y_A to the frontier defined by $Y_{MAX.}$ Thus, in vector notation,

$$RE_i = |Y_{A,i}| / |Y_{MAX,i}| \quad (5)$$

Operationally, if the existence of possible scale effects in projecting Y_{MAX} from X_{MAX} is ignored, Equation (5) can be reduced to:

$$RE_i = \Phi_i [X_{A,i}] * \gamma_i [Z, X_{A,i}] \quad (6)$$

Thus, the relative effectiveness of an organization can be measured approximately as the product of its relative efficiency and its relative efficacy in converting PURs into inputs for its production function.

3.4 *Estimation Approaches*

The estimation of the relationships modeled in this paper draws upon both the econometric and operations research approaches. As suggested by Cooper, Seiford and Tone (2000), regression/canonical correlation techniques can be used to determine if the theorized relationships are valid. Subsequently, DEA can be used to derive the relative effectiveness and efficiency estimates.

Regression techniques are needed to validate the potentially usable resources as indeed resources that can be converted into usable inputs by an organization. This issue is particularly important for nonprofit institutions because the PURs are frequently in forms that are likely to be very different from the actual inputs used. For example, for public agencies the degree of political support from key stakeholders may constitute a very important PUR since that may be converted to the level of funding obtainable from a legislature. Similarly, for private institutions of higher education, key PURs may be factors like the size of the alumni body and its relative wealth, the location of the institution, and the past prestige of the institution. Current inputs, however, may be factors like the total operating costs, size and condition of the physical facilities, and the drawing power of the institution in terms of the entry level Scholastic Aptitude Test (SAT) scores.

Regression techniques are again needed to establish the linkage between the theorized inputs and outputs of the organization. This is also a critical step in the analysis of the performance of non-profit institutions because the definitions of inputs and outputs may be subject to differences of opinion, and they may be conditioning variables that need to be considered. A good example of this conditioning through regression methods can be found in Arnold, Bardhan, and Cooper (1997) where it was found through regression that three input variables had negative effects on an outcome variable, and thus had to be treated differently from the other inputs.

IV. Empirical Illustration

4.1 Data Source and Variables Used

To demonstrate the feasibility of the relative effectiveness measurement approach, the higher-education institutional setting was selected. This selection was motivated by two factors. First, institutions of higher education (IHEs) have sufficient similarities in mission and general objectives (when grouped by the classification criterion advanced by the Carnegie Foundation for the Advancement of Teaching [1983]) to meet the requirement for an adequate sample size. Second, IHEs have been frequently used in the application of DEA techniques, so the input-output relationships required for estimating relative efficiency have been well specified for some time. Thus, this study extends the literature by introducing relative effectiveness as an additional dimension of performance evaluation. The study also provides some degree of continuity in the literature by focusing attention on the notion of potentially utilizable resources, and their role in evaluating institutional performance.[4]

The sample chosen for this study comprises 107 private colleges and universities selected from the 1998 *US News and World Report College Survey.* Based on the previous literature (Ahn, Arnold, Charnes and Cooper,1989), three variables were chosen to measure the inputs of IHEs: Total Operating Expenses (*TOPEXP*); Total Plant, Property and Equipment at book value (*PPE*), and the Average SAT score of entering first-year undergraduate students (*SAT*). *TOPEXP* and *PPE* were gathered from the financial statements of the IHEs for 1997, while *SAT* was obtained from *US News and World Report College Survey.* These variables constitute the set of inputs we referred to earlier as the cost-based inputs (X^C_A).

For the current output measures (Y^C_A), five variables were initially selected based on the prior literature. These variables were the number of full-time equivalent graduate students (*GRAD*), the number of full-time equivalent undergraduate students (*UGRD*), the amount of external research grants obtained by the faculty as reported in the IHE's 1997 annual report (*RESC*), the average graduation rate of undergraduate students (*GRATE*), the academic reputation rating of the institution (*ACREP*), and the alumni giving rate (*ALMGR*). *GRAD* and *UGRD* capture the volume of the IHE's output, while *GRATE* captures the effect of inter-institutional differences in the throughput rate. *ALMGR* is used as a proxy for the level of alumni satisfaction with the institution (reflecting both the level of identification and interest in the furtherance of the institution's objectives). *RESC* proxies for the level of faculty research output, and *ACREP* reflects the reputation that the institution has garnered among its peers for the quality of its academic activities. Except for *RESC*, all the other variables were obtained from *the US News and World Report College Survey*. Because *ACREP* scores are reported by *US News and World Report* by college classification, we introduced the four classes under which the ratings are classified: national university (*NU*), national liberal arts college (*NLA*), regional university (*RU*), and regional liberal arts college (*RLA*).

Six variables were initially identified as suitable PURs (Z) for the IHE setting. The age of the institution (*AGE*) was theorized to be a potentially utilizable resource because of the belief that, the older the institution is, the greater the opportunity the institution has had to build links in the community, build up a reputation, and establish a strong alumni network. Thus, *AGE* was theorized to be positively associated with all the INPUTS.

The LOCATION (by region) of the institution was theorized to be another PUR, with institutions in the *EAST* expected to have higher advantage because of the higher population density than those in the *WEST*, *MIDW* (Midwest), and *SOUTH*. Furthermore, LOCATION SETTING was also theorized to be another PUR. Institutions in *RURAL* settings were theorized to have a competitive advantage over those in *URBAN* or *SURB (suburban)* locations because of the lower density of the immediate population that they can draw upon. This potential disadvantage is offset by the lower cost of living in rural areas, and hence, lower operating costs. IHEs in *SURB* locations were theorized to have an advantage over *URBAN* locations in attracting the best students because of the greater attractiveness of suburban locations, although the lower population density may offset this advantage.

The other three PURs are the estimated maximum number of living alumni (*MAXALM*), the level of endowment from the previous year (*EDW*), and the previously-established level of demand for the institution's services (*DEMD). MAXALM* was estimated as the total number of students enrolled in the institution multiplied by the lesser of 30 or the age of the institution. In general, the higher MAXALM, the higher the resources that can potentially be generated. *EDW* reflects the resources carried forward from previous periods that can be used to support the demand for services. *DEMD* is computed as the complement of the student selectivity ratio (i.e., the number of students rejected to the total number of applications received). It therefore reflects the degree to which institutional reputation for quality of services relative to cost has already been established from past history.

For the purpose of relating the PURs to the inputs, *TOPEXP* was replaced by total revenue (*TREV*). The rationale is that, for measuring efficacy ratings, the focus has be to shifted to the ability of the institution to raise revenues, and using *TREV* instead of *TOTEXP* accomplishes this obejctive. Thus, the X^R_A variables consisted of *TREV*, *PPE* and *SAT*.

To formally demonstrate that the linkages linked theoretically are valid empirical, the following two sets of regression equations were estimated. The first set of equations were designed to determine which hypothesized PURs were related to the three revenue-based inputs. These equations can be written as:

$$\log TREV_t = \alpha_0 + \alpha_1 MIDW + \alpha_2 WEST + \alpha_3 SOUTH + \gamma_1 RURAL + \gamma_2 SUBRB + \lambda_1 \log AGE_{t-1} + \lambda_2 \log MAXALM_{t-1} + \lambda_3 \log EDW_{t-1} + \lambda_4 DEMD_{t-1} \quad (7)$$

$$\log PPE_t = \alpha_0 + \alpha_1 MIDW + \alpha_2 WEST + \alpha_3 SOUTH + \gamma_1 RURAL + \gamma_2 SUBRB + \lambda_1 \log AGE_{t-1} + \lambda_2 \log MAXALM_{t-1} + \lambda_3 \log EDW_{t-1} + \lambda_4 DEMD_{t-1} \quad (8)$$

$$\log SAT_t = \alpha_0 + \alpha_1 MIDW + \alpha_2 WEST + \alpha_3 SOUTH + \gamma_1 RURAL + \gamma_2 SUBRB + \lambda_1 \log AGE_{t-1} + \lambda_2 \log MAXALM_{t-1} + \lambda_3 \log EDW_{t-1} + \lambda_4 DEMD_{t-1} \quad (9)$$

Based on the previous discussion, all the λ parameters in Equations (7) to (9) are expected to be positively signed. The signs of the α (except the intercept) are expected to be negative, reflecting the expectation that, compared to the institutions in the densely-populated East (which represents the base), the institutions in the other regions will be slightly disadvantaged. The signs of the γ coefficients, however, could be either positive or negative (with URBAN as the base), depending on the relative advantages enjoyed by the urban location vis-a-vis rural and suburban locations.

To determine if the cost-based inputs (X^C_A) and current outputs (Y^C_A) selected are related to each other in the directions expected, the following regression equations were estimated:

$$\log TOPEXP_t = \alpha_0 + \alpha_1 NLA + \alpha_2 RU + \alpha_3 RLA + \beta_1 \log GRAD_t + \beta_2 \log UNDG_t + \beta_3 GRATE_t + \beta_4 \log RESC_t + \beta_5 ACREP_t + \beta_6 ALMGR_t \quad (10)$$

$$\log PPE_t = \alpha_0 + \alpha_1 NLA + \alpha_2 RU + \alpha_3 RLA + \beta_1 \log GRAD_t + \beta_2 \log UNDG_t + \beta_3 GRATE_t + \beta_4 \log RESC_t + \beta_5 ACREP_t + \beta_6 ALMGR_t \quad (11)$$

$$\log SAT_t = \alpha_0 + \alpha_1 NLA + \alpha_2 RU + \alpha_3 RLA + \beta_1 \log GRAD_t + \beta_2 \log UNDG_t + \beta_3 GRATE_t + \beta_4 \log RESC_t + \beta_5 ACREP_t + \beta_6 ALMGR_t \quad (12)$$

Based on the previous discussion, the signs of all the β parameters in Equations (10) to (12) are expected to be positive, reflecting the expectation that the generation of these outputs consume resources, on average. The signs of the α parameters are negative, since with NU as the base, one would expect the other types of institutions to have lower costs, given the lack of the heavy administrative overhead costs associated with comprehensive national universities.

4.2 Results of Validation of Variables

To facilitate future referencing, the list of variables used in the empirical part of the study are presented in Table 1. Summary statistics on the distribution of the sample by Carnegie Foundation classification and geographic location and setting are given in Table 2. In general, the sample is widely dispersed over the spectrum of Carnegie Foundation classifications, with 32 NLA colleges, 31 RLA colleges, 28 NU institutions, and 18 RU institutions. To reduce the skewness in the distribution of some of the key variables and also to avoid heteroscedasticity in the estimation of Equations (7) to (12), natural log transformations were applied ten of the 14 continuous variables. The summary statistics reported are based on the transformed variables.

****

Insert Tables 1 and 2 here

****

To enable the log transformation for RESC and GRAD, we added unity to all the scores for these two variables prior to the log transformation. Since the objective of this empirical part of the paper is merely to demonstrate the feasibility of the relative effectiveness techniques advocated here, this transformation is presumed to be innocuous in its effect. For the same reasons, the log-transformed variables were used in both the regression analyses and the subsequent DEA applications.

****

Insert Table 3 here

****

Table 3 presents the results of estimating Equations (7) to (9). This is an attempt to determine if the PURs hypothesized are, in fact, related to the level of inputs. Examining first the results for *TREV* in Panel A of Table 3, the results show that *MIDW* has a positive coefficient relative to the *EAST*. This implies a relative revenue advantage to location in the Midwest relative to the East. The other two geographic dummy variables are not significant, so location in the West or South is not a relevant factor. Furthermore, none of the Location Setting dummy variables are significant, indicating no effect of the location setting of the institution. The coefficient for *AGE* is not statistically significant, while those for *EDW* , *DEMD* and *MAXALM* are all positive and significant (as expected).

Moving on to Panels B and C of Table 3, similar results are observed. With *PPE* as the dependent variable (Panel B), *MAXALM*, *EDW* and *DEMD* all have positive coefficients, as expected. However, *AGE* is not statistically significant here as well. With *SAT* as the dependent variable (Panel C), *AGE*, *EDW* and *DEMD* all have positive and statistically significant coefficients, consistent with prior expectations. *MAXALM*, however, has a statistically significant negative coefficient. Since these are mere associational tests (rather than cause-and-effect relationships), the negative sign may indicate that high SAT entry-score standards tends to contract the pool of alumni of an institution. In other words, for the purpose of increasing the entry level SAT scores, *MAXALM* cannot contribute directly.

Overall, the primary conclusion from these analyses is that the suitable PURs for inclusion in the next-stage DEA analyses are the dummy variables for Regional Location, *AGE*, *MAXALM*, *EDW* and *DEMD*. The negative sign for *MAXALM* in the regression with *SAT* as the dependent variable suggests the need for sensitivity analyses to see how dropping *MAXALM* in the second-stage DEA analyses might affect the relative effectiveness scores. The regional variables for Location Setting can be dropped from any further analyses since they are not significant in any of the regressions. *AGE*, however, is retained because it has the hypothesized effect on *SAT*.

****
Insert Table 4 here
****

Table 4 presents the results of estimating Equations (10) to (12). Panel A of Table 4 shows that, with *TOPEXP* as the dependent variable, all variables (except *ALMGR*) are significant and signed as expected. Relative to *NU*, the operating costs of *NLA*, *RU*, and *RLA* institutions are lower. All the hypothesized output factors (except *ALMGR*) have positive coefficients. Panels B and C presents the results with *PPE* and *SAT* as the dependent variables. In both panels, all the continuous output variables except *RESC* and *ALMGR* are positive and statistically significant. Since *ALMGR* is not statistically significant in any of the regression equations, it is reasonable to conclude that it is not a valid output characteristic and should be excluded.

4.3 Estimating the Relative Effectiveness of the Colleges

The DEA parts of the six-step process described earlier were subsequently applied to the data. Because of possible non-homogeneity, the analysis was carried out by type of college, using the Carnegie Endowment classifications (*NU*, *NLA*, *RU*, and *RLA)*.

In the first DEA step, the efficacy measures (γ_i) were derived using as inputs *AGE*, *MAXALM*, *EDW* and *DEMD*, and associated outputs of *TREV*, *PPE*, *and SAT*. The version of the BCC model implemented was the output-oriented superefficiency DEA model proposed by Tone (2002). The resulting superefficiency scores were converted to a range of zero to unity by scaling by the highest observed score.

The second DEA step involved computing the efficiency measures (Φ_i) using as inputs *TOTEXP*, *PPE* and *SAT* and as outputs *ACREP*, *GRATE*, and *RESC*. The output-oriented superefficiency BCC version proposed by Tone (2002) was again used, and the resulting scores also rescaled to range between zero and unity. Finally, Equation (7) was applied to calculate the relative effectiveness scores for each observation.

****
Table 5 here
****

Table 5 presents the means, standard deviations, and correlations between the PUR efficacy measures, the technical efficiency measures, and the relative effectiveness measures. For all four types of colleges, the relative effectiveness measures are lower than the technical efficiency measures, although the product-moment correlation between them is pretty high (ranging from 0.50 to 0.832). The correlation between the efficacy measures and the efficiency measures are either not significantly different from zero or strongly negative in all four groups. Thus, an insight gained from this analysis is that the private IHE institutions that were most effective in exploiting their potentially utilizable resources were not also the most efficient in using the resources thus obtained. The range of the relative effectiveness scores from 0.41 to 0.64 supports the inference that, on average, there is significant room for achieving more effectiveness within all four types of private colleges.

5. Conclusions

The objective of this study was to develop an internally consistent quantitative measure of the effectiveness that non-profit institutions can use to evaluate their performance. The approach we have presented achieves this objective by demonstrating that while economic efficiency can be considered to be a contributory factor in organizational effectiveness, it is not sufficient. We have shown that an additional measure (an efficacy measure) is needed as well.

The feasiblity of the concepts has been demonstrated by their application to the non-profit private colleges. Our empirical results demonstrate that, while relative effectiveness is correlated with efficiency, efficacy is negatively correlated with efficiency in this context. Thus, from a managerial standpoint, the results show that truly effective nonprofit institutions must excel at both efficacy and efficiency relative to their peers. Excelling on only one dimension is not sufficient to achieve true organizations effectiveness.

REFERENCES

Ahn, T.S., V. Arnold, A. Charnes, and W.W. Cooper. 1989. DEA and ratio efficiency analyses for public institutions of higher learning in Texas. In: J.L. Chan and J.M. Patton., eds. *Research in Governmental and Nonprofit Accounting,* 5: 165-185.

Ahn, T.S., A. Charnes, and W.W. Cooper. 1988. Some statistical and DEA svaluations of relative efficiencies of public and private institutions of higher learning. *Socio-Economic Planning Sciences* 22 (6): 259-269.

Arnold, V.L., I.R. Bardhan, and W.W. Cooper. 1997. A two-stage DEA approach for identifying and rewarding efficiency in Texas secondary schools. in W.W. Cooper, S. Thore, D. Gibson and F. Phillips (Eds.), *IMPACT: How IC^2 Research Affects Public Policy and Business Practices*, Quorum Books, Westport, Connecticut: 131-166.

Banker, R.D. 1984. Estimating most productive scale size using sData Envelopment Analysis. *European Journal of Operational Research* 17: 35-44.

Banker, R.D., A. Charnes, and W.W. Cooper. 1984. Models for the estimation of technical and scale inefficiencies in data envelopment analysis. *Management Science* 30 (9): 1078-1092.

Campbell, J.P. 1977. On the nature of effectiveness. In P.S. Goodman and J.M. Pennings (Eds.), *New Perspectives on Organizational Effectiveness*, Jossey-Bass, San Francisco, CA: 13-55.

Cameron, K. 1986. A study of organizational effectiveness and its predictors. *Management Science* 32(1): 87-112.

Charnes, A. and W.W. Cooper. 1957. On the theory and computation of delegation-type efficiency: K-efficiency, functional efficiency and goals," *Proceedings of the*

Sixth International Meeting of the Institute of Management Science, Pergamon Press, London.

Charnes, A. and W.W. Cooper, 1961. *Management Models and Industrial Applications of Linear Programming*, John Wiley & Sons, Inc., New York.

Charnes, A. and W.W. Cooper, and E. Rhodes. 1978. Measuring the efficiency of decision-making units. *European Journal of Operational Research* 2: 429-444.

Charnes, A. and W.W. Cooper, and E. Rhodes. 1981. Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow-through. *Management Science* 27: 668-697.

Cooper, W.W., L.M. Seiford, and K. Tone. 2000. *Data Envelopment Analysis: A Comprehensive Text With Models, Applications, References and DEA-Solver Software,* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Connolly, T., E.J. Conlon and S.J. Deutsch. 1980. Organization effectiveness: A multiple-constituency approach," *Academy of Management Review* 5(2): 211-217.

Debreu, G. 1951. The coefficient of resource utilization," *Econometrica* 19 (3), 273-292.

Farrell, M.J. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A*, 120 (3): 253-290.

Finks, S.L., R.S. Jenks, and R.D. Willets. 1983. *Designing and Managing Organizations*. Irwin, Homewood, Illinois.

Goodman, P.S., R.S. Atkin and F.D. Schoorman. 1983. "On the demise of organizational effectiveness studies" in K.S. Cameron and D.A. Whetten (Eds.), *Organizational Effectiveness: A Comparison of Multiple Models,* Academic Press, New York: 187-203.

Governmental Accounting Standards Board (GASB. 1990., *Service Efforts and Accomplishments: Its Time Has Come*, H.P.Hatry, J.M. Sullivan, J.M. Mountain and L. Kremer, (Eds.) GASB Research Report, GASB, Norwell, CT.

Hannan, M.T., and J.H. Freeman. 1977. The population ecology of organizations*.* *American Journal of Sociology,* 82: 929-964.

Herman, R.D. and D.O. Renz. 1988. Nonprofit organizational effectiveness: Contrasts between especially effective and less effective organizations*. Nonprofit Management & Leadership* 9(1): 23-38.

Joys, D.J. 2001. The effects of employee satisfaction, organizational citizenship behavior, and turnover on organizational effectiveness: A unit-level, longitudinal study," *Personnel Psychology* 54: 101-114.

Koopmans, T.C., 1951. *Activity Analysis of Production and Allocation*, John Wiley & Sons, New York.

Lewin, A.Y., and J.W. Minton. 1986. Determining organizational effectiveness: Another look and agenda for research," *Management Science,* 32(5): 514-539.

Pfeffer, J., and G.R. Salancik. 1978. *The External Control of Organizations*, Harper and Row, New York.

Quinn, R.E. and Rohrbaugh. 1983. A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis." *Management Science* 29 (3): 363-377.

Rainey, H.G., and P. Steinbauer.1999. Galloping elephants: Developing elements of a theory of effective government organizations," *Journal of Public Administration Research and Theory* 9 (1): 1-32.

Reed, R. 1991. Bimodality in diversification: An efficiency and effectiveness rationale," *Managerial and Decision Economics* 12 (1): 57-66.

Rojas, R.R. 2000. A review of models for measuring organizational effectiveness among for-profit and nonprofit organizations. *Nonprofit Management & Leadership* 11 (1): 97-104.

Scott, W. R. 1977. Effectiveness of organizational effectiveness studies. In P.S. Goodman and J.M. Pennings (Eds.), *New Perspectives on Organizational Effectiveness*, Jossey-Bass, San Francisco, CA: 63-95.

Tone, K. 2001. A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* 130: 498-509.

Van der Heever, P.K. and L.D. Coetsee. 1998. An expert system-based audit instrument to assess organizational effectiveness. *Organization Development Journal* 16 (1): 29-44.

Yutchman, E. and S. E. Seashore. 1967. A system resource approach to organizational effectiveness. *American Sociological Review* 32: 891-903.

Zammuto, R. F. 1982. *Assessing Organizational Effectiveness*, State University of New York Press, Albany, NY.

EXHIBIT I

Diagrammatic Representation of Relative Effectiveness versus Relative Efficiency

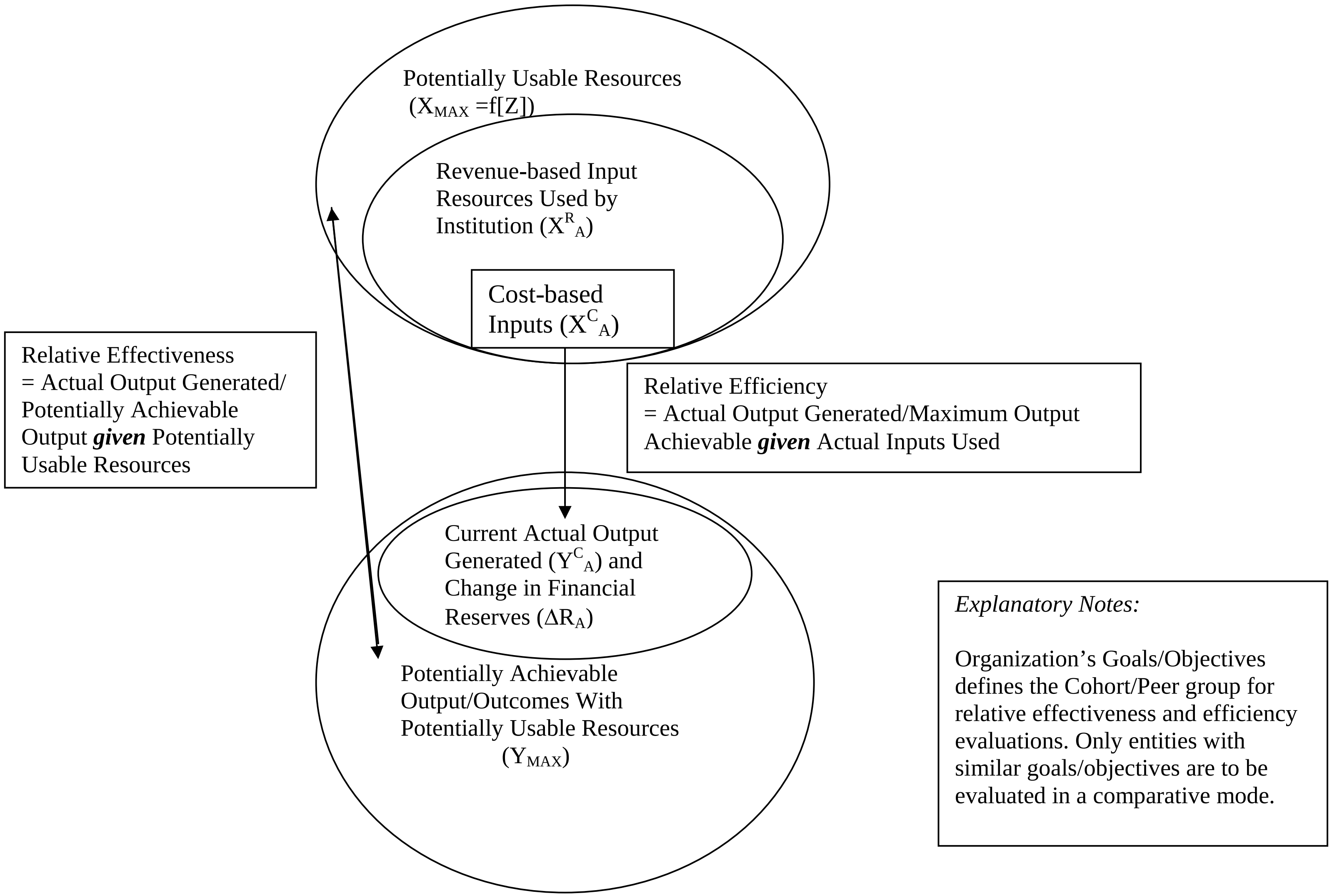
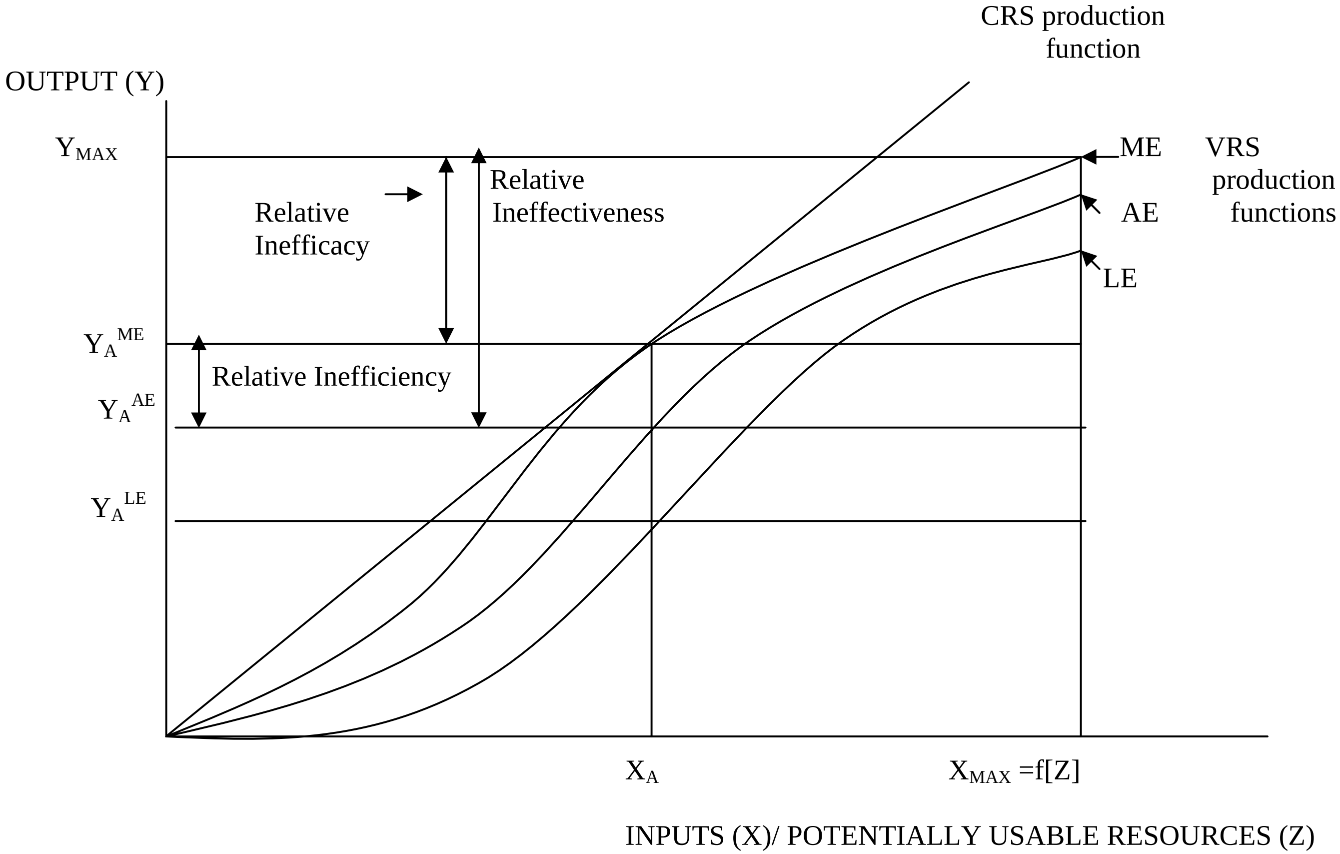


EXHIBIT II

Graphical Representation of the Relationship Between Relative Effectiveness and Relative Efficiency
(Graph of Production Functions)



Explanatory Notes:
The superscripts ME, AE and LE represent Most Efficient, Average Efficiency and Least Efficient, respectively. X stands for Inputs, Z for Potentially Usable Resources, and Y for Outputs. The subscripts A and MAX represent Actual used (for inputs) or generated (for outputs) and Maximum Feasible, respectively.

Table 1

List of variables Used in the Empirical part of study.

EAST	= Dummy variable for institutions located in the NorthEast
MIDW	= Dummy variable for institutions located in the Midwest
WEST	= Dummy variable for institutions located in the West
SOUTH	= Dummy variable for institutions located in the South
RURAL	= Dummy variable for institutions located in a rural setting
SUBRB	= Dummy variable for institutions located in a suburban setting
URBAN	= Dummy variable for institutions located in an urban setting
NU	= Dummy variable for institutions classified as National University
NLA	= Dummy variable for institutions classified as National Liberal Arts College
RU	= Dummy variable for institutions classified as Regional University
RLA	= Dummy variable for institutions classified as Regional Liberal Arts College
Log AGE	= Natural log of the age of the institution
Log MAXALM	= Natural log of the estimated maximum size of alumni body
Log EDW96	= Natural log of the total endowment of the institution in 1996
DEMD96	= Natural log of the total demand for the institution in 1996 estimated as the ratio of students rejected to total applications.
Log RVPS97	= Natural log of total revenue per student enrolled in 1997
Log TEXP97	= Natural log of total operating expenses in 1997
Log PPE97	= Natural log of total property, plant and equipment in 1997
Log SAT97	= Natural log of average SAT score of entering undergraduate students in 1997
Log GRAD97	= Natural log of total graduate students enrolled in 1997
Log UNDG97	= Natural log of total undergraduate students enrolled in 1997
GRATE97	= Average graduation rate of undergraduate students in 1997
Log RESC97	= Natural log of total external research grants generated in 1997
ACREP97	= Academic reputation score from US News and World Report survey in 1997
ALMGR97	= Alumni giving rate (as proxy for alumni satisfaction) in 1997

Table 2
Summary Statistics on Variables in Study

Variable	Sample size	Mean	Standard Deviation	Minimum	Maximum
PURs					
Log AGE	109	4.78	0.46	2.89	5.70
Log MAXALM	109	11.01	0.97	8.99	13.26
Log EDW96	109	11.73	1.66	6.85	15.69
DEMD96	109	0.33	0.22	0.00	0.87
Inputs					
Log TREV97	109	11.25	1.34	0.05	14.60
Log TEXP97	109	11.16	1.26	9.04	14.17
Log PPE97	109	11.06	1.33	8.03	14.43
Log SAT97	109	7.01	0.16	6.60	7.29
Outputs/Outcomes					
Log GRAD97	109	4.96	3.29	0.00	9.67
Log UNDG97	109	7.67	0.78	5.70	9.74
GRATE97	109	0.68	0.17	0.27	0.96
Log RESC97	109	6.73	3.69	0.00	13.52
ACREP97	109	2.97	0.62	1.26	4.00
ALMGR97	109	0.29	0.15	0.02	0.68

Regional Location	*EAST*	*MIDW*	*WEST*	*SOUTH*
Total sample size	63	14	15	16
Location Setting	*RURAL*	*SUBRB*	*URBAN*	
Total sample size	22	37	50	
Class	*NU*	*NLA*	*RU*	*RLA*
Total sample size	28	32	18	31

Table 3
Multiple Regression Analyses of PURs (Independent Variables) and INPUTS (Dependent Variable) As Test of Validity of PURs as Determinants of Levels of Inputs Available

Independent Variables	Expected sign	*log TREV97* Coefficient	t-value	*log PPE97* Coefficient	t-value	*log SAT97* Coefficient	t-value
Intercept		1.862	-3.09 ***	-1.482	-2.46 *	6.378	44.65 ***
Regional Location							
EAST (base)							
MIDW		0.265	2.34 ***	0.109	0.96	-0.113	-4.21 ***
WEST		0.094	0.88	-0.062	-0.57	-0.005	-0.18
SOUTH		0.160	1.56	0.011	0.1	-0.058	-2.38 *
Location Setting							
RURAL		0.033	0.32	-0.001	-0.01	-0.025	-1.04
SUBRB		0.041	0.50	0.081	0.97	0.028	1.39
URBAN (base)							
PURs							
Log AGE	+	0.002	-0.20	-0.011	-0.11	0.048	2.13 *
Log MAXALM	+	0.242	15.19 ***	0.643	13.15 ***	-0.033	-2.83 **
Log EDW96	+	0.0391	11.18 ***	0.454	12.96 ***	0.062	7.52 ***
DEMD96	+	0.881	3.97 ***	0.477	2.31 *	0.161	3.28 **
Adjusted R-squared			0.932		0.931		0.716
F-ratio/ Significance level			165.55 ***		161.66 ***		31.31 ***

Table 4
Regression Analyses of OUTPUTS (Independent Variables) and INPUTS (Dependent Variable)

As Test of Validity of OUTPUTS as Determinants of Levels of Inputs Used

Independent Variables	Expected sign	*log TEXP97* Coefficient	t-value	*log PPE97* Coefficient	t-value	*log SAT97* Coefficient	t-value
Intercept		4.625	8.94 ***	4.207	5.48 ***	6.755	48.69 ***
CLASS Dummy Variables							
NLA		-0.817	-6.17 ***	-0.716	-3.65 ***	-0.007	-0.18
RU		-1.100	-9.86 ***	-1.102	-6.66 ***	-0.091	-3.03 **
RLA		-1.106	-7.95 ***	-1.129	-5.47 ***	-0.126	-3.37 ***
OUTPUT/OUTCOMES							
Log GRAD97	+	0.046	2.95 **	0.063	2.76 **	0.004	0.93
Log UNDG97	+	0.634	8.88 ***	0.645	6.09 ***	-0.032	-1.69
GRATE97	+	1.049	3.5 ***	1.141	2.57 *	0.272	3.39 ***
Log RESC97	+	0.021	1.87 *	0.002	0.14	0.003	1.1
ACREP97	+	0.458	5.61 ***	0.475	3.92 ***	0.113	5.14 ***
ALMGR97	+	-0.128	-0.34	0.343	0.62	-0.011	-0.11
Adjusted R-squared			0.943		0.886		0.729
F-ratio/ Significance level			198.37 ***		94.09 ***		33.3 ***

Table 5
The means and correlations of the various measures in the four college groups

Panel A: Mean (Standard Deviation) of Measures

Group (N)	Efficacy	Efficiency	Effectiveness
NU (28)	0.76(0.13)	0.84 (0.11)	0.64 (0.13)
RU (18)	0.60 (0.13)	0.71 (0.19)	0.41 (0.11)
NLA (32)	0.67 (0.16)	0.67 (0.27)	0.45 (0.21)
RLA (31)	0.62 (0.17	0.73 (0.28)	0.45 (0.21)

Panel B: Product-Moment Correlation Coefficients

	Efficacy vs. Efficiency	Efficacy vs Effectiveness	Efficiency vs. Effectiveness
NU	(0.18)	0.76	0.50
RU	(0.569)	(0.065)	0.832
NLA	0.086	0.611	0.825
RLA	(0.221)	0.407	0.774

ENDNOTES

[1] The highly condensed summary of the effectiveness literature here is based on Cameron (1986) and Rojas (2000). More up-to-date and complete summaries are available in Rojas (2000) and Van der Heever and Coetsee (1998).

[2] The CRS version of the DEA model (see Charnes, Cooper and Rhodes [1978]) can be implemented as a supplementary step to evaluate the sensitivity of the results to the VRS assumption

[3] The relationship between $\Phi_{MAX,i}$ and Φ_{ι}.is difficult to define on an ex ante basis. Under constant returns to scale, the two will be equal, assuming that the movement from X_A to X_{MAX} involved a radial expansion of inputs. Alternative production arrangements may lead to slighly different efficiency outcomes. The use of Φ_{ι} as an approximation for $\Phi_{MAX,I}$ in empirical work is therefore subject to this explicit limitation.

[4] Some of the extensive literature on the application of DEA analyses to IHEs can be found in Ahn, Charnes and Cooper (1988)., and Ahn, Arnold, Charnes and Cooper (1989).