EVALUATING SCHOOL PERFORMANCE:

ARE WE READY FOR PRIME TIME?

Robert Bifulco
William Duncombe

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020, USA
Email: rbifulco@maxwell.syr.edu
Duncombe@maxwell.syr.edu

Prepared for publication in William J. Fowler (ed.), *Developments in School Finance, 1999-2000*, a report published by the National Center for Education Statistics.

Robert Bifulco is a Ph.D. candidate in Public Administration at the Maxwell School of Citizenship and Public Affairs, Syracuse University and a Research Associate at the Center for Policy Research. He has also worked for the New York State Education Department where he was involved in programs to identify and support low-performing schools.

William Duncombe is Associate Professor of Public Administration at the Maxwell School of Citizenship and Public Affairs, Syracuse University and a Senior Research Associate at the Center for Policy Research. He has published numerous articles on school finance, the estimation of educational costs and the evaluation of educational programs.

ABSTRACT

Several states and districts have begun to implement programs designed to reward excellent performance by schools. Programs have also been developed to identify poor performing schools in which states may intervene by mandating changes, replacing staff, or taking over the school. This emphasis on school performance accountability begs the question: can we accurately define and measure school performance? One approach to developing school performance measures is to apply econometric and linear programming techniques that have been developed to measure productive efficiency. This study uses simulated data to assess the adequacy of several of these methods for the purposes of performance-based school reform. Our results suggest that with the complex data sets and technologies typical of education contexts, the most frequently used methods do not provide adequate measures of efficiency.

1. Introduction

Performance-based school reform has received much attention in recent years. Key elements of this reform movement include setting standards of student, teacher and school performance, granting autonomy to local actors in the educational process, and establishing rewards for high performance and remedies for low performance. These elements are prominently featured in the 1994 reauthorization of the federal Title I program as well as several state level reform initiatives.[1]

These reforms have been advanced as a remedy for several perceived problems with existing public education systems. Prominent among these perceived problems are a lack of the incentives and knowledge needed to improve student performance. Some have argued that given current systems for determining compensation, professional advancement and school funding, the incentives of school officials are insufficiently linked to student performance (Hanushek, 1994; Levin, 1997). Performance-based school reform attempts to provide stronger incentives for improving student performance by developing measures of achievement and tying financial and other rewards to those measures. Some also believe that we know very little about how to manage classrooms, schools and districts in ways that consistently result in higher levels of student achievement. By granting local actors the autonomy to experiment with new approaches and providing the means to assess the impact of local experiments on student performance, performance-based school reform is seen as a way to learn how to meet the ever-increasing demands placed on our public education systems (Hanushek, 1994).

Developing valid and reliable measures of school performance is crucial both for efforts to establish incentives and to assess management practices. There is a growing consensus that measures of school performance should be based on the performance of students in the school.

However, there is also recognition that any measure of school performance that is based on the performance of students needs to account for the differences in resources available to and service delivery environments faced by different schools. One approach to developing such measures is to apply the concept of productive efficiency, and techniques for measuring it, that have been developed in the fields of economics and operations research.

Several such techniques have been developed, and several have been applied to estimate the efficiency of educational organizations. These include econometric approaches that utilize ordinary least squares regression and stochastic frontier estimation as well as a group of linear programming approaches falling under the rubric of Data Envelopment Analysis (DEA).[2] The availability of these methods for estimating school efficiency raises two questions. The first is whether or not the methods provide sufficiently accurate estimates of efficiency. The second question is, which method provides the most accurate estimates of efficiency and under what circumstances? Studies that have applied different methods to the same data have found that they provide different results (Banker, *et al*., 1985; Nelson and Waldman, 1986). The problem is that without knowing the true efficiency of the organizations studied, there is no way to determine which measures provide better estimates.

Studies that use simulated data with specified, and thus known, technological relationships and levels of efficiency can help to answer these questions. A limited number of such studies have been conducted, and recently some attempts have been made to use the results of such simulation studies to assess how appropriate existing efficiency measures are for the purposes of performance-based school reform. This paper reviews existing studies and provides new evidence from an analysis using simulated data.

Section 2 briefly describes the two general approaches to measuring productive efficiency used in the economic and operations research literature. Section 3 identifies the specific set of challenges that educational production processes pose for methods of estimating school efficiency. Section 4 reviews existing studies that have used simulated data to evaluate methods of estimating school performance. Section 5 identifies two different regression-based and four different DEA methods for measuring efficiency that we examine in a new simulation study. Section 6 describes how we simulated our data and Section 7 presents an analysis of how well each method did in estimating the known efficiencies of the simulated schools. Section 8 offers concluding remarks concerning the current state-of-the-art in measuring school performance and the implications this has for performance-based school reform efforts.

2. Two Approaches to Measuring Productive Efficiency

Technical efficiency is defined as a feasible combination of inputs and outputs such that it is impossible to increase any output (and/or reduce any input) without simultaneously reducing another output (and/or increasing any other input). In other words, for any given combination of school inputs, a technically efficient school could not produce more of any output (holding the other outputs constant). The curve in Figure 1 represents the combinations of inputs X1 and X2, which if used efficiently, will produce Y units of output. This curve constitutes the efficient production frontier. The combination of inputs used by school 'A' to produce Y units of output places it off the efficient frontier. School A could produce Y units of output with less of either or both inputs. A measure of school A's technical efficiency can be calculated by dividing the length of ray OB by the length of ray OA. This measure represents the proportional amount of each input used by school A required to produce the level of output that it is producing. The fact that this measure is less than one indicates that school A is inefficient.

The different methods for empirically estimating this measure of efficiency can be distinguished by the mathematical models used to estimate the efficient production frontier. Regression-based approaches begin by regressing an aggregate measure of output against a vector of inputs and a vector of environmental variables using Ordinary Least Squares. Next the estimated intercept term is “corrected” so that the estimated equation can be interpreted as a production frontier. The simplest of these methods, typically referred to as Corrected Ordinary Least Squares, increases the intercept term by the amount needed to make the largest residual zero. More complicated methods make use of assumptions about the probability distributions of inefficiency and random error to determine the intercept correction. Because the latter methods attempt to account for the affect of random, i.e. stochastic, factors on the observed relationships between inputs and outputs, they are said to estimate a stochastic production frontier.[3]

Implementing regression-based methods requires the assumption of an explicit functional form, explicit weights for each output and particular distributions of inefficiency and random error. The need to make assumptions that are difficult to verify is the primary disadvantage of these approaches to efficiency measurement. If the assumptions made are valid, however, the residual on the “corrected” regression equation for each school can be interpreted as a measure of inefficiency.[4]

Linear programming techniques for estimating production frontiers fall under the rubric of Data Envelopment Analysis (DEA). All DEA methods start with measures of a set of inputs and outputs for some sample of schools. They then use numerical methods to select, for each school, the set of input and output weightings that maximizes the ratio of weighted outputs to weighted inputs. This maximization problem is subject to the constraint that the weights selected for a given school, when applied to other schools in the sample do not result in one of the other schools having a ratio of weighted outputs to weighted inputs greater than one. This maximum ratio of weighted outputs to weighted inputs is the measure of efficiency. The optimization problem is run for each school separately. Thus, each school will have a different set of input and output weightings. In effect, DEA selects the set of weights that will give a particular school as high an efficiency score as possible, subject to the constraint that no other school would have an efficiency score greater than one given those weights.[5]

DEA does not require *a priori* specification of output weights. Rather the linear programming procedure uses the data to determine relative output weightings for each school individually. Nor does it require assumptions about the functional form of the production frontier or the distribution of inefficiency. Given the uncertainty surrounding these aspects of educational production these are potentially important advantages. The primary disadvantage of DEA is that it is deterministic. That is, it attributes all deviation from the production frontier to inefficiency, and provides no means of accounting for random error.

3. Complications Posed by the Educational Production Process

Analysis of educational production is notoriously difficult.[5] Here, I focus on aspects of education production that complicate the measurement of efficiency. The first difficulty is that education involves joint production of multiple outputs. Not only are schools charged with developing cognitive skills in several subject areas, but they are also charged with developing affective traits, promoting democratic values and furthering other social outcomes. Assumptions that these multiple outcomes are complimentary or even mutually consistent are difficult to maintain, and attempts to develop *a priori* weights that reflect the relative value of various outcomes are problematic. The fact that DEA does not require *a priori* specification of weights is typically touted as one of its primary advantages over regression-based approaches.

The second problem in analyzing educational production concerns the difficulty of measuring educational outputs. Standardized tests of cognitive skills are typically used to measure educational output. However, standardized tests are not always aligned with curricular goals, subjects such as science, social studies and the arts are not often tested, and even in tested subjects, higher order thinking and problem solving skills are often not assessed [Darling-Hammond, 1991]. Valid and reliable measures of affective traits, democratic values and social outcomes may be even more difficult to obtain. The presence of this type of measurement error can push a school off the production frontier even if it is truly efficient or make it appear efficient when it is not. In so far as DEA attributes all deviation from the production frontier to inefficiency, its estimates of efficiency will be distorted by measurement error.

Further complicating analysis is the fact that our knowledge about which factors affect educational outputs is inadequate. In addition, measuring factors that are known to effect educational outputs, such as student motivation or teacher quality, can be difficult. Consequently, attempts to analyze educational production suffer from the presence of unobserved inputs. Because input levels are typically correlated with each other as well as with environmental factors, the problem of unobserved variables can cause the statistical estimation of model coefficients to be biased. This is will cause regression-based techniques to misplace the production frontier, thereby biasing measures of efficiency. More generally, a school's distance from the production frontier will be determined by variation on the omitted inputs, as well as inefficiency. Failure to account for this other source of variation in a school's distance from the estimated production function will lead both DEA and regression-based approaches to provide biased estimates of efficiency.

Education production is also characterized by simultaneous relationships between inputs and outputs. In the case of certain student inputs that affect the learning process this is clear. Student motivation, for instance, both influences and is influenced by the level of educational output. Orme and Smith (1996) suggest that there may be feedback from outputs to institutional inputs as well. School districts in which test scores are low might come under pressure to promote improved performance, which might lead to increased resource provision and thus higher levels of inputs. To some extent this process is institutionalized in legislative programs. The federal Title I program, for instance, targets significant amounts of funds to schools with large numbers of students who show low levels of achievement. Such feedback is also likely to bias the estimation of regression coefficients, and Orme and Smith argue that it can bias DEA estimates of efficiency as well.

Finally, environmental factors, such as the family background of the students served by the school, can substantially influence the level of output that schools obtain. Environmental factors are conceptually different than production inputs because they are beyond the control of school officials. If environmental factors can be represented as simple additive terms in a school's production function, then it may be acceptable to treat them as another set of inputs. In this case, environmental factors might not significantly complicate the estimation of efficiency. If, however, these factors interact with controllable inputs and technologies in non-additive ways, then incorporating environmental factors into efficiency analysis will be complicated.

4. Existing Simulation Studies

There have been several studies of both regression-based and DEA methods of estimating productive efficiency.[7] Most of these are concerned with frontier and efficiency estimation generally, and do not specifically ask whether or not a given method provides measures of efficiency that are accurate enough for the purposes of performance-based school reform. Are the estimates of efficiency provided by existing methods accurate enough to serve as a basis for awarding financial incentives or targeting remedial efforts? Can these methods help us determine what managerial and resource allocation practices help to foster improved performance? Two recent studies have examined these questions and suggest that simple versions of regression and linear programming approaches are inadequate.

Brooks (2000) examines a regression-based approach for developing adjusted performance measures for schools that is similar to simple regression-based measures of efficiency used in the productivity literature. He focuses on the effect that correlation between efficiency (or in his terms "merit") and school inputs has on the accuracy of adjusted performance measures. Examining the case of one observable input and one output, he finds that an increase in the correlation between efficiency and the observed input of 0.10 decreases the rank correlation between the adjusted performance measure and the schools true merit by 0.065. He also finds that as the random error associated with the production of student performance increases, the adjusted performance measure becomes even more inaccurate. In cases where the correlation between the input and "merit" is high (above 0.50) and random error is relatively large, he finds that the rank correlation between the adjusted performance measure and true merit will most likely be statistically indistinguishable from zero. In such cases, adjusted performance measures are unlikely to provide useful rankings of schools by performance.

Given what we know about the education production process, we expect correlation between inputs and efficiency of the kind examined by Brooks. Important unobserved factors such as student and teacher motivation are determined simultaneously with school efficiency. The more efficient a school, the higher student performance and consequently the more motivated teachers and students are to work harder, thereby improving efficiency. Also, higher levels of observed inputs, such as more teachers allowing reduced class sizes, may increase teacher and student motivation. If this is true, we can expect efficiency to be correlated with observed inputs. Thus, Brooks' findings suggest that simple regression procedures for developing efficiency measures are probably not adequate for performance-based school reform.

Bifulco and Bretschneider (2001) examine Corrected Ordinary Least Squares (COLS) and the Charnes, Cooper and Rhodes (1978) formulation of DEA. They find that in cases with simultaneous relationships between inputs and outputs and with measurement error, the rank correlation between efficiency measures and true efficiency levels are no higher than 0.24. In these cases neither DEA nor COLS is able to place more than 31 percent of schools in the correct performance quintile. Schools assigned to the bottom 20 percent of performers are as likely to have actual performance levels above the median as in the bottom quintile. This confirms Brooks' suggestion that simple regression-based procedures provide inadequate measures of efficiency in educational contexts, and implies a similar conclusion for simple versions of DEA.

In the study presented below, we examine the accuracy of more sophisticated versions of regression-based and DEA methods for measuring efficiency. These methods are more complicated and thus more difficult to introduce into program and policy practice. However, they address some of the shortcomings of the simpler methods examined by Brooks (2000) and Bifulco and Bretschneider (2001). It is worth examining whether or not these improvements allow measures of school efficiency that are adequate for the purposes of performance-based accountability.

5. Methods Examined in this Study

We examine the performance of two regression-based methods of estimating efficiency and four versions of DEA. The first regression-based method we examine is Corrected Ordinary Least Squares (COLS) in which the intercept term from the ordinary least square regression is increased by the amount needed to make the largest residual zero. This is the method examined in Bifulco and Bretscneider (2001). The second regression-based method, referred to here as a stochastic frontier estimator (SFE), makes use of assumptions about the distributions of inefficiency and random error to determine the intercept correction. Details on this method are provided in Olsen *et al.* (1980).

The primary advantage typically advanced for regression-based approaches is their potential for addressing measurement error by treating efficient frontiers as stochastic phenomena. That is, these methods attempt to decompose the deviation of actual production from the estimated frontier into a component that is due to inefficiency and a component that is due to random error. The regression-based methods applied here, however, are not fully stochastic. COLS is entirely deterministic. In both adjusting the intercept and interpreting the residual from the adjusted regression equation, it is assumed that all deviation from the production frontier is due to inefficiency. The stochastic frontier estimation method examined here uses assumptions about the distribution of inefficiency and random error in determining how much to adjust the intercept of the regression equations. Once the intercept is adjusted, however, deviations from the frontier are assumed to either be due entirely to inefficiency or entirely to random error. If an observation is on the efficient side of the frontier, all of the deviation from the frontier is assumed to be due to random measurement error. If an observation is on the inefficient side of the frontier, all of the deviation is attributed to inefficiency.[8]

The first of four DEA methods examined, referred to here as DEA I, is the input minimizing formulation of the Charnes, Cooper and Rhodes (1978) version of DEA. Following the practice typical of early applications of DEA (Bessent, 1980; Bessent *et al.* 1982, 1983), only those factors over which school officials have control are included as inputs. This method, which is the one examined in Bifulco and Bretschneider (2001), has been criticized for ignoring the impact of environment on outputs (Ruggiero, 1996), a particularly important issue in education.

The other three DEA-based methods examined attempt to develop estimates of efficiency that control for the influence of environment on production outcomes. One of these methods, referred to in this paper as DEA II, attempts to control for environment by including it as an input in the standard DEA linear programming problem. Although this approach fails to recognize the important conceptual distinction between environmental factors and production inputs, it may provide a practical means of accounting for the environment in DEA estimates of efficiency.

Another approach is a two-stage method that uses regression in an attempt to separate those parts of the DEA estimates that are due to the effect of environment on output from those parts that are due to inefficiency. In the first stage of this method, DEA I is applied, using only discretionary inputs, to develop preliminary efficiency estimates. In the second stage, these preliminary efficiency estimates are regressed on environmental variables. An adjusted efficiency estimate is then computed for each observation by multiplying the coefficients from the second stage regression by the mean value of each environmental variable, and adding this to the observed regression residual. Methods similar to this have been applied to educational organizations by Ray (1991) and by Kirjavanien and Loikkanen (1998).[9] In this paper we referred to this method as "Two-Stage DEA".

The final method we examine is that developed by Ruggiero (1996). This method modifies the standard DEA programming problem to find the minimum level of inefficiency for a given school relative to other schools that face environments no better than the one it faces. Throughout this paper we refer to this as the Ruggiero approach.

6. Our Data Simulations

In order to examine the performance of these methods, we generated 12 different data sets that incorporate various aspects of education production discussed in Section 3. Each data set consists of 200 simulated observations. The relationships between inputs and outputs (i.e. the technologies) underlying each of these data sets are described here.

A Cobb-Douglas Technology with Additive Environment

The first sets of simulated data were generated from the following system of underlying technological relationships:

(1a) $y_1 = x_1^{0.25}\ x_2^{0.25}\ n^{0.50}\ e^{v1-u1}$

(1b) $y_2 = x_1^{0.20}\ x_2^{0.20}\ y_1^{0.20}\ n^{0.40}\ e^{v2-u2}$

The first output (y_1) is related by a Cobb-Douglas production function to two inputs (x_1 and x_2), an environmental factor (n), a random component representing measurement error (v1), and an inefficiency term (u1). The amount of the second output (y_2) is determined by the same two inputs, the same environmental factor, and the amount of the first input produced. The second equation also has a Cobb-Douglas form and includes its own measurement error (v2) and inefficiency terms (u2).

We randomly generated observations for each of the inputs from a uniform distribution on the interval (90, 110). We also randomly generated observations for the environmental variable from a uniform distribution on the interval (0, 200). Given the relative degree of standardization among schools with respect to input measures such as class size, and the large differences in environmental conditions, such as student poverty, faced by different schools, we might expect substantially more variation in the environmental variable than the input variables. The coefficients on both the input and environmental factors were chosen so that the effect of environment on output levels is large relative to the impact of the discretionary inputs. Again, this is intended as a rough approximation of real life educational production.

In each case, observations for the measurement error terms, v1 and v2, were generated from a normal distribution, $N(0,s_v^2)$. The value of s_v was varied to generate three different data sets. In the first data-set, s_v was set equal to zero to simulate a data-set with no measurement error. In the second and third cases, s_v was set equal to 0.1 and 0.3 to simulate cases with 'small' measurement error and 'large' measurement error respectively.

The inefficiency terms, u1 and u2, were each generated from a truncated normal distribution, $N(0,s_u^2)$. The distribution was truncated by setting all negative values equal to zero. For both u1 and u2, s_u, was set equal to 0.3. The overall level of efficiency for each observation was calculated as follows. First, the observed values of y_1 and y_2 were computed from equations (1a) and (1b) and the randomly generated values of x_1, x_2, n, v1, v2, u1 and u2. Then, efficient values of y1 and y2 were generated for each observation by setting u1 and u2 equal to zero. An efficiency value was then calculated as follows:

(2) $$\text{Efficiency} = [w_1(y1) + w_2(y2)] / [w_1(y1^*) + w_2(y2^*)]$$

Where y1 and y2 represent observed values, y1* and y2* represent efficient values, and w_1 and w_2 are weights that represent the relative importance of each output. In all cases, w1 and w2 are both set at 0.50. The mean efficiency values for the data sets with no, small and large measurement errors were 0.891, 0.892, and 0.897, respectively.

Cobb-Douglas Technology with Additive Environment and Endogeneity

The second set of samples was generated from the same technology as was just described with one exception. The observations for x_2 were replaced by observations linked to the inefficiency terms. Specifically, we used the following to generate observations of x_2:

(3) $$x_2 = 95 + 12u1 + 12u2 + e$$

Where e is a normally distributed variable, N (0, 9). This resulted in a distribution of x_2 similar to that generated in the above described data sets, and correlations between x_2 and the efficiency value ranging from –0.679 in the case with no measurement error to –0.671 in the case with large measurement error.

A negative correlation between inputs and efficiency values can be one of the by-products of the type of feedback from outputs to inputs discussed by Orme and Smith (1996). Correlation between the general composed error terms, u1+v1 and v2+u2, can also be the result of omitted variables. Thus, incorporating this correlation into the simulated data allows us to explore the impact of such feedback processes or omitted variables, i.e. endogeneity, on efficiency measurement.

Cobb-Douglas Technology with Interactive Environment

The following productive relationships were used to simulate the third group of data sets:

(4a) $$y_1 = x_1^{(0.25+0.25n)}\ x_2^{(0.25+0.25n)}\ e^{v1-u1}$$

(4b) $$y_2 = x_1^{(0.20+0.20n)}\ x_2^{(0.20+0.20n)}\ y_1^{0.20}\ e^{v2-u2}$$

This system of equations is similar to (1a) and (1b) except that the environmental term enters into the equation non-linearly. Here the environment affects the level of output by modifying the effect of each input. The same observations for x_1, x_2, v1, v2, u1, and u2 that were generated for the first set of samples were used for these data sets. Observations for n were randomly generated from a uniform distribution on the interval (0,1). As in the above cases, samples with no, small and large measurement error were simulated from this underlying technology. The final three samples generated incorporate endogeneity into the above technology. This is done in the same way as described above.

Taken collectively, the 12 data sets simulated for this study incorporate several important aspects of educational production such as multiple outputs, environmental factors, measurement error, and endogeneity. This will allow us to examine how these aspects of educational production affect our ability to measure school efficiency.

7. Results

Our discussion of the results is divided into four sections. First, we discuss the performance of the regression-based methods. Next, we discuss the performance of the DEA-based methods. Then, we compare the performance of the regression-based approaches with the linear programming methods. Finally, we present some analysis aimed at determining whether or not the most effective regression and DEA methods are adequate for the purposes of performance-based school reform initiatives.

The Performance of Regression-Based Methods

Regression-based methods require specification of assumptions concerning the functional form of technological relationships as well as the distribution of measurement error and inefficiency. In the multiple output case, assumptions also have to be made concerning the relative weighting of the various outputs. In applying COLS and the stochastic frontier estimator (SFE) to our simulated data we specify the same set of assumptions regardless of the data set being used. In some cases, these assumptions match the specifications of the true underlying technology, and in other cases they do not. Thus, we can see how misspecification affects the performance of regression-based procedures.

More specifically, in applying both COLS and SFE, we use a Cobb-Douglas functional form with the environment entering additively. In cases where this is in fact the functional form of the underlying technology, the regression models are well-specified. However, the data sets in which the environment enters interactively with the discretionary inputs, represent cases where the regression models are misspecified. In cases with endogeneity, a different type of misspecification is introduced.

The output weights used in forming the aggregate outcome measure used in the regressions were chosen to match those used to calculate the true efficiency value. In real situations these weights might in fact differ from school to school, and in any case, are difficult to specify. By matching the weights used in applying our estimators with those used in generating the true efficiency values, we ignore this potential difficulty in applying regression-based methods. This fact should be kept in mind when evaluating their performance.

Table 1 presents Kendall-Tau rank correlation coefficients between estimated efficiencies and true efficiency values. This measure captures the ability of each method to correctly rank observations. An important component of performance-based school reform is identification of the highest and the lowest performing schools in a jurisdiction. The highest performing schools can then be rewarded and corrective actions can be targeted to the lowest performing schools. Identifying groups of high and low performing schools can also be useful for determining whether certain management or resource allocation practices consistently lead to either higher or lower levels of performance. Thus, the ability of a method to correctly rank schools is an important criterion for assessing the usefulness of the methods for the purposes of performance-based school reform. A high rank correlation suggests that the measure performs well in identifying differential efficiency.

In all cases, the rank correlations for COLS and SFE are virtually identical. Both methods use the same OLS estimates to determine the production frontier slope parameters, and differ from each other only in the way they adjust the intercept to locate the production frontier. Thus, this finding is expected, and confirms similar findings by Ondrich and Ruggiero (1997). This finding suggests that although COLS might provide different cardinal measures of efficiency than SFE (and in fact does),[10] it is equivalent to SFE as an ordinal measure.[11] Thus, in situations where only the ordinal ranking of schools are required, the simpler COLS method might be preferable.

The Performance of DEA-Based Methods

As expected, DEA I provides poor estimates of efficiency in all cases. This method does not control for the influence of the environment on production outcomes and thereby confounds the affects of inefficiency with the affects of environment. Somewhat surprisingly, regressing the estimates from DEA I against the environmental variable in the "Two-Stage DEA" does not substantially improve the performance of DEA I on the rank correlation criteria. This result may be due to a misspecification of the second stage regression model. We used OLS to estimate a linear regression model. However, the distribution of DEA I efficiency estimates, which provide the dependent variable in the second stage regression, is truncated at 1. Thus, a Tobit model may be more appropriate.

In cases without measurement error, DEA II and the Ruggiero approach provide improved efficiency estimates. DEA II achieves improved estimates by including the environmental variable as an input in the standard DEA program. This improves matters only in cases where environment affects production in an additive fashion. The Ruggiero approach achieves improved estimates by modifying the DEA program so that each school is compared only to schools that face an environment no better than the one it faces. The Ruggiero approach achieves improved measures of efficiency when the environment affects production in an interactive way, as well as when environment enters additively. However, the performance of both DEA II and the Ruggiero approach is substantially undermined by the presence of measurement error.

Comparisons of Regression-Based Methods with DEA-Based Methods

The primary advantage typically touted for regression-based approaches is that they provide a means of handling measurement error. Regression-based approaches also provide well established means of controlling for the effect of environmental factors. However, we have seen that the ordinal measures of efficiency provided by COLS and SFE are equivalent. This raises doubts about the ability of SFE to separate measurement error from inefficiency in a truly informative way. Nonetheless, in cases where the SFE model is well-specified and measurement error is present, we might expect SFE to provide more accurate estimates of efficiency than DEA.

Regression-based methods of estimating efficiency require specification of a functional form for the production function. If this is misspecified, then the regression estimates, upon which the efficiency estimates are ultimately based, will be biased. DEA on the other hand constructs a piecewise linear production frontier. This is a highly flexible functional form that can approximate most actual technologies. Thus, we would expect DEA to provide better estimates of efficiency in cases in which the functional form of the COLS model is misspecified.

Correlation between inputs and inefficiency will also bias OLS estimates of production function coefficients. Orme and Smith (1996) argue that the presence of such correlation can also bias DEA efficiency estimates. However, Bifulco and Bretschneider (2001) do not find any support for Orme and Smith's argument. Thus, we might expect the presence of endogeneity to have a larger impact on the performance of COLS and SFE efficiency estimates than DEA estimates

These expectations are by and large confirmed by the results reported in Tables 1. In cases where the COLS model is well-specified (i.e. the first three rows of Table 1), COLS and SFE performs better than DEA on the rank correlation criteria. DEA tends to perform better in cases where the regression model is misspecified. In cases with endogeneity but no measurement error, DEA II and the Ruggiero approach both outperform COLS and SFE. When the functional form of the regression model is also misspecified (i.e. the last three rows of Table 1), the Ruggiero method achieves higher rank correlations than COLS and SFE.

In cases where measurement error is present and the regression model is misspecified, both DEA and the regression-based methods perform poorly. It might be argued that these are the conditions most likely to be encountered in attempts to measure the efficiency of educational organizations. In these cases, the presence of measurement error substantially diminishes the performance of DEA, and the combination of measurement error and misspecification significantly diminishes the performance of COLS. In cases with measurement error, endogeneity and misspecified functional forms, rank correlations higher than 0.306 are never achieved, and in half of these cases rank corrlations are below 0.15. It is doubtful that rank correlations of this magnitude are adequate for the purposes of awarding performance bonuses or targeting remedial resources.

Adequacy of Efficiency Estimates for Purposes of Performance-Based Reform

The results reported in Tables 1 raise doubts about whether our ability to estimate school efficiency is adequate for the purposes of performance-based school reform. To investigate this issue further we divided the observations in each of the 12 data sets into quintiles based on their true efficiency score. We then examined the ability of the most effective methods in the above analyses—SFE, DEA II and the Ruggiero approach—to place observations in the appropriate quintiles. We also examined the true efficiency rankings of the schools identified by these methods as being in the lowest efficiency quintile. The results of these analyses are presented in Tables 2 and 3.

SFE did well in cases where the underlying regression model was well-specified (see first three rows of Tables 2 and 3). Particularly in cases with low measurement error, SFE assigned 74 percent of schools to the appropriate quintile and only 2 out of 200 schools were assigned to a quintile two or more away from their true quintile. The method also did well identifying the lowest performing schools. Of the schools assigned to the bottom quintile by SFE, 95% were actually in the bottom efficiency quintile and none of the schools had true efficiency values that ranked them higher in efficiency than the median.

However, the SFE method did not do as well in cases where the underlying regression model is mispecified. In cases where the functional form is misspecified, SFE places more schools in quintiles two or more away from their true quintile than it places in the correct quintile. In cases where endogeneity is present and the SFE model is misspecified, less than half of the schools identified as being among the schools with the lowest level of efficiency are actually in the bottom efficiency quintile, and at least 14 of the 40 schools placed in the bottom quintile have true efficiency values that rank them above the median.

The two DEA methods did not do well in placing students in the correct quintile in any of the data sets. In no case did either DEA II or the Ruggiero approach place as many as half the schools in the correct quintile. In the majority of cases these methods place as many or more schools in a quintile two or more away from the true quintile as they place in the correct quintile. DEA II and the Ruggiero approach did reasonably well identifying low-efficiency schools, but only in the unrealistic cases where there is no measurement error.

It appears that if the underlying regression model is well-specified, then SFE can provide efficiency estimates that are adequate for at least some purposes. However, the past 35 years of experience in trying to analyze educational production suggests that the functional relationships between educational outcomes, school inputs and environmental factors are complex and that we know little about the forms these relationships take. Unfortunately, SFE does much worse when the underlying regression model is misspecified. DEA has been advanced has a method of estimating efficiency that does not depend on restrictive assumptions about the form of productive relationships. However, the estimates of efficiency provided by DEA, particularly in the presence of measurement error, do not appear to be adequate.

Whether the performance of SFE when the underlying regression model is misspecified or the DEA methods in the presence of measurement error is adequate for the purposes of school-based reform is a matter of judgment. However, it is difficult to argue that the results in Tables 2 and 3 are adequate. In cases with endogeneity, measurement error and a more complex production function (i.e. the last three rows), the best we were able to do is place 31 percent of the schools in the correct quintile. In these cases, at least 58 out of 200 schools were placed a quintile two or more away from their true group. If such a method were relied on to determine financial awards or target corrective action, a large number of schools that lose out on additional resources or face burdensome requirements would have legitimate complaints. It also seems unlikely that analyzing the practices of groups identified as high or low performing by these methods would be very informative. If less than half of the schools that are identified as low performing are actually inefficient, and 30 percent are actually achieving above average levels of efficiency, then it is difficult to say that the managerial practices or patterns of resource allocation found in those schools are ineffective.

8. Conclusions

Existing studies as well as the new evidence presented here suggest that for the complex production processes found in schools, i.e. processes characterized by complex functional forms, endogenous relationships between inputs and outputs, and substantial measurement error, the most commonly applied versions of DEA and regression-based methods do not provide adequate measures of efficiency. It would be difficult to defend implementing performance-based financing or management programs with estimates of school performance whose rank correlation with true performance is no higher than 0.30. However, our results need not be interpreted with unequivocal gloom. Not only must our findings be properly qualified, but they also suggest strategies for developing more adequate measures of efficiency.

COLS and SFE perform well in cases where the underlying model is well-specified, particularly when measurement error is small. DEA also performs much better in cases without measurement error. Some forms of DEA, particularly Ruggiero's approach, also appear to be fairly robust with repect to the functional relationships between outcomes, inputs and the environment. This suggests at least three avenues for improving efficiency measurement.

First, efforts to reduce the amount of measurement error characteristic of current educational data sets are needed. Such efforts are well underway. The 1994 reauthorization of the Elementary and Secondary Education Act provided substantial amounts of funding to state educational agencies to develop testing programs that are aligned with explicit curricular goals, that test higher level thinking skills and that can be used for purposes of evaluating school performance. States, such as Kentucky, have led the way in the development of such assessment systems.[12] In addition, several city school districts, including Chicago and New York City, have developed school-based budgeting systems. These systems provide more reliable school-level resource data than has ever before been available (Rubenstein, 1998; Iatarola and Stiefel, 1998).

In addition to reducing measurement error, it might be possible to modify existing methods of estimating efficiency so as to minimize the effect of measurement error and/or endogeneity. For instance, the fact that the performance of COLS and SFE is diminished by correlation between inputs and inefficiency is not surprising. This type of correlation violates the assumptions that are required if ordinary least squares is to provide unbiased coefficient estimates. Bias in these coefficient estimates is the source of the poor performance of COLS and SFE in estimating efficiency. There are, however, well known simultaneous equation methods, such as two-staged least squares, that provide unbiased coefficient estimates in cases where the assumptions of ordinary least squares are violated. If such methods could be used to estimate production frontiers, then efficiency estimates that perform better than those we have examined might be developed.

Finally, efforts to understand the functional forms that characterize educational production are needed. These efforts may be the most important for improving efficiency measurement and the most difficult to achieve. However, with continued efforts to develop theory and test those theories with more complex empirical models, we may be able to make progress on this front. The use of flexible functional forms, such as the translog production function, might also help provide more accurate estimates of efficiency by relaxing some of the restrictive assumptions about production technology made in typical regression models.[13]

In addition, we must not overlook the possibility of augmenting quantitative measures of efficiency with qualitative forms of evaluation. Such qualitative forms of evaluation might involve site visits and audits by professional peers. Research is needed to determine exactly how information acquired through such methods can be combined with existing data and methods to develop more reliable and valid measures of school performance.

Given the data that are currently available, however, our results suggest that the methods for measuring the efficiency of educational organizations that have been used most frequently, may not be adequate for use in implementing performance-based management systems. This is a discouraging result, and suggests that efforts to implement performance-based school reforms should proceed with caution.

Notes

1. For examples and analysis of state level efforts see Richards and Sheu (1992), Elmore, *et al.* (1996) and King and Mathers (1997).

2. Bessent and Bessent (1980) and Bessent *et al.* (1982, 1983) have applied the basic formulation of DEA developed by Charnes, Cooper and Rhodes (1978) to schools in Houston. Fare *et al.*(1989) have applied a version of DEA that allows for variable returns to scale to school districts in Missouri. More recently, Ray (1991), McCarty and Yaisawarng (1993), Ruggiero *et al.* (1995) and Kirjavainen and Loikkanen (1998) have applied DEA-based approaches that attempt to control for the different environmental factors faced by educational organizations. With regard to regression-based approaches, Barrow (1991), Deller and Rudnicki (1993), and Cooper and Cohn (1997) have applied stochastic frontier estimation methods to estimate the efficiency of districts, schools and classes. Stiefel *et al.* (1999) have recently reviewed the various methods available for measuring school efficiency enumerating some of the advantages and disadvantages of each.

3. See Bauer (1990) for a review of a number of techniques for estimating stochastic frontiers.

4. Stiefel, Schwartz and Rubenstein (1999) discuss alternative regression-based measures of efficiency for cases where school-level panel data are available. Repeated observations on individual schools provided by panel data allows the estimation of school fixed-effects. Stiefel, Schwartz & Rubenstein suggest that these fixed-effects may provide a better measure efficiency than the residual from cross-sectional ordinary least squares regressions because not all of the residual variation is attributed to efficiency. As the authors point out, however, a school fixed-effect reflects all systematic variation in outputs that is not explained by

observed inputs, and therefore, is likely to reflect more than just differences in efficiency. In fact, because estimation of school fixed-effects precludes inclusion of time-invariant inputs in the regression equation, the fixed-effects reflects differences in these inputs as well as the impact of factors that are typically difficult to measure and include in cross-sectional regressions. Thus, it is doubtful that this use and interpretation of panel data estimates provides improved measures of efficiency, and may even be more misleading than measures based on the residual of cross-sectional estimators.

5. The production frontier identified by DEA is a piecewise linear surface connecting each school that receives an efficiency rating of one. A school's efficiency score can be interpreted as its distance from this piecewise linear production frontier.

6. For discussions of these difficulties see Bridge et al. (1979) and Monk (1990).

7. Aigner *et al.* (1977) and Olson *et al.* (1980) have used simulated data do compare different econometric methods for estimating stochastic production frontiers. Orme and Smith (1996) have published simulation studies that examine particular properties of a single DEA method, and Ruggiero (1996) and Ruggiero and Bretschneider (1998) have used simulated data to compare different linear programming models. For simulation studies comparing stochastic frontier estimators and DEA see Gong and Sickles (1992) and Banker *et al.* (1993).

8. This interpretation of the residual is admittedly *ad hoc*. A fully stochastic approach would decompose the residual that remains after the intercept correction into an efficiency and random component, again based on *a priori* assumptions about the probability distribution of these two error components. Jondrow, *et al.* (1982) details the procedure for this decomposition. Arguments made by Ruggiero and Ondrich (1997) demonstrate that the measures of efficiency with and without the decomposition are ordinally equivalent.

9. Both Ray (1991) and Kirjavainen and Loikkanen (1998) use a Tobit model to estimate the second stage regression. Given that the DEA estimate is truncated at the value of one this is appropriate. The application of “Two Stage DEA” in this paper, uses OLS to estimate a general linear regression model for the second stage. In past work done by the authors, we have found that this simplification has little effect on the resulting efficiency estimates.
10. In results not reported here we found that the mean absolute difference between the true efficiency and the estimated efficiency scores were larger for SFE than for COLS in cases without measurement error, and considerably smaller for SFE than for COLS in cases with measurement error. This shows that although ordinally equivalent, the efficiency measures provided by these two methods are not cardinally equivalent.
11. Ruggiero and Ondrich (1997) demonstrate that the ordinal measure of efficiency provided by COLS is equivalent to that provided by the fully stochastic frontier of Jondrow, *et al.* (1982) as well.
12. See *Education Week*’s *Quality Counts 2001* publication entitled *A Better Balance: Standards, Tests and the Tools to Succeed* for a discussion of state efforts to develop improved assessments of student performance.
13. See Beattie and Taylor (1985) for details on the use of flexible functional forms in production analysis.

Table 1
Rank correlations between estimated and true efficiency values*

	COLS	SFE	DEA I	DEA II	Two Stage DEA	Ruggiero Approach
Cobb-Douglas Technology with Additive Environment						
Without Endogeneity						
No Measurement Error	0.866	0.866	0.173	0.575	0.156	0.535
Small Measurement Error	0.835	0.835	0.154	0.192	0.144	0.128
Large Measurement Error	0.596	0.596	0.095	0.071	0.085	0.103
With Endogeneity						
No Measurement Error	0.457	0.459	0.247	0.616	0.274	0.585
Small Measurement Error	0.484	0.487	0.239	0.362	0.264	0.215
Large Measurement Error	0.380	0.399	0.273	0.302	0.285	0.244
Cobb-Douglas Technology with Interactive Environment						
Without endogeneity						
No Measurement Error	0.290	0.280	0.138	0.143	0.122	0.505
Small Measurement Error	0.264	0.262	0.141	0.138	0.133	0.118
Large Measurement Error	0.259	0.255	0.108	0.087	0.090	0.095
With Endogeneity						
No Measurement Error	0.104	0.108	0.215	0.275	0.233	0.458
Small Measurement Error	0.099	0.104	0.213	0.264	0.237	0.203
Large Measurement Error	0.076	0.090	0.260	0.306	0.281	0.233

* Correlations are Kendall Tau-b statistics

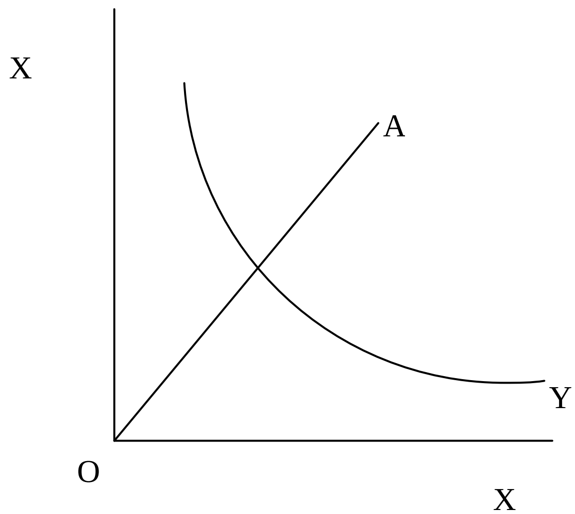
Table 2
Measures of how well various methods do in assigning observations to quintiles

	SFE		DEA II		Ruggiero Approach	
	% Assigned to Correct Quintile	% Assigned Two or More Quintiles from Actual	% Assigned to Correct Quintile	% Assigned Two or More Quintiles from Actual	% Assigned to Correct Quintile	% Assigned Two or More Quintiles from Actual
Cobb-Douglas Technology with Additive Environment						
Without Endogeneity						
No Measurement Error	80.0%	0.0%	36.0%	19.0%	34.0%	26.5%
Small Measurement Error	74.0%	1.0%	27.5%	42.0%	22.0%	37.0%
Large Measurement Error	49.0%	11.5%	20.5%	43.5%	21.0%	38.0%
With Endogeneity						
No Measurement Error	38.5%	24.0%	41.5%	15.5%	39.0%	24.5%
Small Measurement Error	41.5%	19.5%	30.5%	30.0%	26.5%	35.5%
Large Measurement Error	34.0%	28.0%	32.0%	29.0%	26.5%	36.5%
Cobb-Douglas Technology with Interactive Environment						
Without endogeneity						
No Measurement Error	26.5%	38.5%	26.5%	41.0%	35.0%	29.5%
Small Measurement Error	26.5%	37.5%	24.5%	41.5%	22.5%	37.5%
Large Measurement Error	28.5%	38.0%	20.5%	45.0%	22.5%	37.5%
With Endogeneity						
No Measurement Error	27.5%	46.0%	30.5%	30.0%	33.5%	32.0%
Small Measurement Error	27.0%	47.5%	31.0%	30.5%	26.5%	35.5%
Large Measurement Error	24.5%	46.0%	31.0%	29.0%	25.0%	36.5%

Table 3
Measures of how well various measures do in identifying low efficiency schools

	SFE		DEA II		Ruggiero Approach	
	% Assigned to Bottom Quintile Actually in Bottom Quintile	% Assigned to Bottom Quintile Actually Ranked Above Median	% Assigned to Bottom Quintile Actually in Bottom Quintile	% Assigned to Bottom Quintile Actually Ranked Above Median	% Assigned to Bottom Quintile Actually in Bottom Quintile	% Assigned to Bottom Quintile Actually Ranked Above Median
Cobb-Douglas Technology with Additive Environment						
Without Endogeneity						
No Measurement Error	97.5%	0.0%	72.5%	0.0%	67.5%	0.0%
Small Measurement Error	95.0%	0.0%	45.0%	35.0%	30.0%	47.5%
Large Measurement Error	80.0%	0.0%	37.5%	42.5%	27.5%	45.0%
With Endogeneity						
No Measurement Error	67.5%	5.0%	72.5%	0.0%	72.5%	0.0%
Small Measurement Error	72.5%	5.0%	47.5%	12.5%	35.0%	32.5%
Large Measurement Error	62.5%	12.5%	50.0%	22.5%	37.5%	30.0%
Cobb-Douglas Technology with Interactive Environment						
Without Endogeneity						
No Measurement Error	57.5%	10.0%	45.0%	37.5%	77.5%	0.0%
Small Measurement Error	57.5%	10.0%	40.0%	45.0%	30.0%	47.5%
Large Measurement Error	57.5%	10.0%	35.0%	42.5%	27.5%	45.0%
With Endogeneity						
No Measurement Error	40.0%	37.5%	42.5%	22.5%	62.5%	0.0%
Small Measurement Error	37.5%	35.0%	42.5%	22.5%	35.0%	32.5%
Large Measurement Error	35.0%	37.5%	52.5%	22.5%	32.5%	32.5%

Figure 1: Technical Efficiency



. Technical Efficiency for Firm A = OB / OA

References

Aigner, D., Lovell, C.A.K., and Schmidt, P. (1977) Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* **6**, 21-37.

Banker, R.D., Conrad, R.F. and Strauss, R.P. (1985) A comparative application of DEA and translog methods: An illustrative study of hospital production. *Management Science* **32**, 30-44.

Banker, R.D., Gadh, V.M., and Gorr, W.L. (1993) A monte-carlo comparison of two production frontier estimation methods: Corrected ordinary least squares and data envelopment analysis," *European Journal of Operational Research* **67**, 332-343.

Barrow, M.M. (1991) Measuring local education authority performance: A frontier approach. *Economics of Education Review* **10**, 19-27.

Bauer, P.W. (1990) Recent developments in the econometric estimation of frontiers. *Journal of Econometrics* **46**, 39-56.

Beattie, B.R. and Taylor, C.R. (1985) *The Economics of Production*. John Wiley & Sons, New York.

Bessent, A. and Bessent, E. (1980) Determining the comparative efficiency of schools through data envelopment analysis. *Educational Administration Quarterly* **16**, 57-75.

Bessent, A., Bessent E., Charnes, A., Cooper, W. and Thorogood, N.C. (1983) Evaluation of educational program proposals by means of DEA." *Education Administration Quarterly* **19**, 82-107.

Bessent, A., Bessent W., Kennington, J. and Reagan, B. (1982) An application of mathematical programming to assess productivity in the Houston Independent School District", *Management Science* **28**, 1335-66.

Bifulco, R. and Bretschneider, S. (2001) Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review* **20.**

Bridge, G., Judd, C. and Moock, P. (1979) *The Determinants of Educational Outcomes: The Impact of Families, Peers, Teachers and Schools*. Ballinger Publishing Company, Cambridge, MA.

Brooks, A.C. (2000). The use and misuse of adjusted performance measures. *Journal of Policy Analysis and Management* **19**, 323-329.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978) Measuring the efficiency of decision making units." *European Journal of Operational Research* **2**, 429-44.

Cooper, S.T. and Cohn, E. (1997) Estimation of a frontier production function for the South Carolina educational process. *Economics of Education Review* **16**, 313-327.

Darling-Hammond, Linda. (1991) Accountability mechanisms in big city school systems. *ERIC/CUE Digest*, **17**.

Deller, S. and Rudnicki, E. (1993) Production efficiency in elementary education: The case of Maine public schools. *Economics of Education Review* **12**, 45-57.

Elmore, R., Abelmann, C. and Fuhrman, S. (1996) The new accountability in state education reform: From process to performance." In *Holding Schools Accountable: Performance-Based Reform in Education*, ed. H. F. Ladd, pp. 65-98. The Brookings Institution, Washington D.C.

Färe, R., Grosskopf, S. and Weber, W. (1989) Measuring school district performance. *Public Finance Quarterly* **17**, 409-28.

Gong, B. and Sickles, R.C. (1992) Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data, *Journal of Econometrics* **51**, 259-284.

Hanushek, E. *et al*. (1994) *Making Schools Work: Improving Performance and Controlling Costs*. The Brookings Institution, Washington, D.C.

Iatarola, P. and Stiefel, L. (1998) School-Based Budgeting in New York City: Perceptions of School Communities, *Journal of Education Finance* **23**, 557-576.

Jondrow, J., Lovell, C.A.K., Materov, I.S. and Schmidt, P. (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* **19**, 233-238.

King, R., and Mathers, J. (1997) Improving schools through performance-based accountability and financial rewards. *Journal of Education Finance* **23**, 147-176.

Kirjavainen, T. and Loikkanen, H.A. (1998) Efficiency Differences of Finnish Senior Secondary Schools: An application of DEA and Tobit analysis. *Economics of Education Review* **17**, 377-394.

Levin, H.M. (1997) Raising school productivity: An x-efficiency approach. *Economics of Education Review* **16**, 303-311.

McCarty, T. and Yaisawarng, S. (1993) Technical efficiency in New Jersey school districts. In *The Measurement of Productive Efficiency*, eds. H. Fried, K. Lovell, and S. Schmidt, pp. 271-287. Oxford University Press, Oxford.

Monk, David. (1990) *Educational Finance: An Economic Approach*. McGraw-Hill Publishing Company, New York.

Nelson, R.A. and Waldman, D.M. (1986) Measuring technical efficiency: index numbers vs. parametric production frontiers. Unpublished Manuscript.

Olson, J.A., Schmidt, P., and Waldman, D.A., (1980). “A monte-carlo study of estimators of stochastic frontier production functions. *Journal of Econometrics* **13**, 67-82.

Ondrich, J. and Ruggiero, J. (1997) Efficiency measurement in the stochastic frontier model. *Mimeo*, Syracuse University.

Orme, C. and Smith, P. (1996) The potential for endogeneity bias in data envelopment analysis,” *Journal of the Operational Research Society* **47**, 73-83.

Ray, S. C. (1991) Resource use in public schools: A study of Connecticut. *Management Science* **37**, 1520-1628.

Richards, C. and Sheu, T. (1992) The South Carolina School Incentive Reward Program: A policy analysis. *Economics of Education Review* **11**, 71-86.

Rubenstein, R. (1998) Resource Equity in the Chicago Public Schools: A School-Level Approach, *Journal of Education Finance* **23**, 468-489.

Ruggiero, J. (1996) On the Measurement of Technical Efficiency in the Public Sector. *European Journal of Operational Research* **90**, 553-565.

Ruggiero, J. and Bretschneider, S. (1998). The weighted Russell measure of technical efficiency. *European Journal of Operations Research* **108**, 438-451.

Ruggiero, J., Duncombe, W. and Miner, J. (1995) On the measurement and causes of technical inefficiency in local public services: With an application to public education. *Journal of Public Administration Research and Theory* **5**, 403-428.

Stiefel, L., Schwartz, A. and Rubenstein, R. (1999) Measuring school-level efficiency. In *School Based Budgeting*, eds Goertz and Odden, pp. 40-75. Corwin Press, Inc., Thousand Oaks, CA.
.