

## **A Flexible-Weights School Effectiveness Index**

E. Anthon Eff \*

*Middle Tennessee State University, Murfreesboro, TN*

### *Abstract*

Accountability reforms have led to a bewildering array of public school performance measures. Policy makers, parents, and the public would often like to use this information to rank schools by differing degrees of effectiveness. Combining measures from different performance dimensions into a single index should be done in such a way that the resulting index is fair to each school, gives each school the incentive to change in the most desirable directions, and reduces the confusing mass of information to the simplest possible ordinal rank differences. This paper proposes a flexible weights approach, based on a modification of Data Envelopment Analysis. The resulting index maximally reduces information, is fair, and allows policymakers some discretion in guiding the direction schools take in their improvement.

*Key words:* Data Envelopment Analysis; Educational Quality

*JEL category:* I20

\*E. Anthon Eff, Associate Professor Department of Economics and Finance, Middle Tennessee State University, Murfreesboro, TN 37132, phone: 615-898-2387, fax: 615-898-5045, email: [eaeff@mtsu.edu](mailto:eaeff@mtsu.edu)

## A FLEXIBLE-WEIGHTS SCHOOL EFFECTIVENESS INDEX

*Abstract:* Accountability reforms have led to a bewildering array of public school performance measures. Policy makers, parents, and the public would often like to use this information to rank schools by differing degrees of effectiveness. Combining measures from different performance dimensions into a single index should be done in such a way that the resulting index is fair to each school, gives each school the incentive to change in the most desirable directions, and reduces the confusing mass of information to the simplest possible ordinal rank differences. This paper proposes a flexible weights approach, based on a modification of Data Envelopment Analysis. The resulting index maximally reduces information, is fair, and allows policymakers some discretion in guiding the direction schools take in their improvement.

### I. INTRODUCTION

Mandatory testing is a cornerstone of current U.S. educational reform. In the 1960s and 1970s reform efforts targeted inputs, such as teacher student ratios, and sought more equitable spending. By the 1980s, it was clear that these input-based reforms had failed to improve student performance. States then began mandating standardized tests to make schools accountable for student outcomes (Swanson & Stevenson 2002). The 2002 *No Child Left Behind Act* continues this trend: employing mandatory tests to measure how well each school carries out its core mission of educating students (United States Department of Education 2003).

While mandatory testing addresses the need for outcomes assessment, it has been criticized on several counts. First, curricula and class time are often adjusted to emphasize preparation for mandatory tests, so that “teaching to the test” leaves little time for objectives that are arguably more important but less easily measurable—objectives such as fostering critical thinking (Lomax, West, Harmon, Viator & Madaus. 1995). Second, standardized tests implicitly force all students into the same mold, when in fact many educators would maintain that student diversity requires diversity in methods of student assessment (Harris & Ford 1991).

One problem that has not attracted attention, however, is the problem of information overload. Students take so many tests, on so many different subjects, that it is difficult to distill a

meaningful judgment from the blizzard of information.<sup>1</sup> The large number of tests in different subject areas helps administrators and teachers identify areas of strength and weakness, and thus facilitates improvement. However, human cognitive limits are such that no one can readily use the information to make a summary judgment on how well a particular school is doing relative to other schools.

Summary judgments—in which the evaluator ranks a school relative to others—are often desirable. Homebuyers might wish to know which are the best schools in their area. State officials might wish to identify the worst schools, in order to focus remedial efforts. Administrators and researchers might wish to identify the best schools, so that their management practices can be examined and emulated. In the absence of an agreed-upon method for combining the various test scores into a single measure of school quality, partisanship and spin are likely to prevail. A school’s stakeholders can single out the few test scores on which the school performs particularly well, and proclaim the school as one of the best.<sup>2</sup> Detractors are equally privileged: they can single out the few test scores on which the school performs most poorly and proclaim the school as one of the worst. The abundance of performance measures thus may lead to confusion over the actual merits of any school.<sup>3</sup>

---

<sup>1</sup> In the state of Tennessee, for example, each student takes an exam each year in grades three through eight. The exam has five subject areas, and both the normed attainment score and the “value-added” from the student’s previous exam are calculated. From just this one exam, a school with grades three through eight will each year generate 30 attainment scores and 25 value-added scores. In addition, there are writing assessments in grades four, seven, and 11; a math and English competency exam in grade nine; end-of-course exams after many high school courses; and a requirement to take either the ACT, SAT, or the Work-Keys vocational test before graduation (Paige and Xu 1995; Tennessee State Board of Education 2000; Tennessee Department of Education 2002).

<sup>2</sup> Textbooks instructing school administrators in the art of public relations encourage officials to proclaim that they are “pleased” with the test results, pointing to the areas in which they did well, and then announce that they have already begun working on the few areas of weakness (Jones 1978: 144).

<sup>3</sup> The 2002 Tennessee gubernatorial race provides an example of how the same standardized test results can be interpreted in contradictory ways. One candidate—the former mayor of a large city—pointed with pride to excellent test scores; his opponent selected other scores, and described them as a “dismal record” (Humphrey 2002).

The purpose of this paper is to develop a method for combining measures from multiple performance dimensions into a single overall performance index. The proposed method employs flexible weights to create an index that maximally reduces information, and does this in a way that is fair to the evaluated education providers. The paper is organized as follows. The first section discusses the desirable characteristics of a school performance index, and proposes a method based upon a modification of Data Envelopment Analysis. The method is applied to an example data set of nine schools from an urban school district. Finally, the method is examined to determine how well it meets the desirable characteristics.

## **II. A PROPOSED SCHOOL EFFECTIVENESS INDEX**

*School effectiveness* is a measure of how well public schools meet the performance standards set by legislators and administrators.<sup>4</sup> A good school effectiveness index should have several characteristics. First, it should consolidate the bewildering array of school performance data into the smallest possible amount of meaningful information. The information should be meaningful to those who refer to it; it should help parents identify the best schools, and it should give schools a sense of how they stand relative to other schools. Second, the index should be fair. A fair index will be readily accepted, both by the schools that are being evaluated, and by the public. When schools accept the legitimacy of the index, they accept the notion that the index reveals which schools are the best, and they accept the index as a measure of their own improvement. The index will therefore guide improvement. This leads to the third characteristic of a good school effectiveness index, that it call forth the *right* kind of improvement, so that the improvements giving the evaluated school the greatest increase in its index (and which, rationally, it should pursue) are the kinds of improvements that policymakers would most like to see.

---

<sup>4</sup> Effectiveness looks only at outputs—at the performance of an organization in meeting its goals. Efficiency, on the other hand, considers not only outputs but also inputs; comparing performance with the resources used to attain that performance.

An effectiveness index combines performance on different dimensions into a single ordinal scale.

In its most general form:

$$\theta_i = \sum_{r=1}^m \mu_r y_{ri} \quad , \quad \forall i \quad (1)$$

where the effectiveness  $\theta_i$  of the  $i^{\text{th}}$  education provider is the sum of the scores  $y_{ri}$  in  $r=1, \dots, m$  performance dimensions, each score weighted by a weight  $\mu_r$ . Specifying the index requires that one address the following three issues: which education providers  $i$  to include; which test scores  $r$  to include; and how to specify the weights  $\mu_r$ .

#### *Determining the Set of Education Providers*

Determining the set of  $i=1, \dots, n$  education providers is mostly a matter of level of aggregation: is it more appropriate to compare schools, school districts, or counties? The entities should be directly comparable, they should have similar missions and similar bodies of students. Thus schools with grades kindergarten through four should only be compared to other schools with grades kindergarten through four.

#### *Determining the Set of Performance Dimensions*

Determining the set of  $r=1, \dots, m$  performance dimensions is relatively easy, since legislation mandates specific tests. The set of tests should represent most of what education providers do, so that they are assessed on their whole effort, and not just on a portion of it. The scores should also be relevant, they should represent dimensions which both the education providers and the public agree are important. The collaborative nature of the legislative process, where legislators respond to the interests of the public and solicit the testimony of experts, suggests that the most important performance dimensions would be singled out for mandatory testing. Additionally, it is fair to use the set of mandated tests to evaluate schools, since school officials know that they are

accountable for the test results and focus their improvement efforts on the performance dimensions defined by the tests.

#### *Determining the Appropriate Weights*

The weights  $\mu_r$  reflect the relative importance of each performance dimension. If all performance measures  $y_{ri}$  are of equal importance and similarly scaled, one could set all  $\mu_r = 1/m$ , so that the index is a simple average. If policy makers think that some performance dimensions are more important, they can set higher weights on those dimensions, thus encouraging education providers to focus their efforts toward improvement on dimensions with high weights. In many cases, though, the appropriate weights are not known, because there is no consensus on the relative importance of performance dimensions.

Figure 1 gives a hypothetical example of eight different schools evaluated on just two dimensions. The figure shows, for example, that school A's strengths lie in reading, while school B's strengths are in mathematics; the differences among the schools may have to do with different missions, or may have to do with the efficient use of differentiated resources. School A would prefer evaluators to put a high weight on reading performance and a low weight on math, to give the school full credit for its specialized mission or for its comparative advantage. If there is no compelling *a priori* reason to impose a set of weights that is the same for each of the entities evaluated, one could use the weights that cast each school's performance in the best possible light. This approach would be fair in the sense that no school could complain that the weighting scheme distorts its true performance (Rouse, Putterill & Ryan 1997: 129).

*<Insert Figure 1 here>*

The problem of assigning weights to each performance dimension can be addressed with Data Envelopment Analysis (DEA) (Charnes, Cooper & Rhodes 1978). DEA does this by wrapping a multi-dimensional envelope or "frontier" around the set of points. Each school is then evaluated

according to how close it lies to the frontier. In the two-dimensional example of Figure 1, this method awards three schools (A, C, and B) the rank of highest performers, with the others performing more or less well according to how close they lie to the frontier. Expressing the school's position as the percentage of the distance from the origin to the frontier gives a distance index ranging between zero and one. Thus, schools A, C, and B have a distance index score of one, while school G's index score would equal the distance ratio  $OG/Oz_1$ , school H's score would be  $OH/Oz_1$ , and school F's score would be  $OF/OA$ .

Though many of DEA's developers acknowledged its potential for measuring effectiveness (Golany & Tamir 1995: 1173), education analysts have used it almost exclusively to calculate efficiency. Nevertheless, education officials and the public are usually more interested in effectiveness measures than in efficiency. Perhaps because "there is no strong or consistent relationship between school inputs and student performance" (Hanushek 1997: 148), the most efficient school districts tend to be those that spend little, often rural districts with weak tax bases. Since education officials are not accustomed to regard poorly funded districts as exemplars, they tend to treat efficiency measures with skepticism and to focus on effectiveness when ranking schools. Parents' interest in public school efficiency is blunted because parents bear only a portion of the cost (their tax bill), while receiving all of the benefit. Within a large urban school district, parents have no incentive to move close to the most efficient school (since their tax rate is the same throughout the district) but every incentive to move close to the most effective. In addition, expenditure data are typically available only at the district level, not at the school level, so that efficiency measures cannot be calculated at the school level. Thus, effectiveness measures often excite more interest than efficiency measures in public school analysis. The formal specification of a DEA effectiveness model is given below.

Assume there are  $n$  schools, denoted by subscript  $i$ , whose performance is measured in  $m$  dimensions, denoted by subscript  $r$ ;  $y_{ri}$  is the measure of the performance of school  $i$  on

dimension  $r$ . The following linear programming problem solves for weights on the individual performance measures ( $\mu_r$ ) in order to assess the effectiveness of the  $k^{\text{th}}$  school.

$$\text{Maximize} \quad \sum_r \mu_r y_{rk} \quad (2.a)$$

$$\text{Subject to} \quad \sum_r \mu_r y_{ri} \leq 1, \quad \forall i \quad (2.b)$$

$$\mu_r y_{rk} \leq b_U \mu_{\bar{r}} y_{\bar{r}k}, \quad \forall r \quad (2.c)$$

$$\mu_r y_{rk} \geq b_L \mu_{\bar{r}} y_{\bar{r}k}, \quad \forall r \quad (2.d)$$

$$\mu_r \geq 0, \quad \forall r \quad (2.e)$$

The objective function (2.a) selects weights in order to maximize the performance score of the  $k^{\text{th}}$  school. Constraint (2.b), however, restricts the weights so that—applied to every one of the  $n$  schools—no school has an efficiency score higher than one. Thus, the highest value that the objective function may take is one, and then only when the  $k^{\text{th}}$  school lies on the high performance frontier. In all other cases, the value of the objective function will be less than one, since the weights that maximize its score give another school a score of one. In these cases—where the score is less than one—the score may be interpreted as the percent of the best school’s score obtained by the  $k^{\text{th}}$  school.

Constraints (2.c) and (2.d) apply restrictions on the weighted share in the objective function of each performance measure relative to a numeraire performance measure’s share ( $\mu_{\bar{r}} y_{\bar{r}k}$ ), so that each weighted share can only differ from any other share by a magnitude of  $b_U/b_L$ . DEA weight restrictions of this specification have been proposed by Pedraja, Salinas & Smith (1997), based on the results of Monte Carlo simulations. While in some contexts it may be perfectly acceptable to allow weights of zero, in most cases flexible weights should be flexible only within a range. Pedraja, et al. (1997) select  $b_U = 2^{1/2}$  and  $b_L = 2^{-1/2}$  so that the weighted shares can only differ from



each other by a magnitude of two, an order of magnitude that seems reasonable for educational evaluation.

The constrained maximization problem in equations (2.a)-(2.e) is solved  $n$  times—once for each of the  $n$  schools. The value of the objective function for a particular school provides the distance index score for that school. The distance index has a nice intuitive interpretation, since each school's score is the percentage of the score attained by the actual best performers. Nevertheless, in many cases, one might instead wish to classify a school with a peer group of similar performance. This objective can be accomplished by wrapping successive frontiers around the set of schools (Figure 2). Higher frontiers correspond to higher levels of performance; schools situated on the same frontier have comparable levels of performance. Barr & Seiford (2000), apply the name 'Tiered DEA' (TDEA) to this sequential frontier technique.

The schools on the highest performance frontier—frontier 1 in Figure 2—are those with a distance index score of one, found by solving the constrained maximization problem in equations (2.a)-(2.e)  $n$  times. Dropping these high performance schools from the set of schools, one recalculates the distance index score for each of the remaining schools; those with a score of one are on frontier 2. Again, the frontier schools are dropped and the distance index is recalculated for each of the remaining schools; schools with a score of one are on frontier 3. One continues the procedure until all schools have been assigned to a frontier.

*<Insert Figure 2 here>*

For purposes of performance evaluation, TDEA may have certain advantages. When the purpose of evaluation is to classify schools into peer groups, TDEA will find natural breakpoints, and thus reduce the subjectivity of the classification. In addition, a school might complain that the inclusion of a particular high performing school pushed the frontier outward, so that the school's distance index became unfairly low. Such a complaint might be justified in cases where one suspects the performance of a frontier school has been measured with error, or in cases where the

decision to include a particular frontier school in the analysis is arbitrary. In such cases, schools would prefer to be compared with schools below the frontier, which is what TDEA does. Thus, TDEA seems to be a superior method for evaluations whose objective is classification into rank-ordered peer groups, and it seems to promote fairness since it mitigates one possible complaint that evaluated schools would have about DEA.

The following section applies the flexible-weights method to public school testing data from Tennessee.

### **III. THE FLEXIBLE-WEIGHTS METHOD APPLIED TO AN EXAMPLE DATA SET**

Nine schools are selected from the 39 public schools containing only grades kindergarten through four in the city of Nashville, Tennessee. The data are for the 1999-2000 school year. For these schools, the most important exams are the annual achievement tests in five subject areas: language arts, mathematics, reading, science, and social studies. The achievement tests are given in grades three through eight, and special attention is given to the value-added scores, since they measure the amount students have learned in the past year (Baker & Xu 1995). For each of the nine sampled schools, the value added between grades three and four is calculated, in each of the five subjects. A fourth grade writing assessment and the school promotion rate provide two more performance dimensions (Table 1).

*<Insert Table 1 here>*

Table 1 demonstrates how multiple performance dimensions, even in a small group of schools, casts confusion over school ranks. Charlotte Park can claim to be the best school based on scores for language arts and science. Cole has the best scores for mathematics, Gateway for reading, H.G. Hill for social studies, and Crieve Hall for promotion and writing. Supporters of each school are likely to single out these best scores, while detractors are likely to focus on their poorest

scores. Crieve Hall, for example, has the lowest mathematics and social studies scores. A fair method for combining scores into a composite index would remove the confusion.

Applying the constrained maximization problem in equations (2.a)-(2.e) results in the weights shown in Table 2. The final row gives, for each performance dimension, the ratio of the maximum weight over the minimum weight. The ratio gives some indication of how much schools vary in the preferred weight on each dimension: for example, the highest weight on writing attainment is over 2.4 times the lowest weight. Since promotion serves as the numeraire, its weights do not vary much across schools.

*<Insert Table 2 here>*

Table 3 reports the resulting TDEA frontier and DEA distance index  $\theta_i$  for each of the nine schools. The TDEA frontier reduces the differences among the schools to six rank-ordered groups. Charlotte Park and Granberry fall into the same TDEA group, though their DEA distance index values appear to be far apart.

*<Insert Table 3 here>*

For comparison purposes, the third column presents an effectiveness measure employing fixed and equal weights, as shown in equation (3):

$$\tilde{\theta}_i = \frac{\sum_{r=1}^m y_{ri}}{\max_i \left( \sum_{r=1}^m y_{ri} \right)} \quad (3)$$

Equation (3) is a specification of equation (1), setting all  $\mu_r = 1/758.862$ . The standardization allows each school's score to be interpreted as the percentage of the score attained by the highest scoring school. The final column compares the DEA distance index  $\theta_i$  with the fixed and equal weights index  $\tilde{\theta}_i$ . In every case, a school's measured effectiveness is higher with the flexible weights  $\theta_i$  than with the fixed and equal weights index  $\tilde{\theta}_i$ . In fact,  $\theta_i \geq \tilde{\theta}_i$  for all sets of fixed weights, with

the equality holding only when the fixed weights in  $\tilde{\theta}_i$  are the same as the optimal weights in  $\theta_i$ —a situation that can benefit only one school at a time. Schools will therefore regard the flexible weights index as fair, since the method assigns the weights that cast the school's performance in the best possible light.

By examining elasticities, one can better understand how the index might shape a particular school's quest for improvement. From equations 2.a-2.d the performance measure  $r$ 's elasticity of  $\theta_k$  is

$$\xi_r^{\theta_k} = \frac{\partial \theta_k}{\partial y_{rk}} \frac{y_{rk}}{\theta_k}, \quad \forall r, k \quad (4.a)$$

where

$$\frac{\partial \theta_k}{\partial y_{rk}} = \mu_r (1 - \lambda_k - \pi_{U,r} + \pi_{L,r}), \quad \forall r \neq \bar{r} \quad (4.b)$$

$$\frac{\partial \theta_k}{\partial y_{\bar{r}k}} = \mu_{\bar{r}} \left( 1 - \lambda_k - \pi_{U,\bar{r}} + \pi_{L,\bar{r}} - \sum_r \pi_{L,r} b_L + \sum_r \pi_{U,r} b_U \right), \quad r = \bar{r} \quad (4.c)$$

$\lambda_k$ ,  $\pi_{U,r}$ , and  $\pi_{L,r}$  are Lagrangian multipliers for constraints (2.b), (2.c), and (2.d), respectively. The Kuhn-Tucker conditions require that the multipliers equal zero if their constraints are non-binding (Chiang 1984: 725). Thus,  $\lambda_k$  will be zero unless  $\theta_k = 1$ . Both  $\pi_{U,r}$  and  $\pi_{L,r}$  cannot be nonzero, since a weight  $\mu_r$  cannot be binding on both the upper and lower constraints. When  $y_{rk}$  is low, then the objective function (2.a) would set  $\mu_r$  low (and  $\xi_r^{\theta_k}$  low), making it likely that  $\pi_{L,r} > 0$  (so that  $\xi_r^{\theta_k}$  is high). Similarly, when  $y_{rk}$  is high, then the objective function (2.a) would set  $\mu_r$  high (and  $\xi_r^{\theta_k}$  high), making it likely that  $\pi_{U,r} > 0$  (so that  $\xi_r^{\theta_k}$  is low). Thus, the weight constraints

(2.c) and (2.d) and the objective function (2.a) tend to push the elasticities  $\xi_r^{\theta_k}$  in opposite directions.

The objective function (2.a) puts higher weights  $\mu_r$  on performance dimensions in which school  $k$  has a comparative advantage. Thus, the effect of the objective function (2.a) on the elasticities  $\xi_r^{\theta_k}$  is to give schools a higher return on the dimensions in which they have a comparative advantage, so that schools improving along their paths of greatest returns are likely to become increasingly specialized.

On the other hand, the weight constraints (2.c) and (2.d) raise elasticities  $\xi_r^{\theta_k}$  on the performance dimensions in which schools do *not* have a comparative advantage. The effect of the weight constraints is to give schools an incentive to show improvements on a broad spectrum of performance dimensions, so that schools do not become specialized.

*<Insert Figure 3 here>*

Figure 3 shows how the elasticities change as the weight constraints are progressively relaxed, by allowing the order of magnitude (the ratio  $b_U/b_L$ ) to vary from one to 99. Since the elasticities are equal to zero for schools on the frontier, Figure 3 examines only the four schools that lie below the frontier, even with the most relaxed constraints (Cockrill, Goodlettsville, Gower, and Granberry). After dropping the numeraire, the performance measures for each school are divided into two groups: those three measures representing the school's comparative strengths, and those three representing the school's comparative weaknesses. The average elasticity in each group is then calculated. As Figure 3 shows, the elasticity for measures representing a school's comparative strengths increases as the constraints are progressively relaxed, while the elasticity for measures representing a school's comparative weaknesses decreases. Relaxing constraints

thus encourages schools to specialize, while tightening constraints encourages schools to perform well on all dimensions.

Policymakers in different school districts may have different views about the desirability of specialization. In some districts they may welcome specialization, particularly where students have a choice in the school they attend. Allowing schools to improve along the lines of their comparative advantage may also be most efficient, since schools with demonstrated excellence on a particular dimension are likely to have the resources and expertise that make improvements easier to accomplish on that dimension than on other dimensions. Districts wishing to encourage specialization need only set the ratio  $b_U/b_L$  to high values in weight constraints (2.c) and (2.d).

In districts where students cannot choose the school they attend, policymakers are likely to be especially concerned that all schools teach all subjects well. In such a district, specialization will be eschewed in favor of broad-based improvements. If performance improvements are characterized by diminishing returns, then it may also be more efficient to encourage schools to improve on their weakest dimensions. Broad-based improvements can be encouraged by setting the ratio  $b_U/b_L$  to low values in weight constraints (2.c) and (2.d).

#### **IV. SUMMARY AND CONCLUSIONS**

Public school mandatory testing is designed to improve accountability, but the deluge of test scores have made it difficult for parents and policymakers to rank-order schools by degrees of effectiveness. In the absence of an agreed-upon method for converting multiple performance measures into a one-dimensional ordinal scale, no one really knows how well any particular school is performing. Accountability therefore requires the creation of a school effectiveness index.

A good effectiveness index has three features. First, it reduces the confusing mass of information to a readily understood one-dimensional ordinal scale. Second, the index is fair, so that the

evaluated schools willingly accept the judgment placed on them by the index, and parents and policymakers feel no need to search out less biased information. Third, when schools attempt to increase their index score, the kinds of improvement elicited are desirable, in that they fit with the wishes of policymakers to encourage specialization among schools or to ensure that all schools teach all subjects well.

The flexible weights effectiveness index introduced here consolidates multiple performance measures into two kinds of ordinal scales, using a modification of Data Envelopment Analysis. The index ensures fairness since it views each school's performance in the most favorable light. No school can claim that the weights unfavorably bias its performance, since no set of weights exist that would give the school a higher effectiveness index.

Due to the weight constraints, all performance dimensions must make a minimum contribution to the effectiveness index. Hence, a school's quest for improvement must encompass all performance dimensions. When the weight constraints are relaxed, elasticities are higher for dimensions in which the school has a comparative advantage. Thus, by relaxing the weight constraints, policymakers can encourage schools to improve along the lines of their comparative advantage, and schools become increasingly specialized. By tightening the weight constraints, on the other hand, policymakers can encourage schools to improve their weakest performance dimensions, so that all schools teach all subjects well. The method is therefore flexible enough that policymakers can use it to elicit a variety of behaviors.

The flexible weights method presented in this paper can potentially be used in any situation where managers or analysts must combine multi-dimensional performance measures to create an ordinal performance index. The method is particularly valuable whenever fairness considerations are salient. One example in education would be faculty assessment, where research, public service, and teaching performance could be combined into an effectiveness index for purposes of promotion or merit pay. One example outside education would be quality-of-life indices, where

city-level data on crime, cost-of-living, employment, and climate are combined into a composite index. In both these cases, an index will only be accepted—by assessed faculty or assessed cities—if it combines measures in a fair way.



## REFERENCES

- Baker, A. Paige, and Deng Ke Xu. (1995). *The Measure of Education: A Review of the Tennessee Value Added Assessment System*. Report to the Comptroller of the Treasury, State Government of Tennessee. April 1995.
- Barr, R., Durchholz, M. and Seiford, L. (2000). *Peeling the DEA Onion: Layering and Rank-Ordering DMUS Using Tiered DEA*. Southern Methodist University Technical Report.
- Chiang, Alpha C. (1984). *Fundamental Methods of Mathematical Economics*. New York: McGraw-Hill, Inc.
- Charnes, Abraham, William W. Cooper, and E. Rhodes. (1978). Measuring the Efficiency of Decision-Making Units. *European Journal of Operations Research*. 2(6): 429-444.
- Cooper, William W., Lawrence M. Seiford, and Kaoru Tone. (2000). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Boston: Kluwer Academic Publishers.
- Golany, B., and E. Tamir. (1995). Evaluating Efficiency-Effectiveness-Equality Trade-Offs: A Data Envelopment Analysis Approach. *Management Science*. 41(7): 1172-84.
- Harris, John J. III, and Donna Y. Ford. (1991). Identifying and Nurturing the Promise of Gifted Black American Children. *The Journal of Negro Education*. 60(1): 3-18.
- Hanushek, Eric A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*. 19(2): 141-164.
- Humphrey, Tom. (2002). Campaign rhetoric flies between dueling political camps; Squabbles include education, debates *Knoxville News-Sentinel*. August 26, 2002. Monday Final Edition. Page B5.
- Jones, William J. (1978). *Building Public Confidence for Your Schools*. Arlington, Virginia: National School Public Relations Association.
- Lomax, Richard G., Mary Maxwell West, Maryellen C. Harmon, Katherine A. Viator, and George F. Madaus. (1995). The Impact of Mandated Standardized Testing on Minority Students. *The Journal of Negro Education*. 64(2): 171-185.
- Pedraja Chaparro, F., J. Salinas Jimenez, Peter Smith (1997). On the Role of Weight Restrictions in Data Envelopment Analysis. *Journal of Productivity Analysis* 8(2): 215-30.
- Rouse, Paul, Martin Putterill, and David Ryan. (1997). Towards a General Managerial Framework for Performance Measurement: A Comprehensive Highway Maintenance Application. *Journal of Productivity Analysis*. 8(2): 127-49.
- Swanson, Christopher B., and David Lee Stevenson. (2002). Standards-Based Reform in Practice: Evidence on State Policy and Classroom Instruction from the NAEP State Assessments. *Educational Evaluation and Policy Analysis*. 24(1): 1-27.
- Tennessee Department of Education. (2002). *Twenty-first Century Report Card*. <http://www.state.tn.us/education/mstat.htm>
- Tennessee State Board of Education. (2000). *Performance Model* (April 2000) <http://www.state.tn.us/sbe/performance1countabilitymodel.htm>
- United States Department of Education. (2003). No Child Left Behind. Accessed June 20, 2003 <http://www.nclb.gov/start/facts/testing.html>.



FIGURES 2 OF 3

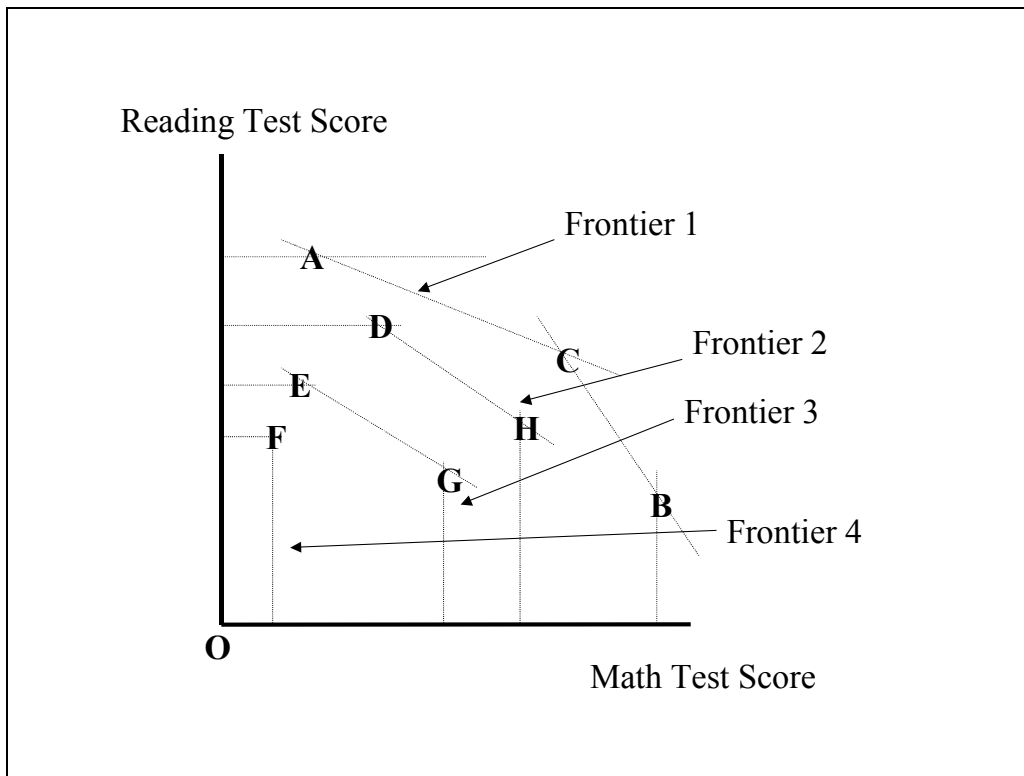


Figure 2: *Wrapping Successive Frontiers: Eight Schools, Two Performance Dimensions*

FIGURES 3 OF 3

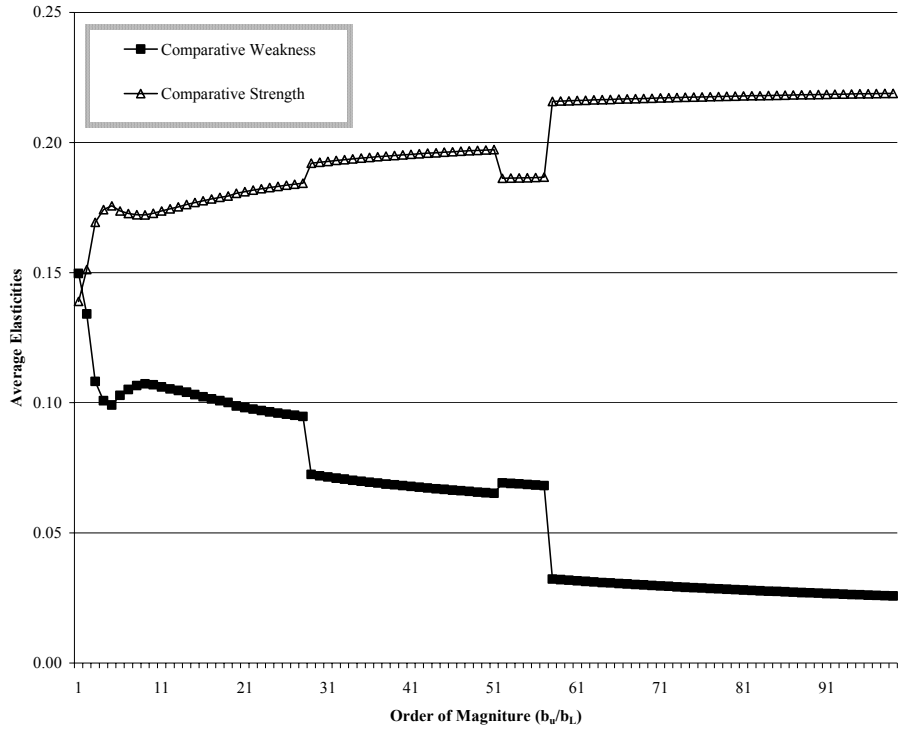


Figure 3: Average Elasticities ( $\xi^r$ ) for School Comparative Strengths and Comparative Weaknesses, as Weight Constraints (2.c) and (2.d) Relaxed.

**TABLES 1 OF 3**

**Table 1: Seven Performance Measures for Nine K-4 Schools**

School Name	Value Added between Grades Three and Four					Grades K-4 Promotion	4 <sup>th</sup> Grade Writing Attainment
	Language	Mathematics	Reading	Science	Social Studies		
Charlotte Park	115.084	100.229	105.977	117.735	116.380	90.741	95.210
Cockrill	98.844	93.310	93.970	100.978	90.709	93.876	81.625
Cole	97.347	115.319	95.882	104.248	99.112	106.372	88.418
Crieve Hall	94.952	91.222	97.718	90.352	86.160	113.551	122.381
Gateway	109.733	100.139	119.857	102.715	113.450	107.568	105.399
Goodlettsville	74.521	97.366	84.181	89.194	90.632	100.603	98.607
Gower	104.606	94.622	107.201	102.545	100.268	107.724	95.210
Granbery	101.089	96.650	111.827	97.300	102.928	98.914	112.192
H.G.Hill	102.212	110.906	114.848	111.127	118.615	103.377	88.418

*Notes:* All seven performance dimensions are standardized scores—where the district-wide mean equals 100 and the district-wide standard deviation equals 10. School year is 1999-2000. Valued added scores are three year means.

TABLES 2 OF 3

**Table 2: Calculated Weights for Seven Performance Measures**

School Name	Value Added between Grades Three and Four					Grades K-4 Promotion	4 <sup>th</sup> Grade Writing Attainment
	Language	Mathematics	Reading	Science	Social Studies		
Charlotte Park	0.001540	0.001564	0.000836	0.001506	0.001523	0.001381	0.000931
Cockrill	0.001709	0.001811	0.000899	0.001673	0.000931	0.001273	0.001035
Cole	0.001028	0.001735	0.001044	0.001920	0.001010	0.001330	0.001267
Crieve Hall	0.001171	0.001219	0.001138	0.001230	0.001290	0.001385	0.001817
Gateway	0.001083	0.001187	0.000992	0.001157	0.001048	0.001563	0.002255
Goodlettsville	0.001104	0.001690	0.000977	0.001844	0.000908	0.001156	0.001668
Gower	0.001926	0.001065	0.000940	0.001964	0.001005	0.001322	0.001058
Granbery	0.001009	0.002111	0.000912	0.001048	0.000991	0.001458	0.001818
H.G.Hill	0.001073	0.001978	0.000955	0.001681	0.000925	0.001501	0.001241
Maximum/Minimum	1.909	1.983	1.360	1.874	1.678	1.351	2.423

*Notes:* Weights are restricted so that  $b_U = 2^{1/2}$  and  $b_L = 2^{-1/2}$  (equations 2.c and 2.d). Promotion serves as the numeraire.

TABLES 3 OF 3

**Table 3: Calculated TDEA Frontier and Distance Index**

School Name	TDEA frontier	$\theta_i$ (equations 2.a-2.e)	$\tilde{\theta}_i$ (equation 3)	$\theta_i - \tilde{\theta}_i$
Gateway	1	1.00000	1.00000	0
H.G.Hill	1	1.00000	0.98767	0.012334
Charlotte Park	2	0.99112	0.97693	0.014191
Granbery	2	0.96022	0.94997	0.010247
Cole	3	0.95395	0.93126	0.022691
Gower	3	0.94821	0.93848	0.009732
Crieve Hall	4	0.93536	0.91761	0.017755
Cockrill	5	0.87965	0.86091	0.018741
Goodlettsville	6	0.85656	0.83691	0.019645

*Notes:* Schools are sorted with the most effective at the top. Lower TDEA frontier numbers correspond to more effective schools.