

Chapter 19

Performance Benchmarking of School Districts in New York State

Thomas R. Sexton, Christie Comunale, Michael Shane Higuera
and Kelly Stickle

Abstract We used DEA to measure the performance of New York State school districts and provide alternative improvement targets for each district. We found that 201 of the 624 (32.2%) districts with one or more high schools and 28 of the 31 (90.3%) districts with no high school were on the performance frontier. We found evidence that NYS could reduce FTE teachers by 8.4%, FTE teacher support by 17.2%, and FTE administration and professional staff by 9.4%. We also found that NYS could increase percentage of students who pass the English exam by 4.9 percentage points, the mathematics exam by 5.0 percentage points, and the science exam by 5.8 percentage points, while increasing the average graduation rate by 5.4 percentage points.

Keywords School districts · Benchmarking · Data envelopment analysis · New York State

19.1 Introduction

In 2011, New York State's 695 school districts (New York State Education Department n.d.) spent \$53.7 billion (U.S. Census Bureau 2011, Table 6) to educate almost 2.7 million elementary and secondary pupils (U.S. Census Bureau, 2011, Table 19), a cost of over \$19,000 per pupil (U.S. Census Bureau 2011, Table 8). Elementary and secondary education accounts for nearly one-quarter of all state and local expenditures in New York State (U.S. Government Spending n.d.). While New York State has some excellent school districts, others struggle with poor standardized test scores and low graduation rates. Many of the reasons for the differences among school districts are widely accepted. These include differences in wealth, English proficiency, and inefficient use of resources.

T.R. Sexton (✉) · C. Comunale · M.S. Higuera · K. Stickle
College of Business, Stony Brook University, 317 Harriman Hall,
Stony Brook, NY 11794-3775, USA
e-mail: Thomas.Sexton@stonybrook.edu

Given the high cost of public education and its critical importance for the future of New York and the nation, it is natural for taxpayers, legislators, and administration officials to hold public education institutions accountable for producing high quality outcomes. To do so, we must measure the performance of each school district in an objective, data-informed manner. Commonly used methods for performance measurement under these circumstances are often called *benchmarking models*. When applied to school districts, a benchmark model identifies leading school districts, called *benchmark school districts*, and it facilitates the comparison of other school districts to the benchmark school districts. Nonbenchmark school districts can focus on specific ways to improve their performance and thereby that of the overall statewide school system.

In this paper, we utilize Data Envelopment Analysis (DEA) to measure the performance of New York State school districts in the 2011–2012 academic year, and provide detailed alternative improvement pathways for each school district.

19.2 Literature Review

DEA has been used since the 1950s in a wide variety of applications, including health care, banking, pupil transportation, and most recently, education. DEA's mathematical development may be traced to Charnes et al. (1978), who built on the work of Farrell (1957) and others. The technique is well documented in the management science literature (Charnes et al. 1978, 1979, 1981; Sexton 1986; Sexton et al. 1986; Cooper et al. 1999), and it has received increasing attention as researchers have wrestled with problems of productivity measurement in the services and nonmarket sectors of the economy. Emrouznejad et al. (2008) provided a review of more than 4000 DEA articles. See Emrouznejad (2014) for an extensive bibliography of DEA publications as well as a DEA tutorial and DEA software.

We are not the first to apply DEA to school districts. Färe et al. (1989) applied DEA to evaluate the performance of a sample of Missouri school districts for the 1985–1986 school year. Kirjavainen and Loikkanen (1998) studied efficiency differences among Finnish senior secondary schools. They found that schools with small classes and heterogeneous student bodies were inefficient whereas school size did not affect efficiency. Surprisingly, private schools were inefficient relative to public schools. Kang and Greene (2002) used DEA to evaluate the impacts of institutional arrangements on various measures of high school output. Driscoll et al. (2003) used statistical methods to estimate a production function for California schools, which showed that smaller districts, smaller schools, and smaller class sizes were associated with higher academic achievement scores. Ruggiero (2007) used DEA to examine efficiency, costs and adequacy of 607 Ohio school districts using school year 2000 data. The results indicate that adequacy standards can be met by improving the performance of inefficient school districts and reallocating existing resources without increasing total expenditures. Thanassoulis et al. (2016)

review applications of DEA in secondary and tertiary education, focusing on the opportunities that this offers for benchmarking at institutional level.

19.3 Data Envelopment Analysis

DEA has proven to be a successful tool in performance benchmarking. It is particularly well suited when measuring the performance of units along multiple dimensions, as is the case with complex organizations such as school districts. DEA empirically identifies the best performers by forming the performance frontier based on observed indicators from all units. Consequently, DEA bases the resulting performance scores and potential performance improvements entirely on the actual performance of other DMUs, free of any questionable assumptions regarding the mathematical form of the underlying production function. On balance, many analysts view DEA as preferable to other forms of performance measurement.

Figures 19.1 and 19.2 illustrate the performance frontier for a simple model of school districts. We can use this simple model, which is clearly inadequate for capturing the complexity of school districts, to demonstrate the fundamental concepts of DEA. In this model, we assume that each school district employs only one type of resource, full-time equivalent (FTE) teachers, and prepares students for only one type of standardized test, mathematics at the appropriate grade level, measured

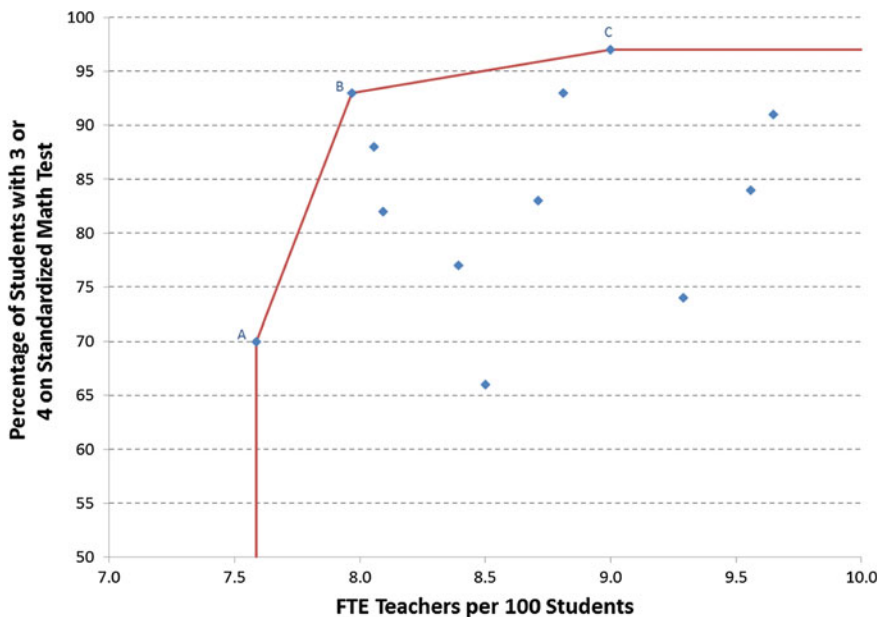


Fig. 19.1 The performance frontier for a simple example

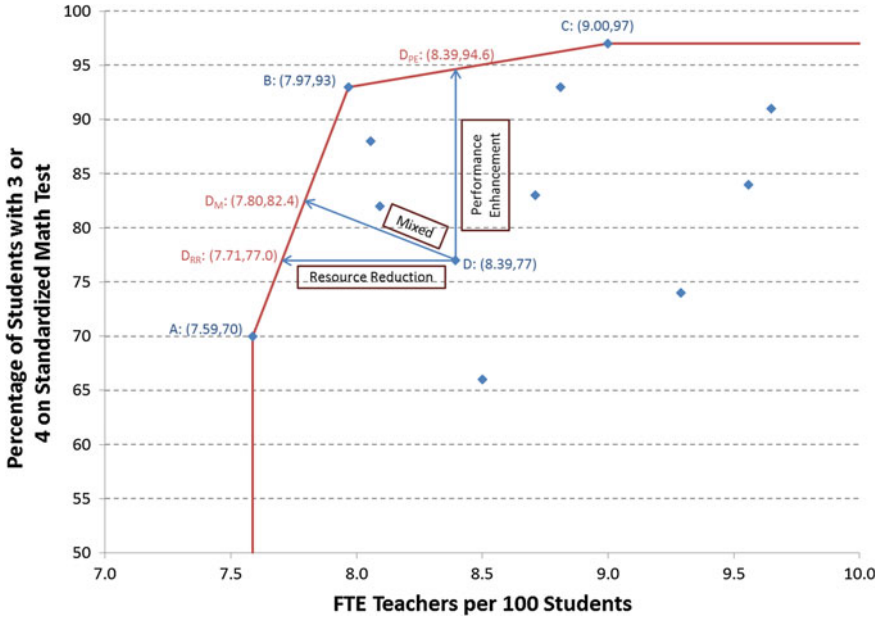


Fig. 19.2 Several ways for school district D to move to the performance frontier

as the percentage of students who score at a given level or higher. Each school district is represented by a point in the scatterplot.

In Fig. 19.1, school districts A, B, and C define the performance frontier. In each case, there is no school district or weighted average of school districts that has fewer FTE teachers per 100 students and has a higher percentage of students who scored 3 or 4 on the standardized mathematics test. Such school districts, if they existed, would lie to the Northwest of A, B, or C, and no such districts, or straight lines between any two districts, exists.

School district D, in Fig. 19.2, does not lie on the performance frontier and therefore its performance can improve. In principle, D can choose to move anywhere on the performance frontier. If school district D chooses to focus on resource reduction without test performance change, it would move to the left, reaching the performance frontier at point D_{RR} . This move would require a reduction from 8.39 to 7.71 FTE teachers per 100 students. If school district D enrolls 10,000 students, this reduction would be from 839 to 771 teachers, a percentage reduction of 8.1%. We refer to this strategy as the *resource reduction orientation*.

If school district D chooses to focus on performance enhancement without resource reduction, it would move upward, reaching the performance frontier at point D_{PE} . This move would require 94.6% of its students to score 3 or 4 on the standardized mathematics test, up from 77%. If 1000 students in school district D sat for the standardized mathematics test, students scoring 3 or 4 would increase would from 770 to 946, or by 22.9%. We refer to this strategy as the *performance enhancement orientation*.

School district D might prefer an intermediate approach that includes both resource reduction and performance enhancement and move to point D_M . This entails both a reduction in FTE teachers per 100 students from 8.39 to 7.80 and an increase in the percentage of students who score 3 or 4 on the standardized mathematics test from 77 to 82.4%. If school district D enrolls 10,000 students, this reduction would be from 839 to 780 teachers, or by 7.0%, and an increase in students scoring 3 or 4 from 770 to 824, or 7.0%. We refer to this strategy as the *mixed orientation*. The mixed orientation has the feature that the percentage decrease in each resource equals the percentage increase in each performance measure.

The three points D_{RR} , D_{PE} , and D_M are called *targets* for school district D because they represent three possible goals for D to achieve to reach the performance frontier. School district D can choose its target anywhere on the performance frontier, but these three points represent reasonable reference points for D as it improves its overall performance.

Of course, this model does not consider other resources used by school districts such as teacher support personnel and other staff, nor does it consider standardized test scores in science or English. It also ignores graduation rates in school districts with one or more high schools. Moreover, it does not recognize differences in important district characteristics such as the number of elementary and secondary students, the percentage of students who qualify for free or reduced price lunch or who have limited English proficiency, or the district's combined wealth ratio.

When other measures are included in the model, we can no longer rely on a simple graphical method to identify a school district's target school district. For this purpose, we rely on the linear programming model that we describe in detail in Technical Appendix. Nonetheless, the target school district will have the same basic interpretation. Relative to the school district in question, the target school district consumes the same or less of each resource, its students perform the same or better on each standardized test, its graduation rate is at least as high (if applicable), it educates the same number or more students, and it operates under the same or worse district characteristics.

19.4 A DEA Model for School District Performance in New York State

To apply the DEA methodology to measure the performance of New York State school districts, we began by identifying three categories of important school district measurements. They were:

- resources consumed;
- performance measures; and
- district characteristics

We defined the resources consumed as:

- FTE teachers;
- FTE teacher support (teacher assistants + teacher aides); and
- building administration and professional staff (principals + assistant principals + other professional staff + paraprofessionals).

For school districts with no high school, we defined the performance measures as:

- percentage of students scoring at or above level 3 on ELA grade 6;
- percentage of students scoring at or above level 3 on math grade 6; and
- percentage of students scoring at or above level 3 on science grade 4.

For school districts with one or more high schools, we defined the performance measures as:

- total cohort results in secondary level English after 4 years of instruction: percentage scoring at levels 3–4;
- total cohort results in secondary level math after 4 years of instruction: percentage scoring at levels 3–4;
- grade 8 science: percentage scoring at levels 3–4 all students; and
- 4-year graduation rate as of August.

We defined the district characteristics as:

- number of elementary school students;
- number of secondary school students;
- percentage of students with free or reduced price lunch;
- percentage of students with limited English proficiency; and
- school district's combined wealth ratio.

We recognize that other choices of variables are possible. We use this particular set of variables because it captures a reasonable range of resources consumed, performance dimensions to be measured, and district characteristics to be taken into account. Other variables may be added if statewide data are available for every school district. Our objective is to illustrate the model and its ability to provide school districts with useful feedback for strategic planning and other purposes.

Our goal is to guide school district managers to the performance frontier and we recognize that there are infinitely many points on the performance frontier. We also believe that school district managers are in the best position to decide which direction to take toward the performance frontier, and that different school districts may choose different strategies based on their own circumstances.

Toward that end, we consider three possible orientations, or directions to the performance frontier, for each school district. The resource reduction orientation seeks to reduce resource consumption as much as possible while maintaining performance measures at their current levels. The performance enhancement orientation seeks to improve performance measures as much as possible while

maintaining resource consumption at current levels. The mixed orientation seeks to improve performance measures and reduce resource consumption simultaneously in a balanced way.

We present the results of all three orientations to provide school district administrators with alternative options for reaching the performance frontier. One district might elect to focus on resource reduction; another might opt for increases in test scores and graduation rate, while a third might prefer a blended strategy that combines these two objectives. Since there are infinitely many points on the performance frontier toward which a district may move, the three that we present are designed to highlight three possible alternatives.

We point out that the performance frontier is unaffected by the choice of orientation. Any district that lies on the performance frontier in one orientation will also lie on it in any other orientation. Orientation only determines the location of the target district on the performance frontier.

19.5 Data and Results

We obtained complete data for 624 public school districts with one or more high schools and 31 public school districts with no high school for the academic year 2011–2012. Complete data were unavailable for certain districts. All data were obtained from the New York State Education Department.

19.6 Results for Three Example Districts

Table 19.1 shows the results for three districts based on the model described above. These districts were selected to illustrate the manner in which the model results can be presented to school districts and how they might be interpreted.

School district A would reduce all three resources by 18.3% using the resource reduction orientation and by 4.0% under the mixed orientation, but would not reduce any resources under the performance enhancement orientation. Improvements in English and science would be virtually the same using all three orientations (in the range of 4%) but the improvements in math and graduation rate are notably higher using either the performance enhancement or mixed orientations. The message for school district A is that it can raise all three test measures by about 4% and graduation rate by about 8% with little or no reduction in resources. Alternatively, it can improve English and science (but not math) by about 4% and graduation rate by 4–5% even with significant resource reductions. The choice of strategy would be influenced by many other factors not reflected in the model.

School district B can reduce its FTE teachers by at least 6.9% but its greater opportunity lies in teacher support, which it can reduce by at least 27.4%. Despite these reductions, it can improve English by almost 7% and math by almost 4%.

Table 19.1 Results for three example districts under three orientations (in percentages of current actual values)

Dist	Orientation	FTE teachers	FTE teacher support	Bld adm and prof staff	Secondary level English (%)	Secondary level math (%)	Grade 8 science (%)	Grad rate (%)
A	Res red	81.7	81.7	81.7	103.9	100.0	103.8	104.6
	Perf enhan	100.0	100.0	100.0	104.3	104.3	104.3	108.5
	Mixed	96.0	96.0	96.0	104.3	104.0	104.0	108.2
B	Res red	90.2	65.8	90.2	105.3	101.5	100.0	100.0
	Perf enhan	93.1	72.6	100.0	106.8	103.8	101.8	101.8
	Mixed	92.8	72.6	98.4	106.7	103.6	101.6	101.6
C	Res red	99.7	99.7	99.7	101.1	113.8	100.0	100.9
	Perf enhan	100.0	100.0	100.0	101.1	113.8	100.1	101.1
	Mixed	99.9	99.9	99.9	101.1	113.8	100.1	101.0

The values in bold draw attention to specific areas for improvement

School district C is performing very well regardless of orientation with the exception of math, which it can improve by almost 14%.

19.7 Statewide Results

We found no evidence that 201 of the 624 (32.2%) school districts with one or more high schools can reduce resource consumption or improve performance. The same statement applies to 28 of the 31 (90.3%) school districts with no high school. Put another way, each of these school districts serves as its own target school district. Based on the observed performance of all school districts in New York State in the academic year 2011–2012, none of these school districts can simultaneously reduce *each* of its resources and improve *each* of its performance measures while operating under the same district characteristics.

It is important to recognize that DEA is an empirical method that measures a school district's performance *relative to the performances of other school districts*. It makes no theoretical assumptions about the shape of the production possibility frontier. Thus, it may in fact be possible for school districts on the frontier to make improvements but current data provide no evidence to suggest this possibility or indicate the extent of such improvements.

19.8 Districts With One or More High Schools

The 624 school districts with one or more high schools employed 126,470 FTE teachers, 33,035 FTE teacher support personnel, and 25,492.5 FTE building administration and professional staff in the academic year 2011–2012. The average

Table 19.2 Data and statewide results for all three orientations for school districts with one or more high schools

	FTE teachers	FTE teacher support	Building admin and prof staff	Secondary level English (%)	Secondary level math (%)	Grade 8 science (%)	Grad rate (%)
Actual	126,470	33,035	25,493	84.4	86.0	81.6	84.2
Mixed orientation							
Target	115,812	27,359	23,091	89.3	91.0	87.4	89.6
Change	10,658	5676	2402	4.9	5.0	5.8	5.4
% Change	8.4	17.2	9.4	5.8	5.8	7.1	6.4
Resource reduction orientation							
Target	<i>102,314</i>	<i>25,653</i>	<i>20,567</i>	86.7	88.4	85.3	86.5
Change	<i>24,156</i>	<i>7382</i>	<i>4925</i>	2.2	2.4	3.7	2.3
% Change	<i>19.1</i>	<i>22.3</i>	<i>19.3</i>	2.6	2.8	4.5	2.7
Performance enhancement orientation							
Target	119,311	27,913	23,687	89.7	<i>91.3</i>	<i>87.6</i>	<i>89.9</i>
Change	7159	5122	1805	5.3	<i>5.3</i>	<i>6.0</i>	<i>5.7</i>
% Change	5.7	15.5	7.1	6.3	<i>6.1</i>	<i>7.3</i>	<i>6.8</i>

The italicized values represent values in the direction of the model’s orientation

percentage of students who scored 3 or 4 on the English exam was 84.4%; on the mathematics exam, the average was 86.0%, and on the science exam, the average was 81.6%. The average graduation rate was 84.2%. See Table 19.2.

Using a mixed orientation, we found evidence that the number of FTE teachers can be reduced by 8.4%, the number of FTE teacher support personnel can be reduced by 17.2%, and the number of FTE building administration and professional staff personnel can be reduced by 9.4%. In addition, that the average¹ percentage of students who score 3 or 4 on the English exam can rise by 4.9 percentage points, by 5.0 percentage points on the mathematics exam, and by 5.8 percentage points on the science exam. Moreover, the average² graduation rate can rise by 5.4 percentage points.

Using a resource reduction orientation, we found evidence that the number of FTE teachers can be reduced by 19.1%, the number of FTE teacher support personnel can be reduced by 22.3%, and the number of FTE building administration and professional staff personnel can be reduced by 19.3%. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 2.2 percentage points, by 2.4 percentage points on the mathematics exam, and by 3.7 percentage points on the science exam. Moreover, the average graduation rate can

¹These are unweighted averages and therefore they do not represent the statewide percentages.

²See previous footnote.

rise by 2.3 percentage points. We point out that, even though we have used a resource reduction orientation, we can still identify potential improvements in the performance measures. This is because one or more of the performance measure constraints may not be binding at optimality; the corresponding slacks represent potential improvements known as nonradial improvements.

Finally, using a performance enhancement orientation, we found evidence that the number of FTE teachers can be reduced by 5.7%, the number of FTE teacher support personnel by 15.5%, and the number of FTE building administration and professional staff personnel by 7.1%. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 5.3 percentage points, by 5.3 percentage points on the mathematics exam, and by 6.0 percentage points on the science exam. Moreover, the average graduation rate can rise by 6.8 percentage points. Once again, even though we have used a performance enhancement orientation, we can still identify potential reductions in the resource measures. This is because one or more of the resource constraints may not be binding at optimality; the corresponding slacks represent potential nonradial reductions in resource consumption.

Figures 19.3, 19.4, and 19.5 illustrate the potential improvements in the three resource categories. For districts that lie on the diagonal of one of these graphs, there is no evidence that they could reduce their use of this resource category. Other districts have the potential to reduce resource consumption by the amount that they lay below the diagonal.

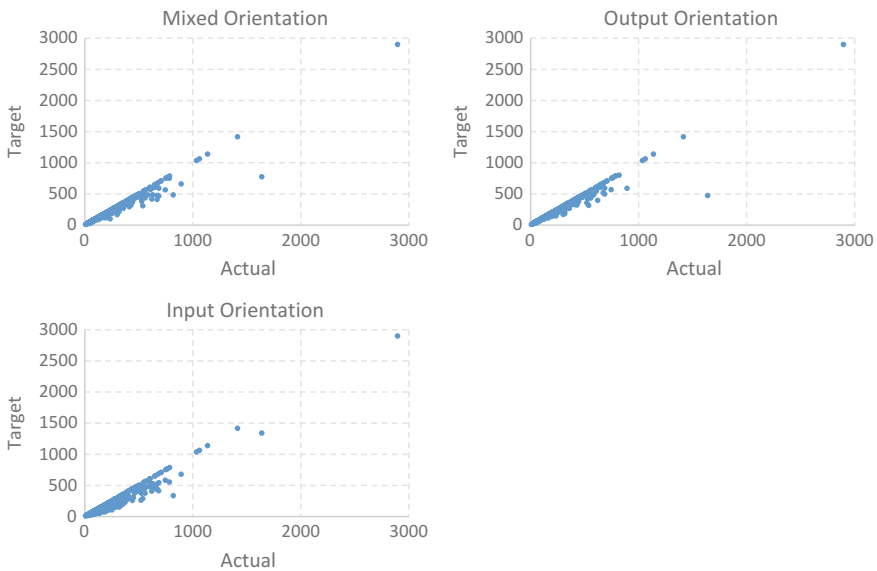


Fig. 19.3 Target versus actual FTE teachers under each of the three orientations for school districts with at least one high school

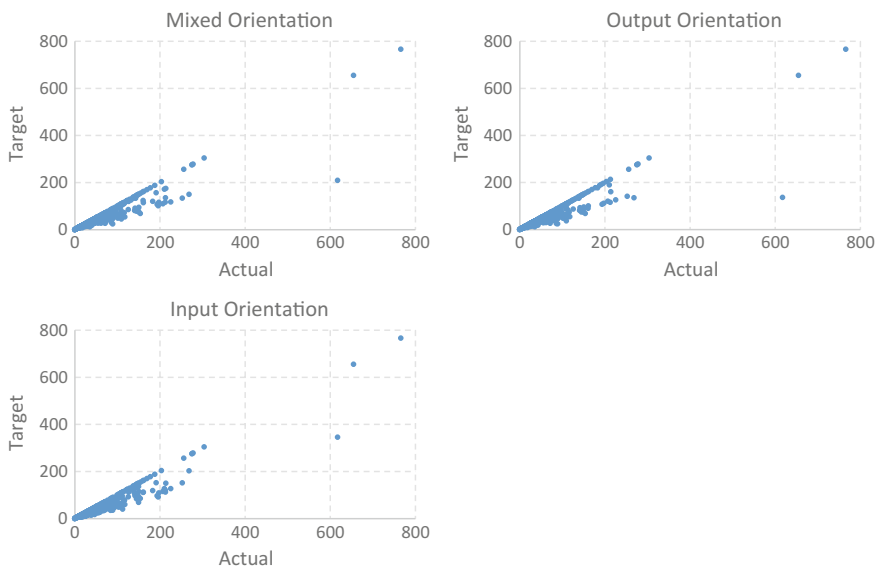


Fig. 19.4 Target versus actual FTE teacher support under each of the three orientations for school districts with at least one high school

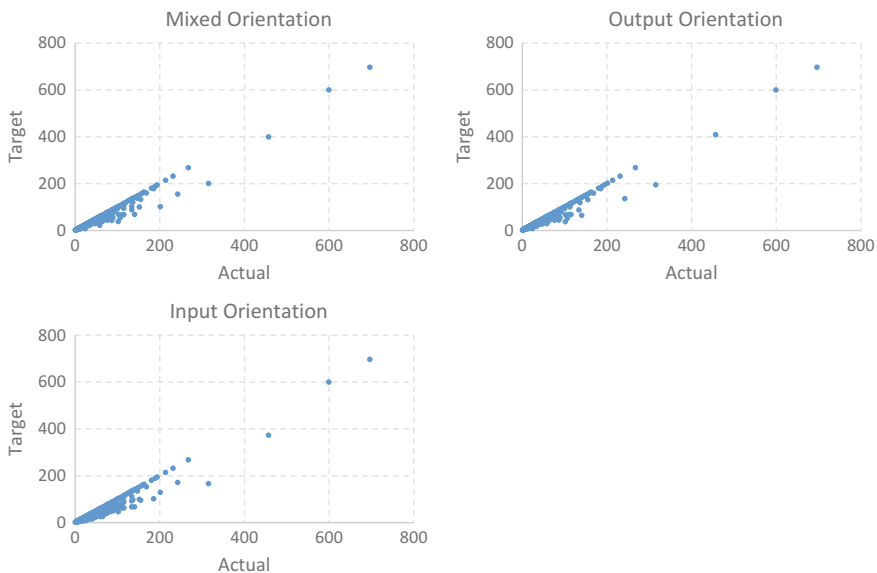


Fig. 19.5 Target versus actual FTE building and administrative professional staff under each of the three orientations for school districts with at least one high school

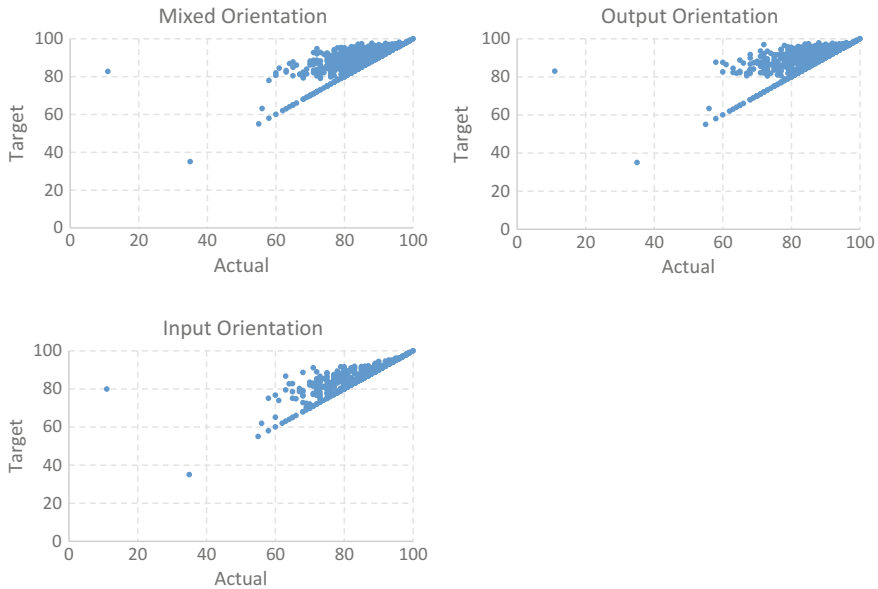


Fig. 19.6 Target versus actual percentage of students scoring 3 or 4 on the secondary level English standardized test under each of the three orientations for school districts with at least one high school

Figures 19.6, 19.7, 19.8 and 19.9 illustrate the potential improvements in the four performance measures. For districts that lie on the diagonal of one of these graphs, there is no evidence that they could improve their performance in this dimension. Other districts have the potential to improve by the amount that they lay above the diagonal.

Figure 19.10 shows the histograms of the school districts for each of the three factor performances associated with the resources, excluding those districts for which no improvement is possible. Figure 19.11 shows the histograms of the school districts for each of the four factor performances associated with the performance measures, again excluding those for which no improvement is possible.

19.9 Districts Without a High School

The 31 school districts with no high school employed 2233 FTE teachers, 762 FTE teacher support personnel, and 416 FTE building administration and professional staff in the academic year 2011–2012. The average percentage of students who scored 3 or 4 on the English exam was 84.4%; on the mathematics exam, the average was 86.0%, and on the science exam, the average was 81.6%. See Table 19.3.

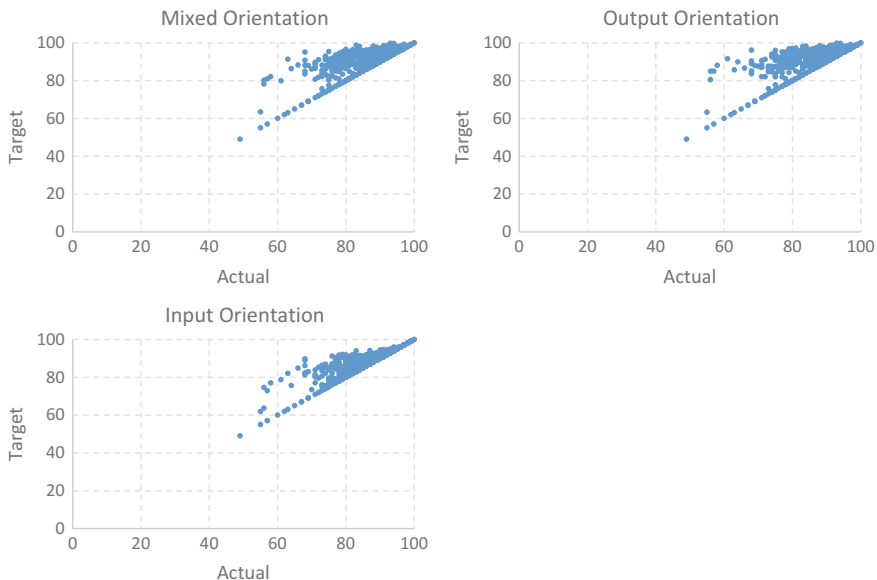


Fig. 19.7 Target versus actual percentage of students scoring 3 or 4 on the secondary level mathematics standardized test under each of the three orientations for school districts with at least one high school

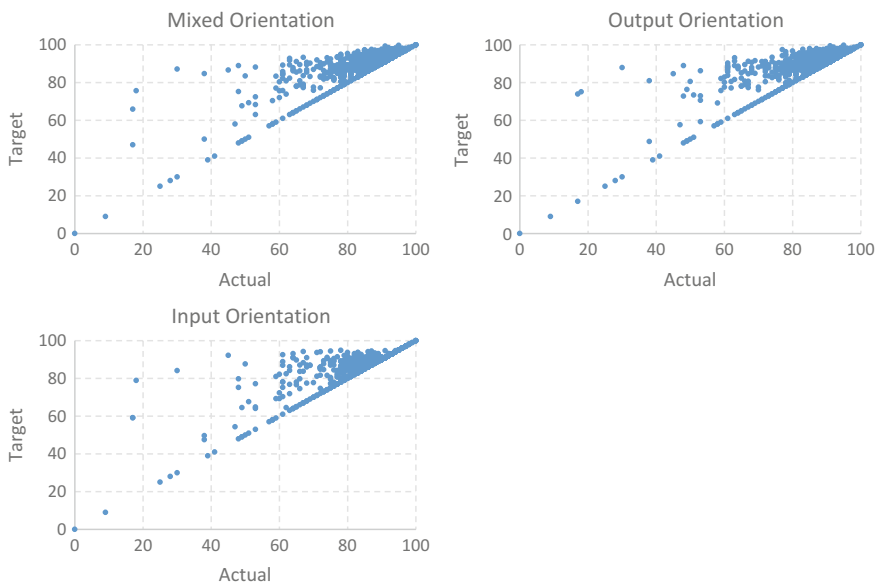


Fig. 19.8 Target versus actual percentage of students scoring 3 or 4 on the grade 8 science standardized test under each of the three orientations for school districts with at least one high school

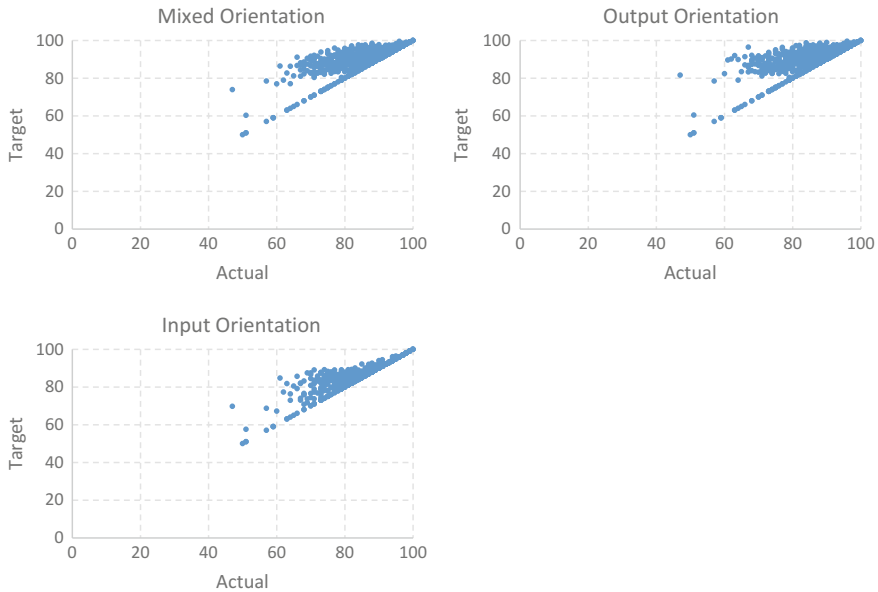


Fig. 19.9 Target versus actual percentage of 4-year graduation rate under each of the three orientations for school districts with at least one high school

Using a mixed orientation, we found evidence that the number of FTE teachers can be reduced by 0.2%, the number of FTE teacher support personnel by 4.3%, and the number of FTE building administration and professional staff personnel by 3.3%. In addition, the average³ percentage of students who score 3 or 4 on the English exam can rise by 0.4 percentage points, by 0.9 percentage points on the mathematics exam, and by 0.3 percentage points on the science exam.

Using a resource reduction orientation, we found evidence that the number of FTE teachers can be reduced by 0.8%, the number of FTE teacher support personnel by 4.6%, and the number of FTE building administration and professional staff personnel by 4.8%. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 0.6 percentage points, by 0.6 percentage points on the mathematics exam, and by 0.0 percentage points on the science exam.

Finally, using a performance enhancement orientation, we found evidence that the number of FTE teachers can be reduced by 0.0%, the number of FTE teacher support personnel by 4.3%, and the number of FTE building administration and professional staff personnel by 3.0%. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 0.4 percentage points, by 0.9 percentage points on the mathematics exam, and by 0.3 percentage points on the science exam.

³These are unweighted averages and therefore they do not represent the statewide percentages.

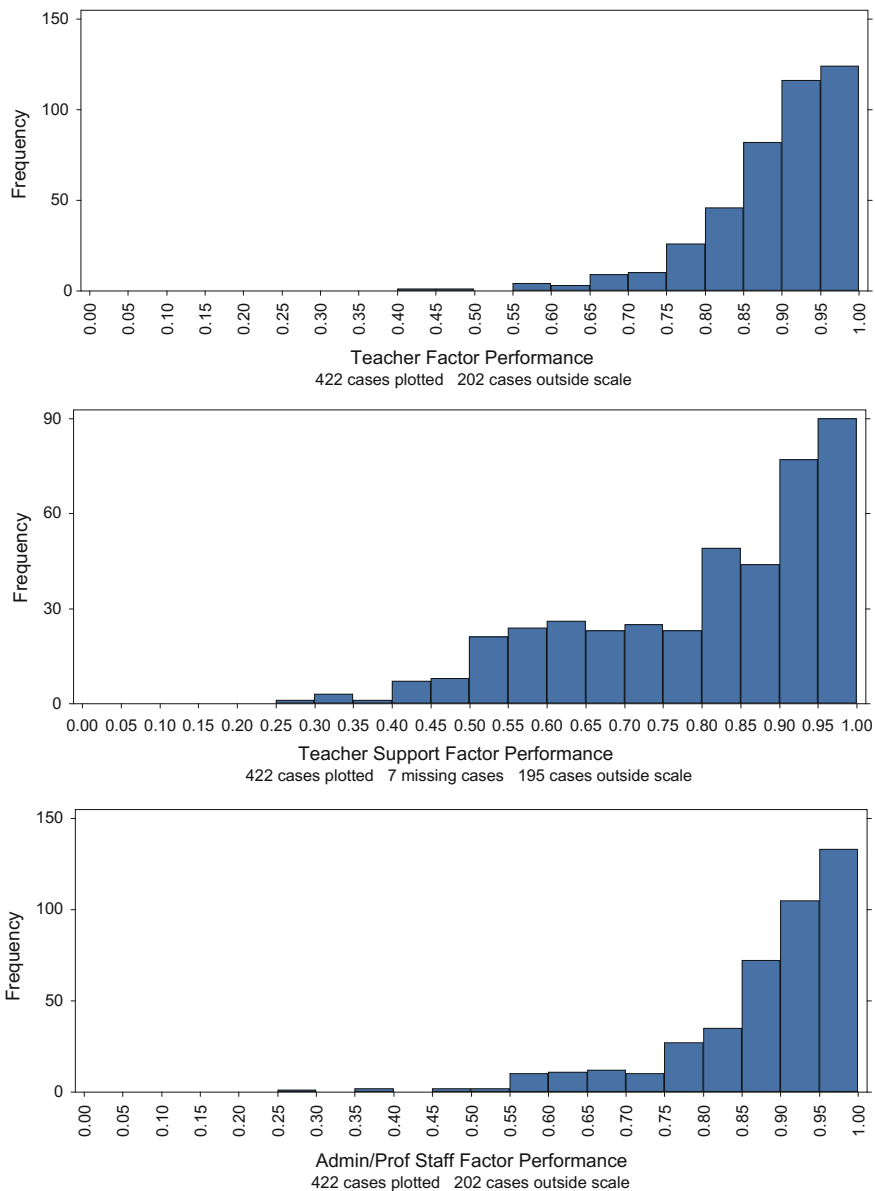


Fig. 19.10 Histograms of the school districts with at least one high school for each of the three factor performances associated with the resources, excluding those districts for which no improvement is possible

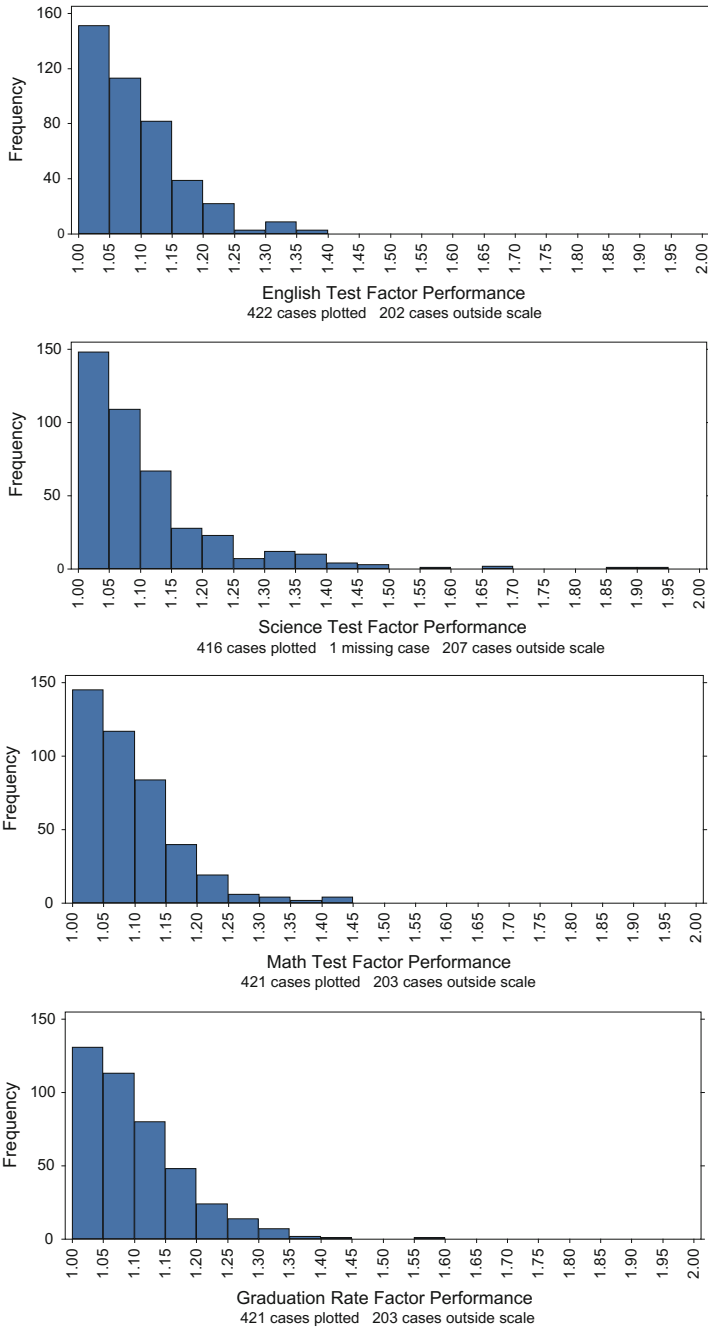


Fig. 19.11 Histograms of the school districts with at least one high school for each of the four factor performances associated with the performance measures, excluding those for which no improvement is possible

Table 19.3 Statewide results for all three orientations for School Districts without a high school

	FTE teachers	FTE teacher support	Building admin and prof staff	Grade 6 ELA (%)	Grade 6 math (%)	Grade 4 science (%)
Actual	2233	762	417	77.7	83.1	94.6
Mixed orientation						
Target	2228	729	403	78.0	83.8	94.8
Change	5	33	14	0.3	0.7	0.3
% Change	0.2	4.3	3.3	0.4	0.9	0.3
Resource reduction orientation						
Target	<i>2216</i>	<i>727</i>	<i>397</i>	<i>78.2</i>	<i>83.6</i>	<i>94.6</i>
Change	<i>17</i>	<i>35</i>	<i>20</i>	<i>0.5</i>	<i>0.5</i>	<i>0.0</i>
% Change	<i>0.8</i>	<i>4.6</i>	<i>4.8</i>	<i>0.6</i>	<i>0.6</i>	<i>0.0</i>
Performance enhancement orientation						
Target	2233	729	404	<i>78.1</i>	<i>83.9</i>	<i>94.9</i>
Change	–	33	13	<i>0.3</i>	<i>0.7</i>	<i>0.3</i>
% Change	0.0	4.3	3.0	<i>0.4</i>	<i>0.9</i>	<i>0.3</i>

The italicized values represent values in the direction of the model’s orientation

19.10 Implementation

We reiterate that other choices of variables are possible. An important first step is for the school districts and the New York State Education Department (NYSED) to work together to modify this model as necessary. For example, the current model does not include data on Regents exam scores. In principle, the only requirement is that complete data exists for all school districts for the specified school year. In addition, it is important to provide a complete data set so that all school districts, especially those in New York City, can be included. This data set needs to be compiled for the latest school year for which complete data are available.

The NYSED would need to determine the distribution of model results. Perhaps the initial distribution during a pilot phase should be restricted to the school districts and NYSED. This would allow school districts the opportunity to understand the full meaning of their own results better and to begin to incorporate the results into their operations and planning. The pilot phase would also allow school districts and NYSED to suggest further improvements in the model.

Ultimately, the model can serve as a key element in a quality improvement cycle. By providing direct feedback to each school district about its performance along multiple dimensions, it supports school district decisions about how to improve and allows them to demonstrate that their decisions have in fact had the desirable effects.

19.11 Intended Use of the Model

It is axiomatic that public school systems, being funded with public money, must be held accountable for the efficient use of their funding. They must use their available resources to provide the highest levels of learning possible given their myriad regulatory, faculty, student, and societal constraints—in effect, public school systems must solve a highly complex constrained optimization problem.

State and school system policymakers have not been able to hold school systems accountable for their performance in a way that precludes reasonable-sounding arguments about why any specific school district cannot perform at the performance level sought by policymakers. It has been too easy for school districts to claim that the target performance levels were set arbitrarily by unknowing policymakers, and that their school district has unique constraints that preclude it from meeting these arbitrary target performance levels. Of course, these claims could be true; we are not taking sides on this issue. We are only trying to solve it.

The model presented in this paper is designed and intended for the use of policymakers to effectively eliminate such claims. The model empowers policymakers to demonstrate that any school district not performing at the frontier level established by the model is underperforming. The school district(s) performing at frontier levels are doing so under conditions that are the same or worse than those school districts that are not performing at frontier levels.

Through the use of our model, policymakers will be empowered to establish data-informed performance levels for accountability, and school districts will not be able to avoid accountability by claiming uniqueness. School districts will then be forced to improve their efficiency or be identified openly and accurately as underperforming, in which case, the responsibility for underperforming will sit squarely with school district decision-makers. As those schools performing at frontier levels continue to innovate and improve, the frontier will continue to move outward, thereby eliminating any possibility of complacency within the accountability system by introducing healthy competition to remain on the frontier.

Our data-informed accountability model will incentivize the leaders of school districts not performing at frontier levels to improve their optimization of learning efforts or suffer being replaced with more effective leaders. This incentive has been missing within public school systems for decades and it will be transformative of public education if implemented.

19.12 Conclusions

We have presented a flexible model that allows school districts and NYSED to measure school district performance throughout New York State. The model provides multiple, mathematically derived performance measures that allow school districts to detect specific areas for improvement. The model also enables NYSED

to identify school districts that are the top performers in the state and others that most require improvement.

The results of a preliminary version of the model applied to data from the 2011 to 2012 school year shows that approximately one-third of the school districts in New York State are performing as well as can be expected given their local school district characteristics. Another 26.8–42.3%, depending on the specific resource or performance measure, can improve by no more than 10%.

Nonetheless, substantial statewide improvements are possible. Using the mixed orientation, for example, if every school district was to match to its target, New York State would have between 8 and 17% fewer personnel, 6 to 7% more students scoring 3 or 4 on standardized tests, and 6% more students graduating within 4 years.

Public education is critically important to the future of New York State and the nation. This model offers the potential to support public school education leaders in recognizing where improvements are possible and in taking appropriate action to implement those improvements.

Technical Appendix: The Mathematics of the DEA Model

We use two slightly different DEA models in this paper, one for school districts with one or more high schools, and one for school districts without a high school. The differences lie in the performance measures (different points at which test scores are measured, and no graduation rate for school districts with no high school). In addition, each model is employed with three different orientations (resource reduction, performance enhancement, and mixed). The text that follows describes the model for school districts with one or more high schools.

Let $n = 624$ be the number of school districts to be analyzed. The DEA literature refers to units under analysis as *decision-making units*, or DMUs. Let X_{ij} be amount of resource i consumed by DMU j , for $i = 1, 2, 3$, and $j = 1, 2, \dots, 624$. In particular, let X_{1j} be the FTE teachers in DMU j , let X_{2j} be the FTE teacher support in DMU j , and let X_{3j} be the FTE building administration and professional staff in DMU j .

Let Y_{rj} be performance measure r achieved by DMU j , for $r = 1, 2, 3, 4$ and $j = 1, 2, \dots, 624$. In particular, let Y_{1j} be the percentage of students scoring at levels 3 or 4 in secondary level English after 4 years of instruction in DMU j , let Y_{2j} be the percentage of students scoring at levels 3 or 4 in secondary level math after 4 years of instruction in DMU j , let Y_{3j} be the percentage of students scoring at levels 3 or 4 in Grade 8 Science in DMU j , and let Y_{4j} be the 4-year graduation rate as of August in DMU j , for $j = 1, 2, \dots, 624$.

Let S_{kj} be the value of site characteristic k at DMU j , for $k = 1, 2, 3, 4, 5$ and $j = 1, 2, \dots, 624$. In particular, let S_{1j} be the number of elementary school students in DMU j , let S_{2j} be the number of secondary school students in DMU j , let S_{3j} be the percentage of students with free or reduced price lunch in DMU j , let S_{4j} be the

percentage of students with limited English proficiency in DMU j , and let S_{5j} be the combined wealth ratio in DMU j , for $j = 1, 2, \dots, 624$.

The Resource Reduction DEA Model

The resource reduction DEA model with variable returns to scale, for DMU d , $d = 1, 2, \dots, 624$, is below. We must solve $n = 624$ linear programs to perform the entire DEA.

Min E_d	(1)	
Subject to		
$\sum_{j=1}^n \lambda_j X_{1j} \leq E_d X_{1d}$	(2.1)	FTE teachers
$\sum_{j=1}^n \lambda_j X_{2j} \leq E_d X_{2d}$	(2.2)	FTE teacher support
$\sum_{j=1}^n \lambda_j X_{3j} \leq E_d X_{3d}$	(2.3)	Building administration and professional staff
$\sum_{j=1}^n \lambda_j Y_{1j} \geq Y_{1d}$	(3.1)	Secondary level English (%)
$\sum_{j=1}^n \lambda_j Y_{2j} \geq Y_{2d}$	(3.2)	Secondary level math (%)
$\sum_{j=1}^n \lambda_j Y_{3j} \geq Y_{3d}$	(3.3)	Grade 8 science (%)
$\sum_{j=1}^n \lambda_j Y_{4j} \geq Y_{4d}$	(3.4)	Graduation rate (%)
$\sum_{j=1}^n \lambda_j S_{1j} \geq S_{1d}$	(4.1)	Number of elementary school students
$\sum_{j=1}^n \lambda_j S_{2j} \geq S_{2d}$	(4.2)	Number of secondary school students
$\sum_{j=1}^n \lambda_j S_{3j} \geq S_{3d}$	(4.3)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^n \lambda_j S_{4j} \geq S_{4d}$	(4.4)	Percentage of students with limited English proficiency
$\sum_{j=1}^n \lambda_j S_{5j} \leq S_{5d}$	(4.5)	School district's combined wealth ratio
$\sum_{j=1}^n \lambda_j = 1$	(5)	Variable returns to scale
$\lambda_j \geq 0$ for $j = 1, 2, \dots, 624$	(6)	Nonnegativity
$E_d \geq 0$	(7)	Nonnegativity

We observe that setting $\lambda_d = 1$, $\lambda_j = 0$ for $j \neq d$, and $E_d = 1$ is a feasible, but not necessarily optimal, solution to the linear program for DMU d . This implies that E_d^* , the optimal value of E_d , must be less than or equal to 1. The optimal value, E_d^* , is the *overall efficiency* of DMU j . The left-hand sides of Eqs. (2)–(4) are weighted averages, because of Eq. (5), of the resources, performance measures, and site characteristics, respectively, of the 524 DMUs. At optimality, that is with the λ_j replaced by λ_j^* , we call the left-hand sides of Eqs. (2.1)–(4.5) the *target resources*, *target performance measures*, and *target site characteristics*, respectively, for DMU d .

Equations (2.1)–(2.3) imply that each target resource will be less than or equal to the actual level of that resource at DMU d . Similarly, Eqs. (3.1)–(3.4) imply that each target performance measure will be greater than or equal to the actual level of that performance measure at DMU d .

The nature of each site characteristic inequality in Eqs. (4.1)–(4.5) depends on the manner in which the site characteristic influences efficiency. Equations (4.1)–(4.4) correspond to unfavorable site characteristics (larger values imply a greater need for resources to obtain a given performance level, on average); therefore, we use the greater than or equal to sign. Equation (4.5) corresponds to a favorable site characteristic (larger values imply a lesser need for resources to obtain a given performance level, on average); therefore we use the less than or equal to sign. Thus, Eqs. (4.1)–(4.5) imply that the value of each target site characteristic will be the same as or worse than the actual value of that site characteristic at DMU d .

Thus, the optimal solution to the linear program for DMU d identifies a hypothetical target DMU d^* that, relative to DMU d , (a) consumes the same or less of every resource, (b) achieves the same or greater level of every performance measure, and (c) operates under the same or worse site characteristics. Moreover, the objective function expressed in Eq. (1) ensures that the target DMU d^* consumes resources levels that are reduced as much as possible in across-the-board percentage terms.

Of course, to proceed we must assume that a DMU could in fact operate exactly as does DMU d^* . In the theory of production, this is the assumption, made universally by economists, that the production possibility set is convex. In this context, the *production possibility set* is the set of all vectors $\{X_i, Y_r | S_k\}$ of resources, performance measures, and site characteristics such that it is possible for a DMU to use resource levels X_i to produce performance measures Y_r under site characteristics S_k . The convexity assumption assures that DMU d^* is feasible and that it is reasonable to expect that DMU d could modify its performance to match that of d^* .

We use the Premium Solver Pro[®] add-in (Frontline Systems, Inc., Incline Village, NV) in Microsoft Excel[®] to solve the linear programs. We use a macro written in Visual Basic for Applications[®] (VBA) to solve the 624 linear programs sequentially and save the results within the spreadsheet. Both the Basic Solver[®] and VBA[®] are available in all versions of Microsoft Excel[®]. However, the Basic Solver[®] is limited to 200 variables and 100 constraints, which limits the size of the problems to no more than 199 DMU and no more than 99 resources, performance measures, and site characteristics combined. We use the Premium Solver Pro[®], available from Frontline Systems, Inc., for this application.

The Performance Enhancement DEA Model

The performance enhancement DEA model with variable returns to scale, for DMU d , $d = 1, 2, \dots, 624$, is below. In this model, we eliminate E_d as the objective function (1) and from the resource constraints (2.1)–(2.3) and introduce θ_d as the new objective function (now to be maximized) and into the performance enhancement constraints (3.1)–(3.4). The parameter θ_d will now be greater than or equal to one, and it is called the *inverse efficiency* of DMU d .

Max θ_d	(1)	
Subject to		
$\sum_{j=1}^n \lambda_j X_{1j} \leq X_{1d}$	(2.1)	FTE teachers
$\sum_{j=1}^n \lambda_j X_{2j} \leq X_{2d}$	(2.2)	FTE teacher support
$\sum_{j=1}^n \lambda_j X_{3j} \leq X_{3d}$	(2.3)	Building administration and professional staff
$\sum_{j=1}^n \lambda_j Y_{1j} \geq \theta_d Y_{1d}$	(3.1)	Secondary level English (%)
$\sum_{j=1}^n \lambda_j Y_{2j} \geq \theta_d Y_{2d}$	(3.2)	Secondary level math (%)
$\sum_{j=1}^n \lambda_j Y_{3j} \geq \theta_d Y_{3d}$	(3.3)	Grade 8 science (%)
$\sum_{j=1}^n \lambda_j Y_{4j} \geq \theta_d Y_{4d}$	(3.4)	Graduation rate (%)
$\sum_{j=1}^n \lambda_j S_{1j} \geq S_{1d}$	(4.1)	Number of elementary school students
$\sum_{j=1}^n \lambda_j S_{2j} \geq S_{2d}$	(4.2)	Number of secondary school students
$\sum_{j=1}^n \lambda_j S_{3j} \geq S_{3d}$	(4.3)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^n \lambda_j S_{4j} \geq S_{4d}$	(4.4)	Percentage of students with limited English proficiency
$\sum_{j=1}^n \lambda_j S_{5j} \leq S_{5d}$	(4.5)	School district's combined wealth ratio
$\sum_{j=1}^n \lambda_j = 1$	(5)	Variable returns to scale
$\lambda_j \geq 0$ for $j = 1, 2, \dots, 624$	(6)	Nonnegativity
$\theta_d \geq 0$	(7)	Nonnegativity

The Mixed DEA Model

The mixed DEA model with variable returns to scale, for DMU $d, d = 1, 2, \dots, 624$, is below. In this model, we keep both E_d and θ_d in the constraints and we may now choose to either minimize θ_d or maximize θ_d . We introduce a new constraint (6) that ensures balance between the goals of reducing resources and enhancing performance.

Min θ_d or Max θ_d	(1)	
Subject to		
$\sum_{j=1}^n \lambda_j X_{1j} \leq E_d X_{1d}$	(2.1)	FTE teachers
$\sum_{j=1}^n \lambda_j X_{2j} \leq E_d X_{2d}$	(2.2)	FTE teacher support
$\sum_{j=1}^n \lambda_j X_{3j} \leq E_d X_{3d}$	(2.3)	Building administration and professional staff
$\sum_{j=1}^n \lambda_j Y_{1j} \geq \theta_d Y_{1d}$	(3.1)	Secondary level English (%)
$\sum_{j=1}^n \lambda_j Y_{2j} \geq \theta_d Y_{2d}$	(3.2)	Secondary level math (%)
$\sum_{j=1}^n \lambda_j Y_{3j} \geq \theta_d Y_{3d}$	(3.3)	Grade 8 science (%)
$\sum_{j=1}^n \lambda_j Y_{4j} \geq \theta_d Y_{4d}$	(3.4)	Graduation rate (%)
$\sum_{j=1}^n \lambda_j S_{1j} \geq S_{1d}$	(4.1)	Number of elementary school students
$\sum_{j=1}^n \lambda_j S_{2j} \geq S_{2d}$	(4.2)	Number of secondary school students
$\sum_{j=1}^n \lambda_j S_{3j} \geq S_{3d}$	(4.3)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^n \lambda_j S_{4j} \geq S_{4d}$	(4.4)	Percentage of students with limited English proficiency
$\sum_{j=1}^n \lambda_j S_{5j} \leq S_{5d}$	(4.5)	School district's combined wealth ratio
$\sum_{j=1}^n \lambda_j = 1$	(5)	Variable returns to scale
$E_d + \theta_d = 2$	(6)	Balance resource reduction and performance enhancement
$\lambda_j \geq 0$ for $j = 1, 2, \dots, 624$	(7)	Nonnegativity
$E_d, \theta_d \geq 0$	(8)	Nonnegativity

References

- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429–444.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1979). Measuring the efficiency of decision making units: Short communication. *European Journal of Operational Research*, 3, 339.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science*, 27, 668–697.
- Cooper, W. W., Seiford, L. M., & Tone, K. (1999). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver software*. Boston: Kluwer.
- Driscoll, D., Halcoussis, D., & Svorny, S. (2003). School district size and student performance. *Economics of Education Review*, 22(2), 193–201.
- Emrouznejad, A. (2014). DEAzone. <http://www.deazone.com/>. Accessed 16 Sept 2014.
- Emrouznejad, A., Parker, B. R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, 42, 151–157.
- Färe, R., Grosskopf, S., & Weber, W. L. (1989). Measuring school district performance. *Public Finance Review*, 17(4), 409–428.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A*, III, 253–290.
- Kang, B. G., & Greene, K. V. (2002). The effects of monitoring and competition on public education outputs: A stochastic frontier approach. *Public Finance Review*, 30(1), 3–26.
- Kirjavainen, T., & Loikkanen, H. A. (1998). Efficiency differences of Finnish senior secondary schools: An application of DEA and Tobit analysis. *Economics of Education Review*, 17(4), 377–394.
- New York State Department of Education. (n.d.). Guide to reorganization of school districts. http://www.p12.nysed.gov/mgtserv/sch_dist_org/GuideToReorganizationOfSchoolDistricts.htm. Accessed 17 Mar 2014.
- Ruggiero, J. (2007). Measuring the cost of meeting minimum educational standards: An application of data envelopment analysis. *Education Economics*, 15(1), 1–13.
- Sexton, T. R. (1986). The methodology of data envelopment analysis, In R. H. Silkman (Ed.), *Measuring efficiency: An assessment of data envelopment analysis*. New Directions for Program Evaluation No. 32 (pp. 7–29). San Francisco: Jossey-Bass.
- Sexton, T. R., Silkman, R. H., & Hogan, A. (1986). Data envelopment analysis: Critique and extensions, In Silkman, R. H. (Ed.), *Measuring efficiency: An assessment of data envelopment analysis*. New Directions for Program Evaluation No. 32 (pp. 73–105). San Francisco: Jossey-Bass.
- Thanassoulis, E., De Witte, K., Johnes, J., Johnes, G., Karagiannis, G., & Portela, C. S. (2016). Applications of data envelopment analysis in education. In *Data envelopment analysis* (pp. 367–438). New York: Springer.
- U.S. Census Bureau. (2011). *Public education finances*. <http://www2.census.gov/govs/school/11f33pub.pdf>. Accessed 17 Mar 2014.
- U.S. Government Spending. (n.d.). *New York state spending*. http://www.usgovernmentspending.com/New_York_state_spending.html. Accessed 17 Mar 2014.