Measuring Educational Productivity in Standards-Based Accountability Systems: Introducing the SES Return on Spending Index (RoSI)

Martin Hampel

Standard & Poor's

About the Author

Martin Hampel is associate director of Standard & Poor's School Evaluation Services (SES). Prior to joining Standard & Poor's, he worked as a quantitative analyst for a financial analytics software company in New York. He earned his doctorate in elementary particle physics at the Technical University of Aachen in Germany and the German collider research center DESY in Hamburg, where he also held a postdoctoral position. During his tenure at Aachen and DESY, he gained extensive experience in computer modeling and the quantitative analysis of large and complex data samples. His academic honors include a Fulbright Scholarship at SUNY Stony Brook, where he also obtained a master's degree, and two scholarships from the German National Merit Scholarship Foundation.

Measuring Educational Productivity in Standards-Based Accountability Systems: Introducing the SES Return on Spending Index (RoSI)

Martin Hampel

Standard & Poor's

I. Introduction and Motivation

For many years the school system in the United States has measured success by the number of dollars spent, computers and textbooks purchased, and programs created. Moreover, the measures of success have not focused on academic achievement. Since 1965, American taxpayers have spent more than $321 billion in federal funds on K–12 public education, yet the average reading scores for 17-year-olds have not improved since the 1970s, according to the U.S. Department of Education.[1] In an era where standards, testing, and accountability are at the forefront of the education community, parents, educators, administrators, legislators, and stakeholders are requiring an objective way of ascertaining the progress of public schools throughout the United States.

There is a rich body of literature studying the relationship between resources spent on education and educational outcomes such as performance on achievement tests, graduation rates, and other assessment indicators. Since there are several hundreds of studies investigating this topic, it is quite impossible to provide an exhaustive review of the literature, and any overview could not be comprehensive. However, a recent book by Armor (2003) provides a fairly representative synopsis of various groups of studies and ongoing discussions, in particular, investigations looking into a "production function" approach, i.e., the relationship between "input" variables, such as spending, and "output" indicators, such as performance on standardized tests. Armor had worked as a graduate student on the classic "Coleman" study (Coleman et. al 1966), which pioneered the identification of the relationship between socioeconomic background and student performance. His main thesis states that these family effects are greater than school grade level achievement, and therefore any influence of spending variables is typically less pronounced.[2]

Another literature review can be found in Monk, Walberg, and Wang (2001). Schweke (2004) provides an additional overview.

[1] U.S. Department of Education (n.d.). *How* No Child Left Behind *Ensures Schools Get Results*. Available at http://www.ed.gov/nclb/accountability/results/getting_results.pdf.

[2] One example of a discussion of the question "Does Money Matter?" is the exchange between Hanushek and Hedges. A thread of several related references can be found in Armor (2003).

While most academic research is obviously focused on identifying relationships between quantitative indicators, the methodology introduced here uses these underlying relationships as background variables but focuses on identifying the relative position of individual entities, such as school districts, to these environmental variables, which is important from a methodology perspective.[3]

Under the No Child Left Behind Act of 2001 (NCLB), states and school districts now have more flexibility in how they use federal education funds. Accordingly, Standard & Poor's School Evaluation Services (SES) introduces the "Return on Spending Index" (RoSI), which provides diagnostic information about the comparative educational return on resources generated by school districts in the United States. Used in combination with the "Error Band" method and the "Risk-Adjusted Performance Index" (described in two earlier SES reports, Gazzerro and Hampel [2004] and Hampel [2005], respectively), RoSI helps to identify school districts that achieve better educational performance for a given level of spending, while taking into consideration the proportional enrollment of economically disadvantaged students.

The "Return on Spending Index" (RoSI) provides diagnostic information about the comparative educational return on resources generated by school districts in the United States.

While the NCLB establishes the goal of educational proficiency in reading, math, and science, such proficiency is usually measured by cutoff scores that are used in a binary fashion, measuring a student's performance either above or below the standard. To rely upon standardized test scores to identify best practices in the classroom, more comprehensive measures of academic achievement are desirable.

"Gain scores" are measures of the progress that students make between the beginning and end of a school year. They are measures of the "return" on education resources and the public's investment in education. One way of analyzing gain scores is to use a costly system of annual value-added assessments that employ complex statistical models. The system also might require the use of unique student identifiers, so that the gains of student groups can be tracked over time. So far, cost, complexity, and in some cases, even mistrust, have kept most states from implementing value-added assessment systems.

Getting more out of test data

To assist states and districts that do not currently have value-added assessment systems but wish to get more out of their existing test data, SES offers a technique known as the "Error Band" analysis (Gazzerro and Hampel 2004). It determines whether a school is performing above or below the achievement range (the Error Band) typically associated with a concentration of disadvantaged and at-risk students.[4] Schools that consistently perform above this range may shed light on best practices that could be benchmarked and replicated by lower performing schools. This might be thought of as a bridge between traditional standardized testing and value-added assessment, with the benefit of meeting three elusive educational goals:

1. Accountability for school performance that takes into account different challenge levels for the purpose of measuring "Risk-Adjusted Performance" (Hampel 2005);
2. Diagnostic information that can be used to manage instruction; and
3. A potential source of best practices that work in practice, not just theory.

While this is a worthwhile approach in its own right and provides interesting and actionable insights, the "input" variable—poverty—cannot be controlled directly by

[3] An early study, which takes a somewhat similar approach, attempting to identify effective schools based on the analysis of residuals, can be found in Klitgaard and Hall (1973).

[4] The Error Band methodology performs a regression analysis and determines an index which is based on the distribution of schools' distances in performance from the regression line; these distances are commonly referred to as residuals. To make this approach more intuitively understandable for a lay audience, the performance Error Band is also referred to as "performance zone" in documents addressed to the general public.

education decisionmakers. It is therefore desirable to be able to analyze parameters that can actually be influenced, such as spending. Additionally, an important question to ask is what "return" in educational performance does a certain level of spending achieve?

Defining a return on spending index

Standard & Poor's methodology to analyze the return on educational spending will therefore be introduced in the following steps:

- Choosing an appropriate performance indicator.
- Selecting the appropriate corresponding spending variable.
- Defining a "Return on Spending Index" (RoSI). This indicator will provide a general productivity measure as a proxy for average educational return, given a certain spending level.
- Performing comparative "return" analysis. This entails transferring the principles of the Error Band and Risk-Adjusted Performance methodology to analyze the RoSI in relationship to relative poverty. Combining the RoSI and the Risk-Adjusted Performance data in one framework provides a powerful approach to study both simultaneously.[5]

II. Choosing an Appropriate Performance Indicator

An Error Band analysis can be performed for a single return indicator, such as the results of a standardized test, or for a combination of test results and other measures, such as graduation rate and retention rate.

An Error Band analysis can be performed for a single return indicator or for a combination of test results and other measures.

Figure 1 provides an example of a scatter plot showing the New York grade 8 Mathematics Test Proficiency Rate versus enrollment of economically disadvantaged students for 2002 at the district level. While passing and proficiency rates can be calculated at the school level as well, the often limited availability of financial data at the school level makes it necessary to perform the analysis at the district level.

In the *Resource Adequacy Study for the New York State Commission on Education Reform* (2004), Standard & Poor's introduced the "Multiple Performance Measures" (MPM) Index, which combines the weighted results of 13 state tests, averaged over 3 years (in this case, from 2001 to 2003), plus a corresponding graduation rate and retention rate indicator.[6] The corresponding Error Band plot is shown in figure 2.

While the poverty distribution in both plots is obviously identical, the slope of the regression line is much flatter for the MPM Index, and the width of the band is considerably smaller. This is due to the fact that the MPM Index is calculated as a comprehensive average of different performance indicators as well as over time, which reduces the statistical fluctuations. In addition, the aggregation of different tests and performance measures, which are not necessarily correlated and partly have a higher average, results in an increase of the average MPM Index value compared to the grade 8 mathematics test results.

Since financial information is usually only available at the district level and at a considerable degree of aggregation, an indicator such as the MPM Index is therefore more suitable for a productivity analysis than test results at a grade level, particularly when financial data for 1 year are used.[7] For this report, data come from the 2001–02 fiscal year.

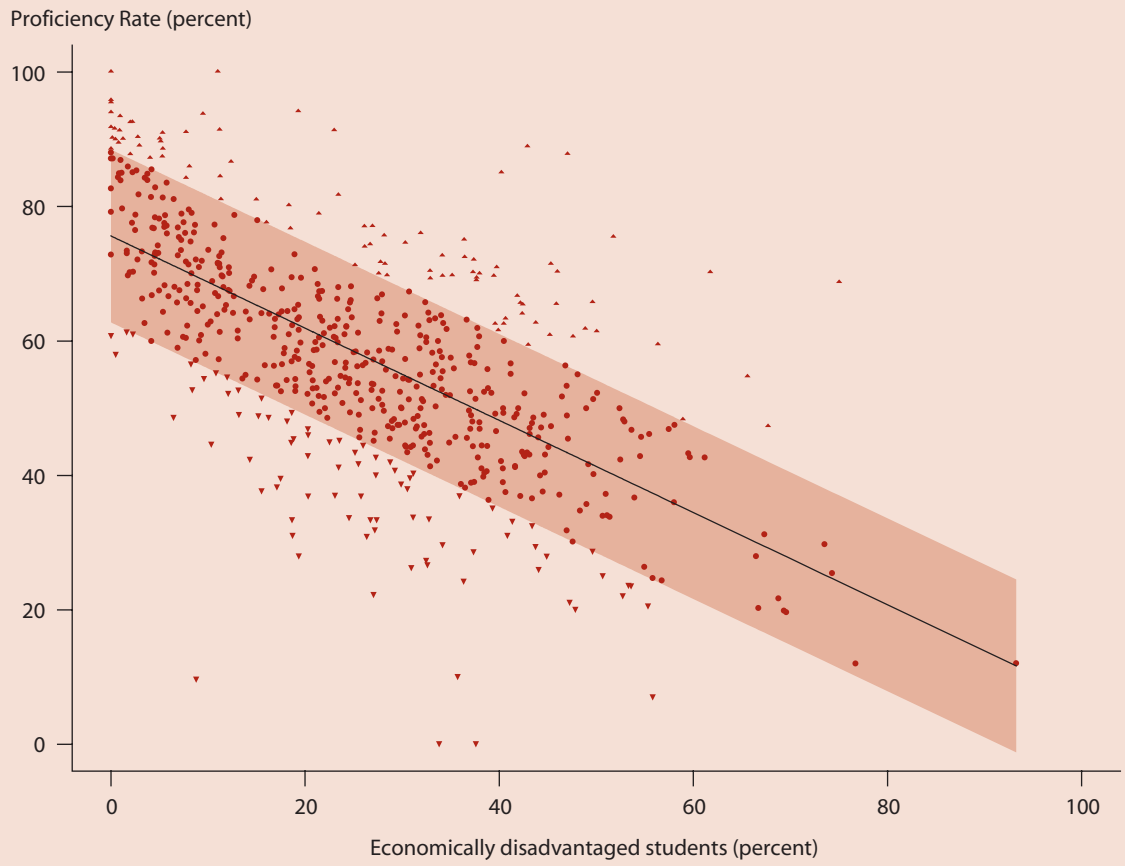
---

[5] In general, the principles of the Error Band and the Risk-Adjusted Return methodology can be applied to a wide range of statistical relationships, as long as some general underlying conditions regarding the data structure, such as conformity with assumptions typically made for regression analysis, are met. For additional information, see footnote 13.

[6] The report *Resource Adequacy Study for the New York State Commission on Education Reform* (2004), which provides further details, such as the exact definition of the MPM index, can be obtained at http://www.SchoolMatters.com. At the time of the publication of the study in March 2004, the latest financial data available were from the 2001–02 fiscal year, which are used in this paper.

[7] An aggregate performance indicator such as the MPM Index can be defined in any state using an analogous approach of combining available educational achievement measures.

Figure 1. Scatter plot of New York grade 8 mathematics test Proficiency Rate versus enrollment of economically disadvantaged students for 2002 (using available data for 635 school districts)



NOTE: The scatter plot includes a linear regression line and an "Error Band" that permits the identification of school districts that lie above, within, or below the band, indicated by upward facing triangles (▲), circles (●), and downward facing triangles (▼), respectively. The Proficiency Rate includes the percentage of students scoring at the proficient level or above.

SOURCE: Author's calculations from New York State data.

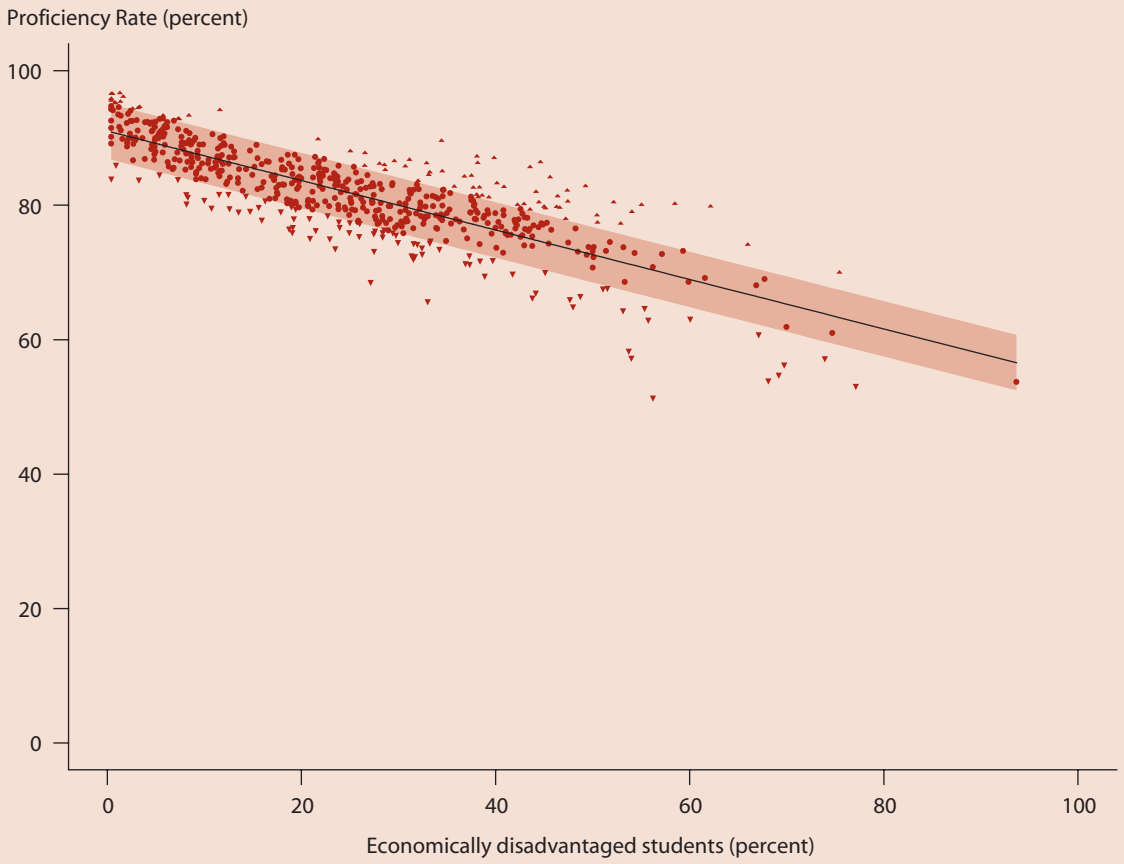
III. Selecting the Appropriate Corresponding Spending Variable

In order to combine the achievement indicator with a spending measure, an appropriate spending variable needs to be determined. Operating expenditures are suitable, since they exclude capital expenditures, which can vary widely from year to year and distort the influence of spending on day-to-day activities. For a similar reason, transportation expenses are excluded as well, as they depend to a large degree on the physical characteristics of each school district.

Another important aspect of the spending indicator is its scope. A "core" spending amount per student, which is defined as the total operating spending for the district divided by the number of enrolled students, provides a reasonable proxy for per student spending, since it includes the additional spending amounts assigned to students with limited English proficiency, students with disabilities, and economically disadvantaged students. Defining the spending variable in this way is particularly meaningful, since the subsequent RoSI Error Band analysis introduced below will explicitly take the proportional enrollment of economically disadvantaged students into account.

To control for in-state, regional differences in the purchasing power of the dollar, a geographic cost adjustment needs to be performed that expresses the spending amount in

Figure 2. Scatter plot of the Multiple Performance Measures Index (MPM) for New York State (using available data for 581 school districts), expressed as a Proficiency Rate



NOTE: As in figure 1, the scatter plot includes a linear regression line and an "Error Band" that permits the identification of school districts that lie above, within, or below the band, indicated by upward facing triangles (▲), circles (●), and downward facing triangles (▼), respectively. The Proficiency Rate includes the percentage of students scoring at the proficient level or above. For comparability purposes, the scale has been kept the same as the scale in figure 1.

SOURCE: Author's calculations from New York State data.

"standardized dollars," which are comparable across different districts. Standardized dollar amounts have a very well defined meaning that allows for a relative comparison of spending. However, since the scale of any cost adjustment is usually defined by normalizing spending to a particular geographic region, it should be recognized that within this context the absolute dollar amount is of limited use.[8]

For the purposes of this methodology paper, the standardized 2002 New York core expenditures per student were used, geographically cost adjusted by the New York Regional Cost Index.[9]

IV. Defining a "Return on Spending Index"

Standard & Poor's introduced the Performance Cost Index®(PCI) as a measure that allows for the comparison of spending and outcome measures in tandem. It was defined by the ratio of spending divided by a performance

[8] Further details about the aspects mentioned in this section can be found in *Resource Adequacy Study* (2004).

[9] The 2002 financial data were the latest data publicly available at the time of the publication of the New York Resource Adequacy Study. Since 2002 denotes the year in the middle of the 3-year period for the definition of the MPM Index, it can be seen as a reasonable spending proxy. To retain the properties of the spending data relative to other districts, no spending projections or inflation adjustments were made.

indicator, yielding the average amount of money spent per unit of achievement measured.[10]

The structure of such a measure with respect to the enrollment of economically disadvantaged students is usually dominated by the performance variable, which is typically much stronger than the relationship with the spending distribution. It is therefore suggested to invert the PCI to create a RoSI, which is defined as a performance indicator divided by a spending variable and can be interpreted as a productivity indicator.[11]

The additional benefit of the RoSI methodology lies in its more intuitive meaning as a measure of productivity. Larger values are often viewed more favorably than smaller values, as they indicate either higher performance, lower spending, or both; it is important to note that there may be exceptions where larger values should not be seen as better, depending on the underlying component values and local circumstances.[12]

V. Performing Comparative "Return" Analysis

The RoSI enables the use of an Error Band approach because when it is plotted against poverty it has a similar structure to the performance measure itself. Again, this behavior stems mainly from the trend of decreasing performance with increasing poverty, rather than the influence of spending. This means that one can identify statistically significant outperformers and underperformers in the RoSI, which combined with additional criteria such as a minimum performance level, signifies whether an entity is using resources efficiently.[13]

The RoSI enables the use of an Error Band approach because when it is plotted against poverty it has a similar structure to the performance measure itself.

Figure 3 shows the overlay of the performance Error Band of figure 2 with a RoSI Error Band in relationship to the enrollment of economically disadvantaged students. The right-hand scale for the RoSI variable has been adjusted such that the two regression lines lie on top of each other.[14] In order to make the plot easier to read, only districts that lie simultaneously above or below both Error Bands are shown; in addition, the Proficiency Rate range shown in the plot has been adjusted. To make the identification of corresponding data points easier, figure 3 shows a connection of the two data points for each district by a vertical line.

As one can see clearly, the Error Band for the RoSI is broader than the Error Band for the MPM Index itself. This is mainly due to the fact that the RoSI was calculated using the MPM Index as one input, increasing the statistical fluctuation in the RoSI value. It will therefore generally be the case that the RoSI Error Band is broader than the performance measure Error Band.

One possible follow-up analysis consists of looking at the Proficiency Rate value and the RoSI value for each district separately.

[10] Before the introduction of the NCLB testing requirements, an additional adjustment for test participation was usually included. Further details about the PCI can be found at http://www.SchoolMatters.com.

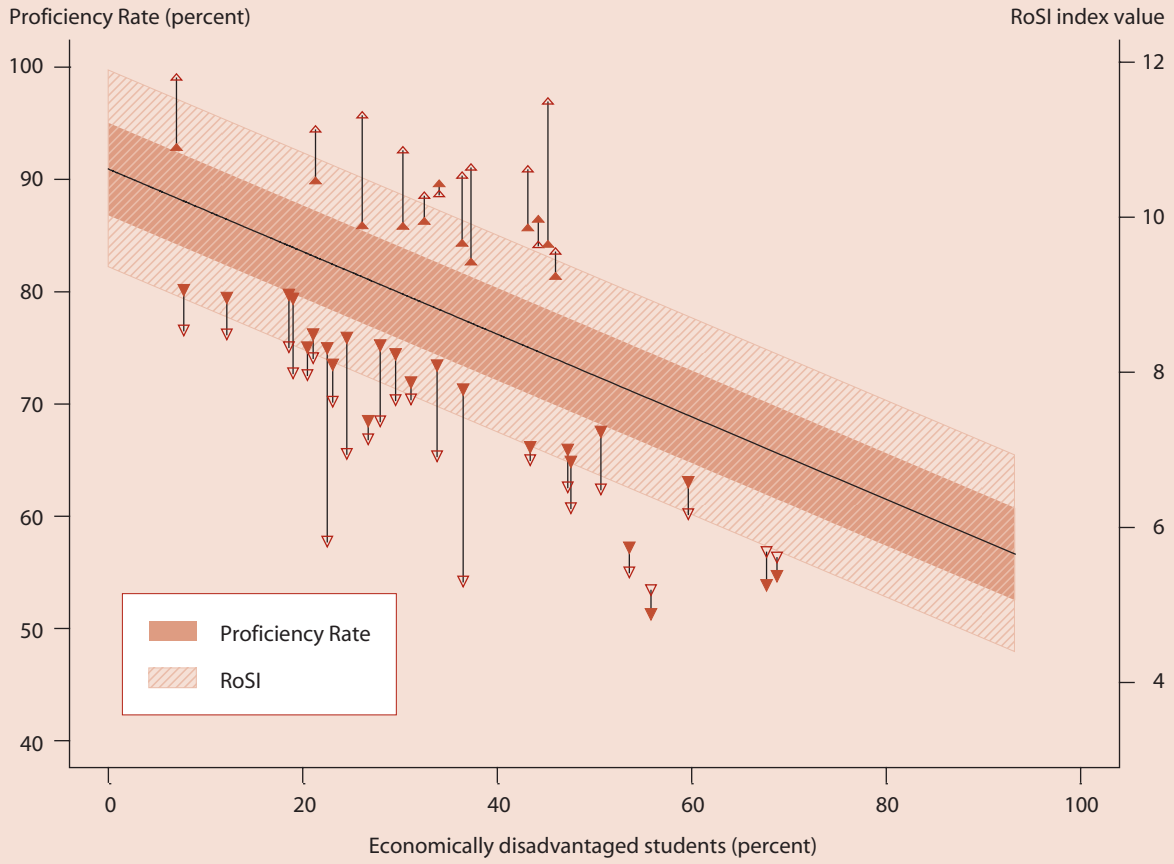
[11] In principle, any performance measure and any spending variable could be used to define a RoSI mathematically. However, a RoSI definition based on indicators with meaningful properties relating to the productivity relationship one is trying to measure is clearly preferable.

[12] Both the PCI and the RoSI are average indicators, not marginal. In the case of the PCI, it measures the average cost of a unit of student performance achieved, while the RoSI measures the average achievement level per unit of spending. It would generally be a mistake to assume that the return on spending or cost of student performance is always constant; in fact, one might expect diminishing returns at certain spending and performance levels. This is an important conceptual distinction, but not of any consequence for the analysis presented here since both spending and performance measures are defined as averages.

[13] As in the analysis of performance measures, the RoSI Error Band analysis needs to ensure that the criteria necessary for a regression analysis are sufficiently met. The goal of identifying outperformers and underperformers also requires the analysis of the data substructure such as by a localized and robust fit. This ensures that no nonlinearities in the relationships distort the results.

[14] This two-scale approach is always possible, as long as the signs of the slope of the regression lines are the same. Strictly, a RoSI has a unit of [%/$] if a passing or proficiency rate is used, but since the RoSI can be interpreted as an index, the unit-free representation is chosen, expressed as per $1,000 of spending. This also corresponds to the fact that the absolute scale of the index value is somewhat arbitrary due to the geographic cost adjustment of the spending indicator. This property (and the fact that each state generally uses its own performance indicator) usually prevents a direct comparison of RoSI values for different states.

Figure 3. Extended plot of scatter plot in figure 2, with the addition of a RoSI Error Band using the scale at the right-hand side of the plot, rescaled such that the linear regression lines overlap



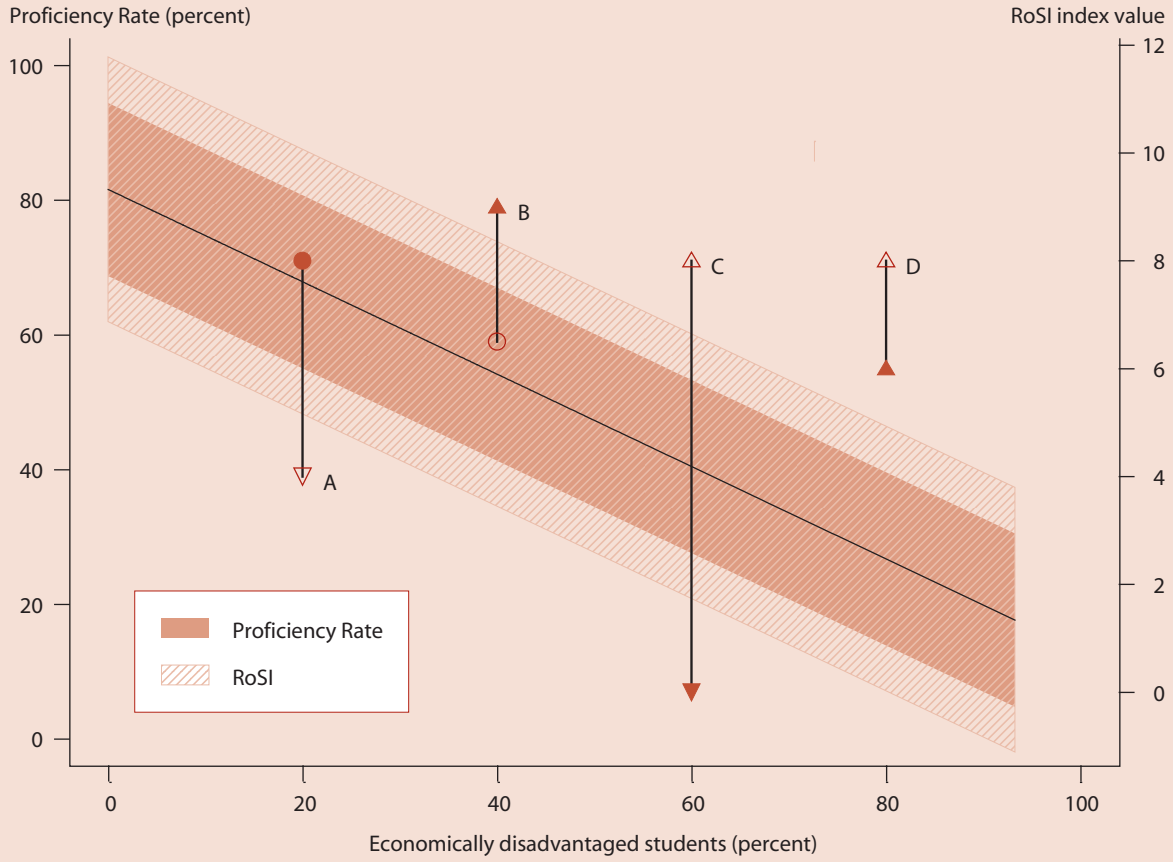
NOTE: Only districts that lie simultaneously above or below both Error Bands are shown; in addition, the Proficiency Rate axis scale has been adjusted. Proficiency Rates above and below the Proficiency Rate band are indicated by upward facing triangles (▲) and downward facing triangles (▼), respectively. RoSI values of districts that lie above or below the RoSI band are indicated by open upward facing triangles (△) and downward facing triangles (▽), respectively. In addition, Proficiency Rate and RoSI data points of each district are connected by a vertical line.

SOURCE: Author's calculations from New York State data.

The data presentation in figure 3 combines a wealth of information into a single plot. To illustrate this relationship more explicitly, a hypothetical example is drawn in figure 4, with four potential combinations of data point pairs A through D.[15] Pair A denotes an entity with performance within the performance Error Band, but a RoSI value that lies below the RoSI Error Band. This could be interpreted as demonstrating performance within statistical expectation accompanied by educational returns on spending below the statistical expectation, i.e., a spending level that is relatively high given the associated performance level and the proportional enrollment of economically disadvantaged students. Correspondingly, pair B shows a proficiency outperformer with a RoSI value within the RoSI Error Band, which could be interpreted as a spending level within statistical expectation. Pair C combines a proficiency underperformer with a RoSI value above the RoSI Error Band, i.e., a spending level significantly below expectation. Finally, pair D shows a proficiency outperformer combined with a RoSI value above the RoSI Error Band. This entity demonstrates arguably the most desirable behavior, which consists of proficiency above the statistical

[15] Since a data point for each Error Band can lie above, within, and below the respective band, a total of nine combinations for each data point pair are possible for the analysis of two simultaneous Error Bands.

Figure 4. Hypothetical example of data point pairs relative to the Risk-Adjusted Performance Error Band and RoSI Error Band



NOTE: Proficiency Rates above, within, and below the Proficiency Rate band are indicated by upward facing triangles (▲), circles (●), and downward facing triangles (▼), respectively. RoSI values of districts that lie above, within, or below the RoSI band are indicated by open upward facing triangles (△), circles (○), and downward facing triangles (▽), respectively. In addition, Proficiency Rate and RoSI data points of each district are connected by a vertical line.

SOURCE: Figure by author.

expectation, while at the same time obtaining this proficiency level with a high level of productivity (i.e., relatively low spending for the given level of performance).

A particular RoSI value could be due to a relatively high performance level and correspondingly high spending level or, conversely, relatively lower performance and lower spending. Therefore, analyzing the RoSI value in connection with the actual performance indicator provides insight into whether a large RoSI value is due to higher performance or just lower spending.

One particularly valuable output of this method is that the RoSI Error Band permits the production of a measure of "Risk-Adjusted Return," i.e., a "Risk-Adjusted Productivity" similar to the Risk-Adjusted Performance Index value for the performance indicator. This way, one can quantify how far away the RoSI value lies from the regression line, given the relative enrollment of economically disadvantaged students.

Since the MPM index is defined as a 3-year average, fluctuations are already smoothed out considerably, which inherently increases the robustness and usefulness of the RoSI analysis. In addition, Error Band analyses could be performed for a sequence of years with a correspondingly adjusted MPM Index definition and spending adjustments, which would allow for the study of the develop-

ment of the productivity of each district over time similar to a multiyear analysis of the performance Error Band.

The RoSI approach presented in this paper expands the Error Band analysis of a performance measure in relationship to the enrollment of economically disadvantaged students to the study of spending and performance. It thus helps to provide actionable information using independent data concerning spending decisions that are under the control of educational decisionmakers.[16]

In addition to presenting the graphical representation, which is instructive in its own right, Standard & Poor's is currently considering integrating this type of analysis as part of its analytical website offering, and adding tools to allow the identification of districts by Risk-Adjusted Performance and Risk-Adjusted Productivity criteria. Further information can be found at http://www.SchoolMatters.com.

Other directions of potential future research include the extension of this approach to school buildings, if financial information at the school building level becomes available. One likely difficulty at the school level would be the probable increase in data uncertainties and fluctuations due to reporting issues and varying interpretations of accounting standards and reporting requirements between schools.

Furthermore, the productivity approach discussed here could be analyzed in more detail by including additional indicators on the spending as well as the performance side, and by also taking demographic environment variables into account. Some of these enhancements might be performed based on the Error Band analysis framework, allowing for a rich view of educational data.

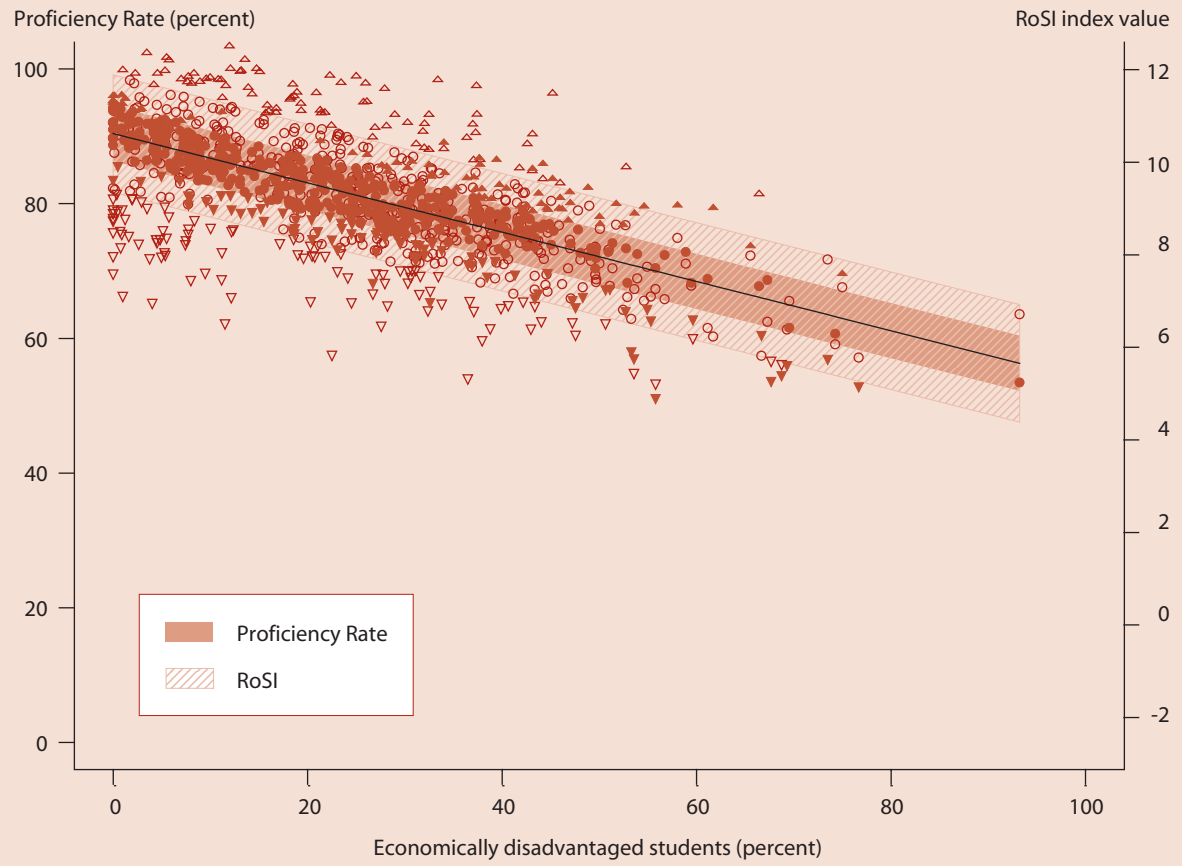
VI. Appendix

This appendix shows two plots containing the full set of data points utilized for the Proficiency Rate and RoSI Error Band analysis. Figure 5 contains essentially the same two Error Bands as figure 3 without the connecting lines between data points. All data points are shown, and the Proficiency Rate scale has been kept the same as in figure 2 to allow for a direct comparison.

Figure 6 shows the same information as figure 5, with corresponding data points connected by vertical lines. Although this plot contains information similar to figure 3, it shows all data points, not only those where both the Proficiency Rate and the RoSI values lie simultaneously above or below the corresponding bands. This way, a direct comparison to figure 2 is possible, and the dramatic effect of the range of possible combinations of Proficiency Rates and RoSI values is illustrated.

[16] One possible extension of this approach could be a true multivariate analysis of either the Proficiency Rate and/or the RoSI as a function of a set of other learning environmental or demographic variables that have been shown to be correlated with student performance. Such an analysis would obviously be more challenging to present graphically, and the relatively small number of available sample data points would likely make the meaningful identification of outperformers in each dimension more difficult, particularly since the analysis is focusing on the distribution of residuals rather than only the accuracy of the regression itself. The current approach takes additional characteristics into account when benchmarking studies are conducted to match underperformers and outperformers, requiring that the entities under consideration are matched with respect to additional variables, thus avoiding the density dilution effect of multidimensionality.

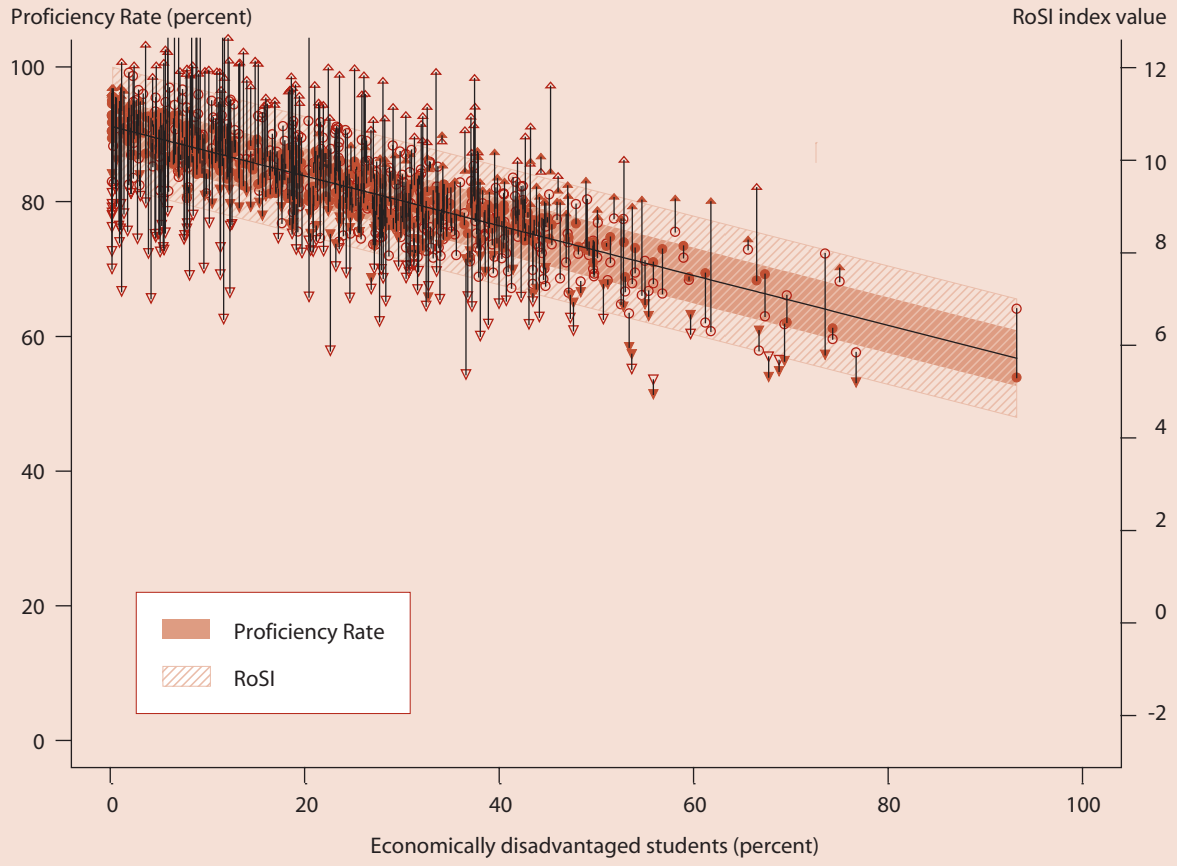
Figure 5. The same plot as in figure 2, with the addition of a RoSI Error Band using the scale at the right-hand side of the plot, rescaled such that the linear regression lines overlap



NOTE: Proficiency Rates above, within, and below the Proficiency Rate band are indicated by upward facing triangles (▲), circles (●), and downward facing triangles (▼), respectively. RoSI values of districts that lie above, within, or below the band are indicated by open upward facing triangles (△), circles (○), and downward facing triangles (▽), respectively. The Proficiency Rate axis scale has been kept the same to allow for direct comparisons with figure 2.

SOURCE: Author's calculations from New York State data.

Figure 6. The same plot as in figure 5, with the addition of performance and RoSI data points of each district connected by a vertical line



SOURCE: Author's calculations from New York State data.

VII. References

Armor, D.J. (2003). *Maximizing Intelligence.* New Brunswick, NJ: Transaction.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., and York, R. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office. (ERIC ED012275)

Gazzerro, P., and Hampel, M. (2004, May). *Identifying Outperforming and Underperforming Schools: Introducing the SES "Error Band" Method* (SES Research Methods Paper No. 1). New York: Standard & Poor's. Available at http://www.SchoolMatters.com.

Hampel, M. (2005, March). *Introducing the SES Risk-Adjusted Performance Index* (SES Research Methods Paper No. 2). New York: Standard & Poor's. Available at http://www.SchoolMatters.com.

Klitgaard, R.E., and Hall, G.R. (1973). *A Statistical Search for Unusually Effective Schools* (RAND R-1210-CC/RC). Santa Monica, CA: RAND.

Monk, D.H., Walberg, H.J., and Wang, M.C. (2001). *Improving Educational Productivity.* Research in Educational Productivity. Greenwich, CT: Information Age.

Schweke, W. (2004). *Smart Money: Education and Economic Development.* Washington, DC: Economic Policy Institute.

Resource Adequacy Study for the New York State Commission on Education Reform (2004, March). New York: Standard & Poor's. Available at http://www.SchoolMatters.com.

A Cost Allocation Model for Shared District Resources: A Means for Comparing Spending Across Schools

Lawrence J. Miller
University of Washington

Marguerite Roza
University of Washington

Claudine Swartz
Policy Consultant

About the Authors

Lawrence J. Miller, MBA, is a research analyst at the Center on Reinventing Public Education, Daniel J. Evans School of Public Affairs, University of Washington. His research focuses on education finance and equity. He can be contacted at ljmiller@u.washington.edu.

Marguerite Roza, PhD, serves as a Research Assistant Professor at the Daniel J. Evans School of Public Affairs at the University of Washington. Her research focuses on quantitative policy analysis, particularly in the area of education finance and labor markets. She can be contacted at margroza@u.washington.edu.

Claudine Swartz, MPP, is a policy consultant focused on analysis of K–12 education issues and working with the Center on Reinventing Public Education at the University of Washington. She focuses on education finance and leadership issues and has presented before the American Education Finance Association, the Education Writer's Association, the American Educational Research Association, and the Association of Public Policy Analysis and Management. Pending publications include studies of within-district resource allocation and the supply of school principals. She can be contacted at claudineswartz@sbcglobal.net.

A Cost Allocation Model for Shared District Resources: A Means for Comparing Spending Across Schools

Lawrence J. Miller
University of Washington

Marguerite Roza
University of Washington

Claudine Swartz
Policy Consultant

Acknowledgments

We thank Jerry Wartgow, Superintendent of the Denver Public Schools, and his staff for their help in gaining access to that district's budget detail. We also thank Paul Hill, Dan Goldhaber, and Leslie Breitner of the Daniel J. Evans School of Public Affairs, University of Washington, for their helpful comments on earlier drafts.

Introduction

According to figures reported by the Cross City Campaign for Urban School Reform (2001), school-level budgets in 10 North American urban school districts consume from 38 to 95 percent of total district appropriations (see table 1).[1] With such enormous variation, one is left wondering whether the differences reflect drastically different spending patterns or simply reflect different accounting methods. In either case, the data fuel the mounting concerns about how to report costs among our nation's schools (Educational Testing Services 2004). Centrally reported costs can represent a significant portion of a district budget, but we have relatively little sense of how these dollars are distributed among or benefit different schools.

Schools receive many shared district resources, which can be important drivers of variation in school spending. Shared resources are the people, equipment, grants, and services housed or supervised by the central office that directly service and benefit schools in their efforts to educate students. Shared resources are reported centrally despite the fact that much of these resources are deployed outside the central office and inside schools. For example, services for non-English-speaking students are often delivered by a team of centrally managed specialists, despite the fact that students receive these services within their own building. Gifted and talented

[1] All but one, the Edmonton, Alberta, district, are U.S. school districts.

NOTE: Earlier versions of this paper were presented at the 2004 American Education Finance Association conference in Salt Lake City, UT, and at the 2004 National Center for Education Statistics summer data conference in Washington, DC.

Table 1. Reported school-level spending varies significantly: By district, 2001–02

School district	Total district appropriations	Percent of district budget reported at school level
Denver, CO	$ 910,555,851	38
Baltimore, MD	881,167,245	46
Chicago, IL	4,400,000,000	52
Oakland, CA	600,000,000	53
Seattle,WA	610,100,000	56
New York, NY	14,900,000,000	63
Philadelphia, PA	1,900,000,000	77
Edmonton, AB	545,000,000	80
Los Angeles, CA	9,800,000,000	99
Milwaukee, WI	1,000,000,000	95

SOURCE: Adapted by authors from data from the Cross City Campaign for Urban School Reform, Annual Decentralization Progress Comparison Across Ten Cities, 2001–02 school year. Appropriations for Baltimore, MD, were for 1999–2000.

programs, many of which include specialists that teach pullout programs within schools, are often controlled centrally. Other central services deployed outside the school building also benefit schools and students. For instance, many districts have centrally run professional development programs aimed at building teaching skills at low-performing schools. Because shared resources are centrally reported, rather than accounted for at the school level, it is difficult to compare spending on shared resources from school to school.

There are at least four reasons why lack of clarity around how shared district resources are distributed among schools is problematic. First, fully accounting for spending by school is critical for accountability reforms. Unmeasured and unreported variations in school resources call into question whether all schools are provided equal resources to meet accountability requirements. Second, the courts have not tolerated between-district inequity and, given that recent studies show significant variation in spending between schools, districts should be similarly concerned about legal remedies in within-district inequity cases (Iatarola and Stiefel 2003; Rubenstein, Schwartz, and Stiefel 2004). Third, district managers and board members currently rely on insufficient school-level accounting data to inform resource allocation decisions. Incomplete school-level funding data increase the probability of misalignment between spending decisions and district strategy. Fourth, the findings of resource effectiveness studies rely on data that, in some districts, capture as few as one-third of the dollars actually spent in the school. New studies that utilize data that more fully account for school-level spending may find a stronger relationship between resources and student outcomes.

To improve our understanding of school spending, a model to fully account for shared resources at the school level is developed here. The model both accounts for resources by schools and classifies resources according to type of students served. The model repairs outdated budgeting and accounting practices, bringing them into alignment with new policies where schools, not districts, are the focus.

Background

Two ways of measuring resources at the school level are currently utilized: a resource-based approach and an accounting approach. The resource-based approach classifies district funds according to the nonmonetary resources purchased (e.g., teacher characteristics, teacher-pupil ratios) and can facilitate answering questions about the effectiveness of different combinations of resources, including teacher qualifications, length of the instructional day, and class size (Chambers 1999). The accounting approach records resources in terms of their cost in dollars. Several researchers have used the accounting approach to compare spending across schools, examining different portions of district spending (Iatarola and Stiefel 2003; Roza and Hill 2004; Rubenstein, Schwartz, and Stiefel 2004). The accounting approach lends itself well to addressing questions involving comparisons of total resources across schools, because it provides a single metric (dollars per pupil) by which to compare all resources.

Districts use the accounting approach to meet the financial reporting requirements of regulators, private creditors, and other external stakeholders (Chambers 1999; Fowler 2001; Hartman, Bolton, and Monk 2001). Such external pressure has resulted in widely available district-level financial information, but there has been little pressure to report much of this spending at the school level. As a result, we often know how much is spent districtwide on instructor salaries and textbooks, but fail to know how these resources are distributed among schools. Efforts to address weaknesses in school-level data have resulted in some improvements; for example, 20 states now require school budgets that enable some degree of school-level comparisons (Fowler 2001). However, school budgets do not include many centrally reported resources that appear instead in consolidated central department budgets, making it unknown which schools ultimately benefit from them.

We often know how much is spent districtwide on instructor salaries and textbooks, but fail to know how these resources are distributed among schools.

One plausible explanation for why school budgets, including shared resources, do not reflect the full cost of educating students is the lack of consensus on the primary objective of school-level information. Researchers have proposed several different school budget models, each with a different objective in mind. For example, school budgets that reflect educational strategies report information in a format that facilitates the comparison of school reform models, instructional strategies, and resource deployment (Odden et al. 2003). Another model uses the locus of control to define costs allocated at the school level, including only those resources over which the school has budgetary authority (Odden and Busch 1998). As a result, a school's budget includes teacher salaries only if the school is given recruiting and staffing authority. While these approaches accomplish their stated objectives, neither seeks to fully account for school-level spending.

Other models suggest a trend toward accounting for more spending at the school level. One model, developed by Coopers & Lybrand, accounts for a greater share of district resources at the school level by allocating costs based on the face-to-face principle. With this model, only the cost of personnel who physically work within schools is reported at the school level, while administration and operations costs associated with central services remain centrally reported (Coopers & Lybrand LLP 1994). While this strategy enables us to report a greater portion of shared resources at the school level than is typically reported, it excludes indirect costs and effectively underprices the marginal cost of shared resources delivered at the school site.[2] As a result, when school-based resources are compared to shared district resources, shared resources appear more cost effective than they actually are and some within-district variation is lost.

The Core Finance Task Force of the National Forum on Education Statistics calls for districts to allocate *all* spending to schools, including district administrative and school board costs. The rationale states that "the provision of educational services through operation of schools is the only product of a school district [and] the allocation of these costs is necessary to full costing of the schools and their programs" (National Forum on Education Statistics 2003). No cost allocation model has yet been developed to execute the recommendation.

In sum, although different accounting models have been proposed to allocate more resources to school budgets, none of them are designed to fully capture shared resources in a way that allows for comparing spending differences between schools. The next section proposes a model by which typical school budget data are supplemented with spending data on shared district resources in order to gain a more complete picture of district spending on each school.

Shared District Resources Cost Model

New methods for accounting for district resources inevitably involve numerous decisions about how and where to record resources. The model proposed here has been

[2] The indirect portion of centrally reported costs can be significant; for example, Denver Public Schools (DPS) Title I costs were $22.2 million in fiscal year (FY) 2002–03, and $1.7 million (8 percent) of those costs were spent on the administration function.

developed specifically to facilitate meaningful spending comparisons among schools within a district. Toward this end, we use a set of principles to guide the design of the model. Most importantly, the model must properly account for resources in terms of the *schools* that they benefit. In addition, the model must generate comparable data (to enable resource comparisons) and thus must convert resources into a common metric (dollars). Moreover, the conversion must use real, instead of average, costs, as average costs mask spending variations between schools (Roza and Hill 2004). Finally, the model must account for spending by student need in order to delineate spending differences among schools with differing student needs.

Using these guiding principles, the model outlined here follows three structured steps: First, we identify shared district resources that benefit different schools and thus ought to be included in spending comparisons across schools. Second, we allocate those costs (in real dollars) to the schools that receive them. Third, we classify costs based on student need.

There are no widely accepted guidelines for determining which costs to report at the district versus the school level.

Step 1: Identifying Shared District Resources to Allocate

There are no widely accepted guidelines for determining which costs to report at the district versus the school level. Historically, costs have been classified as one of two types, central or school based. The vague term "central" necessitates further clarification, as it includes resources used to benefit students (sometimes unevenly) among schools. We divide central costs into two categories in order to identify resources relevant to spending differences among schools and those that are not: shared district resources and resources for district leadership and operations (see figure 1). The addition of typical school site budgets to the combination of these two types of central costs rounds out a district's overall spending framework: (1) school budgets; (2) shared district resources; and (3) resources for district leadership and operations.

School budgets generally report site-based costs, including the cost of the teachers and administrators who work there. Examples of site-based costs include classroom teachers, principals, librarians, and instructional aides. Other site-based costs sometimes reported in school budgets include facilities, operations, supplies, and materials.

Shared district resources, as defined here, include the people, equipment, grants, and services housed and supervised by the central office and used to directly service and benefit students and schools by central office managers or the school board. Shared district resources are currently reported in a consolidated fashion, typically in line-item, department, and program budgets. Examples of shared district resources include itinerant art and music teachers, centrally operated gifted programs, professional development, psychologists, and curriculum services.[3]

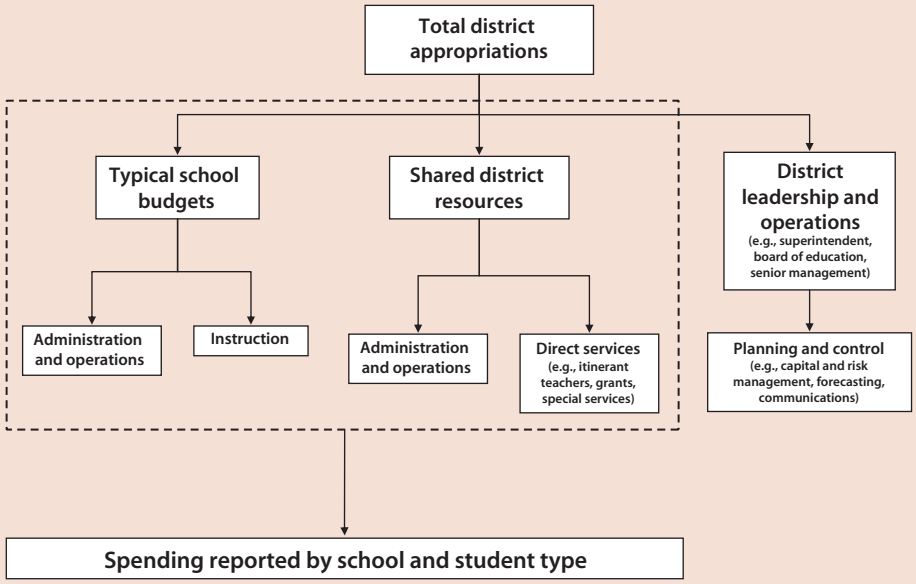
Resources for district leadership and operations, in contrast to shared district resources, do not include services for specific schools or students. District leadership and operation costs are composed entirely of indirect support services that are not used at the school level (e.g., the office of the superintendent, governance costs such as the board of education, and capital and risk management expenses). Indirect services can only be allocated to schools formulaically, typically on a per pupil or per school basis. For instance, because the superintendent's office (in medium and large districts) does not typically direct its services toward one school versus another, these costs could only be allocated to the school level by allocating them in an equal dollar amount per pupil. Such information adds little to our understanding of actual between-school spending variations. For this reason, our model keeps these costs consolidated.

Step 2: Allocating Shared District Resources

Allocating shared district resource costs to schools is challenging for two reasons: First, in contrast to site-based costs, shared resources generally serve multiple schools, and this often necessitates data collection activities to

[3] Although most districts opt to provide such services centrally, they may also choose to decentralize and grant control to schools or procure services from outside providers.

Figure 1. District spending framework



SOURCE: Framework by authors.

trace the schools, students, or school-based personnel on which the resources were ultimately brought to bear. Second, overhead costs related to shared district resources must be disentangled from district leadership and operations costs in order to allocate the full cost of shared services.

While there is no one approach to dividing shared resources among multiple sites, the practice is common in other public and private sectors (Cooper and Kaplan 1999; Horngren, Data, and Foster 2002). For this model, various cost accounting practices were adapted to create a seven-step process to guide the allocation of shared district resources: (1) identify the cost objects to be allocated (e.g., labor hours, program materials, grant dollars); (2) identify the direct costs associated with each cost object; (3) identify indirect costs associated with each cost object; (4) define the cost-allocation basis for allocating indirect costs to the cost object (e.g., flat rate, per pupil weighting); (5) compute the indirect cost per unit; (6) compute the direct cost per unit; and (7) compute the total cost per unit. Additional information and examples of each step are outlined in table 2.

Some discussion of direct and indirect costs can clarify the above steps. Direct costs are costs that can be traced directly to the schools where they are used. For a gifted program, the cost of itinerant gifted teachers is a direct cost because the labor hours for each of these teachers can be traced directly to a school. "Labor hours" then becomes the "cost object." Other examples of cost objects might be the number of students participating in a centrally offered program, and the number of school personnel participating in professional development or receiving support services. Undoubtedly, tracking costs by these new "cost objects" will require additional data collection in some cases.

Shared district resource costs that cannot be traced directly to a school are referred to as indirect costs. In the gifted program example, the costs of the administrator and support staff necessary to run the program are classified as indirect costs because staff time is not traceable to individual schools. What makes the allocation of indirect costs for shared resources unique in this model is that they have direct costs to guide their allocation, which markedly improves the accuracy of reported school-level resources.

Step 3: Classifying Costs According to Student Need

The final step in the model is to classify shared district resources based on the student needs they intend to serve. We do so by classifying resources as categorical or

Table 2. Shared resource cost allocation

Activity	Description	Examples
1. Identify the cost objects to be allocated.	The unit of measure for the product or service the model is costing.	Service hours (e.g., psychologists, nurses, social workers, gifted teachers), pages translated (e.g., translation department), or dollars distributed (centrally controlled school grants).
2. Identify the direct costs associated with each cost object.	Costs that can be traced to their recipients.	The total compensation of itinerant and substitute teachers.
3. Identify the indirect costs associated with each cost object.	Costs related to the cost object that cannot be traced to that cost object in an economically feasible way.	Administration and overhead costs of shared district resource departments, including gifted programs and psychologists, and curriculum and development.
4. Define the cost-allocation basis for allocating indirect costs to the cost object.	There are several methods to consider, including weighting the allocation by direct cost or computing a flat rate per unit produced. Weighting the allocation by direct cost works well in departments where the cost object is heterogeneous, whereas weighting by unit produced works best when cost objects are homogeneous.	Allocating indirect costs in a gifted program can use a flat rate per hour of service because the department provides a single type of service. The curriculum and development department, with multiple service lines and programs, is better served by a weighted direct cost allocation model.
5. Compute the indirect cost per unit.	Divide the indirect cost allocation by the number of units in the cost object.	
6. Compute the direct cost per unit.	Divide the direct costs by the number of units in the cost object.	
7. Compute the total cost per unit.	Combine the direct and indirect costs for each unit.	

SOURCE: Table by authors.

noncategorical. Categorical costs are earmarked to serve specific student needs, and are further classified according to the common student identifiers of poverty, minority, bilingual, gifted and talented, and vocational education.[4] Categorizing funding allows for comparison of schools with different school populations and an assessment of whether a particular school receives greater, or less, than the district average cost for a given type of student need. By default, costs not labeled categorical are classified as noncategorical costs.[5]

Application of the Model

The shared resources cost allocation model was applied to an existing dataset of school-level financial data collected from the Denver Public Schools (DPS) during the 2002–03 school year. The DPS is a large urban district serving approximately 72,000 racially and economically diverse students in its 148 schools.

The model was applied to the DPS dataset according to the steps outlined above: (1) shared district resources were identified using the district spending framework; (2) shared district resource costs were allocated to the schools that received them; and (3) costs were classified according to student need. For illustrative purposes, this paper also reports how application of the model illuminates spending differences (by accounting for shared district resources) in two DPS middle schools (Middle School A and Middle School B). The schools were selected for comparison because of their similar demographics and size (summarized in table 3), and state academic rating (both were labeled "low academic performers").

Step 1: Identifying the DPS's Shared Resources

In the DPS, school budgets represented 45 percent of operating costs, while 55 percent of operating costs were reported centrally. Twenty-five percent of these centrally

[4] For this model, a district would use any student characteristics that would dictate additional need, such as homeless, pregnant, migrant, etc.

[5] Because these dollars do not intend to serve a specific student need, we might expect them to be equally distributed to all students (Berne and Stiefel 1994).

reported costs were identified as shared district resources and allocated to the schools that received them. Where district data were insufficient or unavailable, we were not able to allocate shared district resources to schools, resulting in a significant portion of shared district resources that are not tracked by student or school.

Step 2: Allocating the DPS's Shared Resources

After allocating a portion of shared district resources in the DPS, school-level resources increased by nearly one-third, relative to the original amount reflected in school budgets.[6] The distribution of shared resource costs in the DPS allowed us to gain information about how an additional $92 million was spent from school to school. On average, it added an additional $1,058 in per pupil costs, but these resources were not distributed evenly. The maximum gain from shared resources at a school was $1,985 per pupil, while the maximum loss was $666 per pupil, a $2,651 range.[7]

After allocating a portion of shared district resources in the DPS, school-level resources increased by nearly one-third, relative to the original amount reflected in school budgets.

Comparing two middle schools (see table 3), Middle Schools A and B, before the allocation of shared resources, we found that the former received $8 per pupil ($6,728 total) less than the district average and the latter received $117 per pupil ($84,708 total) more than the district average. In short, comparing school budgets alone, it appears that the DPS spent $125 more per pupil ($91,436 more total) on Middle School B than on Middle School A.

When we looked at how Middle Schools A and B fared after shared resource costs were allocated by student need, a new picture emerged. Middle School A received $331 more per pupil ($278,371 total) than the district average while Middle School B received $549 less per pupil ($397,476 total) than the district average. Comparing the combined resources of school budgets and shared resources reversed our original assessment; a greater share of district resources was expended on Middle School A, which actually received $880 per pupil (or $675,847) more than Middle School B.

Step 3: Classifying the DPS's Costs According to Student Need

Shared resource costs were classified as categorical (e.g., bilingual, gifted) or noncategorical as described in table 3. Classifying costs in this way illuminated variation by student type. We found that the additional resources received by Middle School A were concentrated in two categories: noncategorical and poverty. Conversely, those same two categories represented where Middle School B was shortchanged on most of its shared resources. We identified similarities as well; both schools received less than the district average per pupil cost on bilingual education.

District Implementation

Just how likely is district implementation of a shared resource cost allocation model? This question raises issues about demand for the model, cost of implementation, and other key challenges. As earlier acknowledged, there are clear benefits to measuring shared district resource costs, but district demand for this level of information is not clear. The practice of maintaining central control over a large portion of district resources is widely accepted and the will to untangle, and account for, this money is not now evident. It is likely that demand will only surface with external pressure from interest groups, researchers, and parents who are interested in understanding whether resources are equitably distributed.

[6] It is important to note that the data used here to demonstrate the significance that shared resources have on actual school spending do not represent a full and complete shared resources analysis of the DPS. Of the $371 million in centrally controlled budgets, this database contains $92 million worth of shared resources that have been identified as shared resources and allocated to the schools that received them. The amount of resources present in a school that are centrally controlled and not reported in school budgets is underreported by this data.

[7] Student need is controlled for in this financial analysis by calculating the district average cost for each student type and multiplying the average cost by the number of students in the school. For example, if the district spends $600 per pupil on children of poverty, a school serving 100 students who qualify for free or reduced-price lunch would expect to receive $60,000 in compensatory education funds. To facilitate interpretation, schools that receive the district average are set to zero; schools that receive more than the district average are reported as a positive value, and schools that receive less than the district average are reported as a negative value.

Table 3. A comparison of school allocations to district averages: 2002–03

Characteristic	Middle School A	Middle School B
School type	General education	General education
Demographics		
Enrollment	841	724
Percent minority	94	80
Percent limited English proficient	28	16
Percent poverty	93	74
Percent gifted	8	13
Academic performance	Low	Low
Financials (in dollars per pupil)		
School budget	(8)	117
Shared resources		
Noncategorical	107	(237)
Poverty	214	(387)
Limited English proficient	(127)	(162)
Gifted education	7	(30)
Homeless education	(77)	204
Total shared resource allocation	331	(549)
Combined allocation (school budget and shared resources)	323	(432)

NOTE: Parentheses are used to indicate negative values.
SOURCE: Computed by authors from DPS data.

Accounting for shared resource costs will require some upfront investment, in part to modify current financial software and reporting systems. More significant, perhaps, would be the costs of tracking spending by the new "cost objects." For instance, recording how itinerant staff spends time between schools and how district administrators allocate services school to school creates an additional workload and, consequently, additional expense for the district.

Other implementation challenges revolve around the ability of districts to actually collect information to plug into the model. It is time intensive to collect data on shared resource costs and, as a result, efficiency and efficacy questions must be addressed. Under current systems, data collection is not straightforward, and multiple information sources must be tapped to learn, for example, how Title I money is distributed versus bilingual education spending. Streamlining the accounting process and identifying clear priorities for accounting information is a critical first step in implementation of any new model.

Furthermore, for the model to be useful, districts must ensure that all, or a majority of, shared district resources are measured. Sidestepping accounting challenges by over-categorizing resources as "district leadership and operations" will hinder efforts to capture more spending in school-cost comparisons. As evidenced by school-based, and student-based, budgeting formulas, funding equity cannot be assessed if only a small portion of resources are examined (Miles and Roza 2004). As evidenced by our analysis of DPS data here, it is difficult to make unequivocal statements about equity when only 25 percent of central office shared resources were allocated.

Conclusion

A shared district resource cost allocation model enables more meaningful school-level spending comparisons in that a greater portion of district costs are captured in the school's allocation. Application of the model to the DPS allowed greater understanding of how 25 percent of the central office budget was utilized; we know which schools received shared resources and we know how those resources were spent by student type. A comparison of two middle schools demonstrated significant variation in school spending caused by the inequitable distribution of shared district resources.

There is a clear need for a methodology that accounts for shared district resources and tracks the distribution of these funds. This model has the potential to inform resource allocation decisions because it reveals a more complete school-by-school funding picture. Such information can inform discussions about the variance between intended and actual school funding levels and help decisionmakers as they grapple with the tradeoffs of funding one program over another. Additionally, a greater understanding of how to account for central office resources has the potential to make within-district equity analysis more reliable. Lastly, with this kind of accounting, researchers and policymakers can better compare the cost of different types of schools, including charters, magnets, and alternatives, with better insight into their access to shared resources. Without establishing and implementing a model to include shared resources in school-level analysis, researchers, policymakers, and practitioners will continue to see an eclipsed view of the resources directed to our schools.

References

Berne, R., and Stiefel, L. (1994). Measuring Equity at the School Level: The Finance Perspective. *Educational Evaluation and Policy Analysis, 16*(4): 405–421.

Chambers, J.G. (1999). *Measuring Resources in Education: From Accounting to the Resource Cost Model Approach* (NCES 1999–16). U.S. Department of Education. Washington, DC: National Center for Education Statistics Working Paper.

Cooper, R., and Kaplan, R.S. (1999). *The Design of Cost Management Systems*. Upper Saddle River, NJ: Prentice Hall.

Coopers & Lybrand LLP. (1994). *Resource Allocation in the New York City Public Schools*. New York City: Author.

Cross City Campaign for Urban School Reform. (2001). *Annual Decentralization Progress Comparison Across Ten Cities 2001*. Retrieved August 17, 2004, from http://www.crosscity.org/pdfs/10city2001.pdf.

Educational Testing Services. (2004). *Survey: Americans Seek Fairness in Public School Funding*. Retrieved August 17, 2004, from http://www.ets.org/news/04070101.html.

Fowler, W.J., Jr. (2001). Financial Reporting in the New Millennium. In S. Chaikind and W.J. Fowler, Jr. (Eds.), *Education Finance in the New Millennium. Yearbook of the American Education Finance Association, 2001* (pp. 35–51). Larchmont, NY: Eye On Education.

Hartman, W.T., Bolton, D.G., and Monk, D.H. (2001). A Synthesis of Two Approaches to School-Level Financial Data: The Accounting and Resource Cost Model Approaches. In W.J. Fowler, Jr. (Ed.), *Selected Papers in School Finance, 2000–01* (NCES 2001–378) (pp. 77–119). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Horngren, C.T., Data, S.M., and Foster, G. (2002). *Cost Accounting: A Managerial Emphasis* (11th ed.). Upper Saddle River, NJ: Prentice Hall.

Iatarola, P., and Stiefel, L. (2003). Intradistrict Equity of Public Education Resources and Performance. *Economics of Education Review, 22*(1): 69–78.

Miles, K.H., and Roza, M. (2004). *Understanding Student-Based Budgeting as a Means to Greater School Resource Equity.* Working paper.

National Forum on Education Statistics, Core Finance Data Task Force. (2003). *Financial Accounting for Local and State School Systems: 2003 Edition* (NCES 2004–318). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Odden, A., Archibald, S., Ferminick, M., and Gross, B. (2003). Defining School-Level Expenditure Structures That Reflect Educational Strategies. *Journal of Education Finance, 28*(3): 323–356.

Odden, A., and Busch, C. (1998). *Financing Schools for High Performance: Strategies for Improving the Use of Educational Resources.* The Jossey-Bass Education Series. San Francisco: Jossey-Bass.

Roza, M., and Hill, P.T. (2004). How Within-District Spending Inequities Help Some Schools to Fail. In D. Ravitch (Ed.), *Brookings Papers on Education Policy* (pp. 201–227). Washington, DC: Brookings Institution Press.

Rubenstein, R., Schwartz, A.E., and Stiefel, L. (2004). *From Districts to Schools: The Distribution of Resources Across Schools in Big City School Districts*. Working paper presented at annual conference of the American Education Finance Association, Salt Lake City, UT.

Best Schools, Worst Schools, and School Efficiency: A Reconciliation and Assessment of Alternative Classification Systems

Leanna Stiefel
Hella Bel Hadj Amor
Amy Ellen Schwartz
Robert F. Wagner Graduate School of Public Service
New York University

About the Authors

Leanna Stiefel, Professor of Economics at New York University's Wagner Graduate School of Public Service, teaches courses in applied econometrics, economics of education, and microeconomics. Her current research includes measurement of school efficiency, achievement of immigrant students, effects of school organization on student achievement, links between high school characteristics and college performance, and test score gaps within schools. She is author of *Statistical Analysis for Public and Non-Profit Managers* (1990), coauthor of *The Measurement of Equity in School Finance* (1984), coauthor of *Measuring School Performance and Efficiency* (2005) and publishes regularly in journals such as *Educational Evaluation and Policy Analysis*, *Economics of Education*, and *Journal of Human Resources*. She is past president of the American Education Finance Association, a member of the National Center of Education Statistics Technical Planning Panel, a member of the policy council of the Association for Public Policy and Management, and a governor on the New York State Education Finance Research Consortium. She received her Ph.D. in economics from the University of Wisconsin-Madison (1972) and her A.B. degree, with high honors, from the University of Michigan-Ann Arbor (1967); she also holds an Advanced Professional Certificate in Finance from New York University's Stern School of Business (1984). She can be contacted at leanna.stiefel@nyu.edu.

Hella Bel Hadj Amor is a research scientist at the Institute of Education and Social Policy at New York University. She obtained her Ph.D. in Public Administration from New York University's Wagner Graduate School of Public Service in January 2005. She specializes in education policy, public finance, and applied econometrics. She holds an M.A. in economics from New York University and a B.S. in applied economics from the University of Paris IX Dauphine. She has worked on various projects dealing with issues of education policy and education finance. In particular, she has worked for the Campaign for Fiscal Equity and interned at the regional mission of the World Bank in Bangkok, Thailand, and the National Bank of Cambodia. Her recent research deals with school performance measurement and the role of high schools in the transition from high school to college. Her dissertation work deals with the cost of education in Ohio and the impact of the size and structure of Ohio school districts on school performance. She can be contacted at hella.bel@nyu.edu.

Amy Ellen Schwartz, Professor of Public Policy, Education, and Economics, teaches public economics, finance, and policy at the Wagner School and The Steinhardt School of Education. An applied microeconomist, her research focuses on state and local governments and urban policy, particularly education policy and finance. Ongoing projects

in K–12 education consider the education of immigrant children in New York City; the causes and consequences of educational disparities across racial and ethnic groups; and the measurement of school performance. Research in higher education examines the cost of college and investigates the relationship between high school experiences and success in college. Her research in urban economic development examines the impact of subsidized housing on property values, the relationship between schools and neighborhoods, and the roles of Business Improvement Districts in economic development. Previous research has evaluated the role of public infrastructure in determining state output, growth, and employment, and other issues in public finance. Professor Schwartz's research has been published in a wide range of academic journals, including the *American Economic Review*, *The Journal of Human Resources, Educational Evaluation and Policy Analysis*, and *Journal of Public Economics*. In addition, Professor Schwartz has consulted on various issues of economic and tax policy for nonprofit organizations and governments. She is a member of the Board of Directors of the American Education Finance Association and the Director of Research for New York University's Institute for Education and Social Policy. Professor Schwartz received her Ph.D. in economics from Columbia University. She can be contacted at amy.schwartz@nyu.edu.

The papers in this publication were requested by the National Center for Education Statistics, U.S. Department of Education. They are intended to promote the exchange of ideas among researchers and policymakers. The views are those of the authors, and no official endorsement or support by the U.S. Department of Education is intended or should be inferred.

Best Schools, Worst Schools, and School Efficiency: A Reconciliation and Assessment of Alternative Classification Systems

Leanna Stiefel
Hella Bel Hadj Amor
Amy Ellen Schwartz
Robert F. Wagner Graduate School of Public Service
New York University

Acknowledgments

This research benefited from the generous support of the U.S. Department of Education, Institute for Education Sciences, through Field-Initiated Studies Grant #R305T010115. All responsibility for omissions and errors is ours.

I. Introduction

While academic researchers and policymakers debate the relative merits of ranking public schools and the alternative methodologies for doing so, classifications of public schools have become a feature of the educational landscape. In many cases, the goal is to distinguish the "best" schools or the "worst" schools from the others—for rewards or sanctions, for intervention, or as a guide for parents or students. In New York City, for example, a local not-for-profit education advocacy organization, Advocates for Children of New York, Inc. (AFC), has published two guides to the best public schools (Hemphill 1999, 2002). At the same time, the New York City school district recently identified a set of schools performing so well that they were exempted from a systemwide curriculum and governance reform. At the other end of the spectrum, both New York State and the federal government have identified a set of schools performing so poorly as to require special interventions to spur improvement.

Interestingly, while these lists of best and worst schools differ in their criteria, data, and methodology, none explicitly considers the efficiency with which these public schools use their resources. That is, they effectively ignore the cost of the schools to the taxpayers. Thus, these measures alone may not provide useful guidance to school districts facing resource constraints. As an example, if the "best" schools achieve their high performance because they have garnered especially generous budgets—through grants or donations, perhaps—then looking to them for best practices to replicate in more modestly funded schools may well lead to disappointing results. In this paper, we compare these four lists of best and worst New York City public schools, both to one another and to lists grounded in efficiency measures. We explore the characteristics of schools classified as "best" or "worst," those in which the different methods agree, and those in which they disagree. We then discuss the implications for public policy.

The rest of this paper is organized as follows. The lists of best and worst schools are described individually in section II. In section III, we present the two best schools lists and the two worst schools lists. Section IV

presents current research on the measurement of school performance in a context of increased accountability. In section V, we provide an overview of the New York City public schools and the institutional context in which they operate. In section VI, the efficiency-based lists of schools are introduced, and they are compared with schools in the other lists. In the sixth section, we also conclude with the implications of the results for policymakers and school system participants.

II. Lists of Best and Worst Schools

A. The "Best" Schools

New York City's Best

AFC promotes quality and equal public education services for New York City's poorest families and children who are at greatest risk of discrimination and failure in schools. AFC provides legal services, technical assistance, training about children's educational entitlements and due process rights, organizing, research, and policy analysis.[1]

"Children First" is Mayor Michael Bloomberg's blueprint for reforming the governance and curriculum of the New York City public schools.

Perhaps the most well thought out, highly regarded, accessible information for parents looking for public schools in New York City is the set of guides written by Clara Hemphill of AFC (Hemphill 1999, 2000). Hemphill interviewed teachers and parents, observed schools, and examined school statistics to gather information on atmosphere, homework, student stress, competition among students, quality of the teachers, condition of the building, safety records, class size, test scores, ethnic diversity, admission requirements, and teaching methods.[2] The schools in the books may not be "the best," especially since there is not a formula for selecting them—Hemphill did not necessarily pick schools with one particular feature (high test scores) but a combination of features (e.g., nice building)—but she advertises them as the best.

Hemphill's method is similar to that used in a number of well-known college guidebooks. For example, Fiske (2002) ranks colleges and universities by selecting over 300 of the "best and most interesting institutions in the nation" (out of more than 2,000).[3] At the heart of his methodology is a ranking along three lines, academics, social life, and quality of life, on a scale from one to five.[4] Fiske states that these classifications are subjective and general and that they summarize a write-up for each school that includes information on academics, campus setting, the student body, financial aid, housing, food, social life, and extracurricular activities.[5] Likewise, Hemphill's choice of best schools contains subjective, judgment factors.

Schools Exempted From the "Children First" Instructional Approach

"Children First" is Mayor Michael Bloomberg's blueprint for reforming the governance and curriculum of the New York City public schools beginning with academic year 2002–03. One component of Children First is a new systemwide instructional approach in reading and math curricula, which was phased in starting in September of 2003.[6] The Chancellor's schools are (the only) schools that have received a waiver based on designated performance criteria in reading/writing and math and will not be required to

[1] For more information, see http://advocatesforchildren.org.

[2] This list of schools may be obtained from two sources: Hemphill (2002) and Hemphill (1999). Information on individual schools may be found at http://www.insideschools.org.

[3] Criteria for selection include more than academic strength: there is an effort to achieve geographical diversity and a balance of public and private institutions and to include schools that offer popular or unusual programs and schools that have experienced recent improvements.

[4] Academics include the academic climate of the institution, its reputation, the quality and the seriousness of the faculty (as teachers and researchers) and students, and the quality of facilities and services such as libraries. Quality of life may reflect the level of competition among students, the nature of the social system, the community, the campus, and its location.

[5] Additional information includes the male/female ratio on campus, the range of SAT and ACT scores, the percentage of applicants who are accepted and enroll, and the percentage of students who return and graduate.

[6] http://www.nycenet.edu/childrenfirst/faqs.asp.

implement the new instructional approach. Initially, in April of 2003, 209 schools received such a waiver; they are called Chancellor's schools. [7,8,9]

B. The "Worst" Schools

No Child Left Behind Schools

Prior to 2002, the traditional role of the federal government in education had been to provide aid to disadvantaged pupils and fund research and development. The No Child Left Behind Act (NCLB), which President Bush signed into law in January of 2002, expanded the role of the federal government to stimulate states to raise the achievement of low-performing students. NCLB emphasizes accountability, provides for investment in effective instructional techniques and reform, stresses reading, and allows a more flexible use of federal school funds. It also involves greater choice for parents and students, particularly in instances where students are attending low-performing schools: in such instances, students can transfer out of the low-performing schools while retaining their Title I funds.[10]

The emphasis on accountability, and its accompanying rewards and sanctions, is key. States must put in place comprehensive accountability systems based on ambitious standards in reading and mathematics, annual testing for all students in grades 3 through 8, and annual statewide progress objectives for subgroups based on poverty, ethnicity, disability, and language ability. Practically speaking, a performance index is calculated for each school based on its test scores. If a school's performance index falls below the state standard, then it is assigned an adequate yearly progress goal (AYP), which consists of a set of targets that it needs to reach in the following 3 years to get above the standard. Schools and districts that fail to meet their AYP within 2 years are subject to improvement, and those that fail within 3 years are in need of corrective action.[11] Restructuring measures are taken to make them achieve their goal. Schools that fulfill or go beyond their AYP objectives and those that close performance gaps are rewarded. We call NCLB schools the failing schools under these guidelines.[12]

Under NCLB, if a school's performance index falls below the state standard, then it is assigned an adequate yearly progress goal (AYP).

Schools Under Registration Review (SURR)

The New York State Education Department requires that all schools operating in the state be registered. Schools that are farthest from meeting the state's performance standards are in danger of having their registration revoked if they fail to show adequate

---

[7] This list may be found at http://www.nycenet.edu/PRESS/02-03/HS_CSD_List_by_District_and_school.htm.The following quote from an article describes the criteria for selecting the original exempted schools: "Under the formula to get onto the list, each school was put into a high, middle, or low poverty category, based primarily on the number of free lunches students qualify for. Each school was then scored based on test results and modifying factors such as the number of special education students, non-native English speakers, and recent improvements. The top 20 percent of each category made the list, and schools facing more challenges didn't have to score as high to get on." (Yan 2003) (In fact, 209 schools received waivers, not 208.)

[8] Note that these descriptions reflect the way schools are *said* to be chosen based on these methods, which may differ from what was actually implemented. The number of schools picked is indeed peculiar.

[9] It is anticipated that additional schools will receive waivers as the new systemwide curriculum produces results. In addition, the Chancellor's office has established a petition process whereby schools or programs within schools can request a waiver from either the comprehensive literacy or math portions of the new curriculum—or both. Schools petitioning for waivers will be evaluated based on past and expected student performance as well as the rationale for their request. A list of the schools that petitioned for the waivers and received them as of April 2003 may be found at http://www.nycenet.edu/SchoolsGrantedWaivers.pdf. Note that some of these schools received waivers only in math, some only in reading, and some in both subjects.

[10] More information may be found at http://www.ccsso.org/federal_programs/NCLB/index.cfm and http://www.ed.gov/nclb/overview/intro/execsumm.html.

[11] *Assessment and Accountability in the Fifty States: 1999–2000, New York.* Consortium for Policy Research in Education (CPRE) (July 2000) http://www.cpre.org/Publications/ny.pdf.

[12] A list of NCLB schools as of April 2003 is available on the New York City Department of Education website at http://www.nycenet.edu/nclb/PSChoice.asp.

improvement within 3 years. Schools that have been warned that their registrations may be revoked are considered "Schools Under Registration Review" (SURR).[13]

All public schools are expected to have at least 90 percent of their students scoring above the state standard for their grade on state standardized tests and a dropout rate of less than 5 percent. SURR schools are identified as schools that are farthest from meeting this standard. As an example, in the 1998–99 school year, schools "farthest from the state standard" on a standardized test included schools in which less than two-thirds of the students performed at or above level 2 on the grades 4 and 8 English language arts examination and the grade 4 mathematics examination, and schools in which less than a third of the students performed at or above level 2 on the grade 8 mathematics exam.[14,15] SURR schools are also schools that have a "poor learning environment." Such an environment is one in which the school is the subject of persistent parent complaints or where conditions threaten the health, safety, or educational welfare of students (such as high rates of absenteeism or a high level of violence).[16]

New York State schools that have been warned that their registrations may be revoked are considered "Schools Under Registration Review" (SURR).

The three government lists of best and worst schools would be labeled "high stakes" by most observers. If the lists fail to show significant overlap or if they diverge significantly from the efficiency-based lists, then it will be important to discuss openly the costs of errors of inclusion or omission of schools that might be "misclassified."

III. Literature on Measuring School Performance

Efforts toward identifying good schools is not a new phenomenon. Thirty years ago, in response to the Coleman report (1966), the Effective Schools Movement argued that, while family background matters, schools also play an important role in children's learning (Edmonds 1982). More specifically, a number of features common to schools that successfully educated students with diverse backgrounds were believed to be both associated with school success and under the control of the school system: a strong emphasis on high-quality, focused instruction supported at the highest levels of the school hierarchy; high expectations for all students and regular evaluation of their performance; and safe, well-organized schools (Edmonds 1979, 1982). Some precepts of the Effective Schools Movement are timely today in that they advocate holding schools accountable and reporting disaggregated measures of student performance to verify school success for students of different sexes, races, and poverty levels.[17]

More recently, the measurement of school performance for the purpose of school accountability has been the subject of a small but growing research literature. This literature addresses the properties of a "good" performance measure, as well as the features of the school system that threaten that measure. For example, numerous authors analyze features of measures used for accountability. According to Hanushek and Raymond (2002) and Ladd (2002), the quality of a performance measure depends on whether it reflects the material covered in the classroom and the performance of all the stakeholders—students, teachers, and administrators. They go on to say that, in order to promote accountability, a performance measure should provide a balance between challenge and feasibility. This, in turn, depends on the choice of levels of performance targets or rates of improvement and the incentives and disincentives the different options create. Feasibility also depends on the data requirements of a measure and the possible impact of error.

[13] Note that, by definition, the SURR schools are a subset of the NCLB schools such that we expected to find a substantial overlap, yet in our best effort to identify these schools in our sample, they are not. We continue to explore this perplexing find.

[14] Level 2 is the second lowest of four.

[15] Kadamus (2000). Lists of SURR schools may be obtained from the New York State Education Department website at http://www.emsc.nysed.gov/nyc/regrev.html#SURRList/2001AugSURRlist.pdf or http://www.emsc.nysed.gov/nyc/SURR/SURRJan2003.html.

[16] New York State Education Department Office (1998).

[17] For an update on the Effective Schools Movement today and more detail on the correlates, see http://schools.tdsb.on.ca/albertcampbell/spri/docs/Revolutionary.pdf and http://ali.apple.com/ali_media/Users/1000059/files/others/lezotte_article.pdf.

For many authors, fairness requires that factors over which schools have no control be taken into account when measuring performance. Adjusting performance measures for client and environmental characteristics is generally considered an improvement over raw measures (Rubenstein, Stiefel and Schwartz 2003). However, while it is widely accepted that student poverty raises the educational challenge faced by schools, the effect of, and the appropriateness of including, for example, minority representation in the student body, is more controversial (Clotfelter and Ladd 1996; Ladd 1999, 2002; Ladd and Walsh 2002).

Student mobility, which is especially prevalent in schools with disadvantaged populations, may affect the quality of a performance measure. This problem may be alleviated by comparing specific cohorts across grades (Hanushek and Raymond 2002) or calculating value-added measures only for the students who attend a given school for a minimum number of days during the school year (Ladd and Walsh 2002), rather than comparing successive cohorts in a school. Cohorts may also differ across years if schools exempt some students from testing. Reliance on several tests as well as other measures of school performance, including attendance and dropout rates, as in Dallas (Ladd 1999), helps to circumvent this problem.[18]

Improvements in achievement varied across low- and high-achieving students and across subjects.

Scores can change from one year to the next due to measurement error (Hanushek and Raymond 2002; Kane and Staiger 2002; Ladd 2002; Ladd and Walsh 2002).[19] This is especially true in small schools and when changes or gains are used rather than levels. Averaging and weighting scores over several years, combining data across grades and/or subjects, or using more sophisticated techniques may reduce measurement error at the cost of a loss of transparency.

Several authors study the effects of the measurement of school performance in specific accountability systems on student and school outcomes and, in the process, describe in detail the measurement methods used. An early evaluation of school-based incentive programs on student outcomes is provided in Ladd (1999). She compares gains in student performance in Dallas, which implemented an accountability system starting in 1991, to those in five other large Texas cities and finds that the reform seems to have resulted in positive and relatively large effects for some students in selected grades.

The introduction of high-stakes testing in the Chicago schools raised student achievement, especially for students in the lowest performing schools (Roderick, Jacob, and Bryk 2002). Improvements in achievement varied across low- and high-achieving students and across subjects, which underlines the importance of taking the distributional effects of accountability systems into account in order to fully judge their effectiveness.

School classifications across various performance measures in Dallas and in South Carolina are compared in Clotfelter and Ladd (1996). The authors compare a series of performance measures based on changes in test scores, which turn out to be highly correlated, as well as a number of measures based on residuals. These are correlated as well, but correlations across the two groups are not quite as high.

Overall, researchers agree that a number of choices must be made when designing school accountability systems: whether to use levels, changes, or value-added measures; whether to exempt certain students from testing; whether to take into account factors outside a school's control; etc. There is no consensus on what constitutes a "best" measure, or, more specifically, on the effect of these choices on how accountability systems affect school performance. Despite these issues, early evidence indicates that accountability systems appear to have a positive effect on performance.

[18] Kane and Staiger (2002) state that broadening the range of measures under consideration is also important because the narrow focus on commonly reported subjects such as math and reading disadvantages schools that focus on other outputs.

[19] Ladd and Walsh (2002) focus on measurement error, how it affects school classifications, and how to correct for it. They find evidence of serious measurement error in North Carolina and South Carolina samples, and correcting for this error changes the ranking of schools according to their value-added performance measures. In addition, they find that using averaged residuals or fixed effects as measures of effectiveness seems to cause some changes in the classifications of schools.

IV. Education in New York City

In 2001, New York City's newly elected mayor, Michael Bloomberg, successfully pressed the state to grant control of the city schools to the mayor, beginning in school year 2002–03. Mayor Bloomberg, in association with School Chancellor Joel I. Klein, then launched Children First, a plan to reform governance and curriculum in the New York City public schools over the next several years.

The New York City school system educates over a million students in almost 1,300 schools and programs. About half of these pupils are in elementary schools; about 20 percent are in middle schools and 20 percent in high schools; and the rest are in collaborative or vocational schools or alternative or special education programs. The New York City Department of Education approved a budget of $12 billion for the school year *2002–03,* with a corresponding average cost per pupil of a little above $11,000. In the spring of *2002*, slightly fewer than 40 percent of elementary school students in grades 3 through 8 met or exceeded the grade level on the state and city reading exams, and about 35 percent did the same on the mathematics exam in grades 4 through 7. [20]

We have constructed a rich school-level database using data provided by the New York City Department of Education on elementary schools.

We have constructed a rich school-level database using data provided by the New York City Department of Education on elementary schools. The Annual School Reports and School-Based Budgeting Reports are school-level databases, which we augmented with student-level data in order to construct grade-level variables (e.g., race/ethnicity, immigrant status, free and reduced-price lunch eligibility). Our data contain information on student characteristics, test scores, and school resources for the years 1995–96 through 2000–01. We use a balanced panel of 602 schools with grades 3, 4, and 5 and valid reading and math scores for each grade in each year.[21,22] Descriptive statistics on schools (unweighted by pupils) for the year *2000* are presented in table 1. These statistics are averages across schools and do not take into account differences in school size. Thus, statistics based on students would differ from these.

All test score data are reported as standardized *z*-scores. Data for third and fifth grades come from the CTB/McGraw Hill Test of Basic Skills (CTB) in reading and the California Achievement Test (CAT) in mathematics, while fourth-grade data for 1998–99 and 1999–2000 are from state English Language Arts (ELA) reading and mathematics tests. For comparability, the tests are normalized to citywide averages.[23]

Total expenditure per pupil includes direct services to schools and district and systemwide costs (instructional, administrative, and other). Non-classroom teacher expenditure includes all of these items except classroom instruction.[24] Other direct services encompass instructional support services (counseling and other outreach services, drug prevention programs, after-school activities, parent involvement), school leadership (and their support staff and supplies), ancillary support services (food, transportation, safety, and computer system support), building services (custodial, maintenance, leases, and energy), and district support.

New York City educated about 483,000 students in the 602 sample schools in the 1999–2000 school year. The vast majority of these schools are elementary schools, with almost two-thirds of the schools serving up to grade 5 and almost a third serving up to grade 6. The remaining 9 percent serve grades 7 and 8, as well. The

[20] The figures in this section may be found on the New York City Department of Education website at http://www.ncyenet.edu/Offices/stats/default.htm.

[21] For greater detail on the data, see Schwartz and Zabel (2003) and Schwartz, Stiefel, and Bel Hadj Amor (2003).

[22] Note that descriptions of the Advocates for Children, Chancellor's, NCLB, and SURR schools are limited to such schools for which data are reported by the New York City Department of Education. A count is available from the authors.

[23] Greater detail on the normalizing procedure is available in Stiefel, Schwartz, Bel Hadj Amor, and Kim (2003).

[24] Classroom instruction includes teachers and other educational and classroom staff, textbooks, librarians and library books, instructional supplies, curriculum development, contracted instructional services, and summer and evening school.

Table 1. Descriptive statistics (unweighted) for New York City elementary schools: 1999–2000

N = 602	Mean	Minimum	Maximum
Student characteristics			
Grade 5 mean reading *z*-scores	0.03	–0.85	1.53
Grade 5 mean math *z*-scores	0.04	–1.08	1.35
Grade 5 mean reading and math *z*-scores for poor students	–0.04	–0.93	1.08
Grade 5 mean reading and math *z*-scores for the non-poor	0.41	–2.14	1.78
Grade 5 mean reading and math *z*-scores for Black students	–0.10	–1.54	1.31
Grade 5 mean reading and math *z*-scores for Hispanics	–0.08	–1.47	1.18
Grade 5 mean reading and math *z*-scores for Asian and other	0.39	–1.94	2.22
Percent female students	48.85	41.30	58.20
Percent free lunch students	73.75	6.70	100.00
Percent Black students	35.22	0.10	97.30
Percent Hispanic students	35.30	1.30	98.00
Percent Asian and other students	11.63	0.00	92.50
Percent LEP students	13.56	0.00	57.40
Percent recent immigrant students	6.95	0.00	26.70
Percent students in special education	4.86	0.00	18.40
Percent students in resource room	6.34	1.30	16.90
School characteristics			
Total expenditure per pupil (in dollars)	9,798	5,970	21,893
Non-classroom teacher expenditure (in dollars)	5,653	2,823	17,302
Pupil-teacher ratio	13.76	7.71	20.34
Percent teachers licensed/permanently assigned	81.70	0.00	100.00
Percent teachers with over 5 years' experience	57.66	0.00	93.90
Percent teachers with master's degree	77.27	0.00	100.00
Percent teachers with over 2 years in same school	64.11	0.00	89.70
Enrollment	803	100	2,200
School serves up to grade 5	0.62	0.00	1.00
School serves up to grade 6	0.29	0.00	1.00
School serves up to grade 7	0.00	0.00	1.00
School serves up to grade 8	0.09	0.00	1.00
SURR school	0.03	0.00	1.00
NCLB school	0.22	0.00	1.00
Chancellor's school	0.22	0.00	1.00
AFC school	0.15	0.00	1.00

NOTE: N is smaller for subgroup performance variables because not all schools have students in every subgroup. Here, N for non-poor is 500, for Black, 581, for Hispanic, 598, and for Asian and others, 475.

SOURCE: Authors' calculations based upon data provided by the New York City Department of Education. The list of SURR schools is from the New York State Education Department website. The lists of NCLB and Chancellor's schools are from the New York City Department of Education website. The list of AFC schools is from Hemphill (1999, 2002).

average school enrolls about 800 students; included are some small schools (as low as 100 students) and some very large ones (over 2,000 students).

In 1999–2000, New York City schools spent a little under $10,000 on average on each elementary school student, but there is a wide range across schools, from a low of almost $6,000 to a high of close to $22,000.[25] On average, a little less than half of per pupil spending goes to the classroom. There is a wide range of teacher characteristics across schools, and the average school has over three-quarters of its teachers who are licensed and who hold M.A. degrees, over two-thirds who have been in the same school for at least 2 years, and more than half who have more than 5 years of experience. The range in the number of pupils per teacher is quite wide as well: there can be as few as 8 students per teacher and as many as 20. The average is about 14.

[25] The school with the second highest number spends $16,677 per pupil.

Accordingly, the range of performance is wide. While some schools have average *z*-scores of almost –1, others are as high as 1.44. The average is 0.04. This average masks wide variations across subgroups of students, however. The average *z*-score for poor students (i.e., free lunch students) is –0.04, while that for non-poor students is 0.41. The average *z*-scores for Black and Hispanic students are lower than the whole-school averages (–0.10 and –0.08, respectively) while that for Asian and other non-White students is higher (0.39).[26,27]

There is a wide variation in the representation of poor and minority students in New York City schools. While there are anywhere between 7 and 100 percent poor students, there are no minority students in some schools and close to 100 percent in others. Almost three-quarters of the students are poor and about 82 percent are non-White in the average school. Over 80 percent of the non-White students are divided fairly evenly between Black and Hispanic. On average, seven percent of the students are immigrant students and about 14 percent have limited English proficiency. The representation of students in resource room (that is, receiving part-time special education services) and students in special education is about the same (6 percent and 5 percent, respectively) and their representation is always below 20 percent. About one-half of the students are female.

While there are anywhere between 7 and 100 percent poor students in New York City schools, there are no minority students in some and close to 100 percent in others.

V. Comparing Best and Worst Schools

A. Is Good in the Eye of the Beholder? Advocates for Children Schools Versus Chancellor's Schools

Table 2 compares the best schools that are in our sample to the rest of the city schools. Column (1) presents descriptive statistics for the AFC schools that are not on the Chancellor's list, column (2) for the schools that are on both the AFC and Chancellor's lists, column (3) for the Chancellor's schools that are not on the AFC list, and column (4) for the rest of the city schools.[28] There is a fair amount of agreement regarding which are the best schools when the best schools are compared to the rest of the schools. Clearly, the best schools perform much better than the rest (the average *z*-score is 0.36 to 0.69, compared to –0.14 for the rest). Subgroup *z*-scores show that schools with higher shares of disadvantaged (advantaged) children have lower (higher) average *z*-scores than average.

The best schools are also schools with somewhat more advantaged populations: they have many fewer poor (41 percent to 56 percent), Black (9 percent to 26 percent), and Hispanic students (17 percent to 30 percent) than the rest of the schools (82 percent, 43 percent, and 40 percent, respectively); in 2 out of 3 cases they have fewer LEP students and slightly fewer students in special education. In 2 out of 3 cases, the best schools do, however, have more immigrant students than the rest of the city schools and slightly more students in resource room. The best schools spend slightly less (under $9,600) than the rest of the schools (a little over $10,000) per pupil, yet they tend to have more experienced and educated teachers. They are also smaller than the rest of the schools (722 to 768 students, on average, vs. 824 students in the other schools). The geographical distribution of the schools varies slightly across the subsamples with, generally, more of the best schools in Manhattan, Queens, and Staten Island, and fewer of the best schools in the Bronx and Brooklyn.

While there is a fair amount of agreement regarding which are the best schools when the best schools are compared to the rest of the city schools, the AFC and Chancellor's lists disagree somewhat on which schools are the best schools: of the 92 AFC schools in our sample, 67 are also on the Chancellor's list,

[26] Note that this is close to the average for White students, which is 0.34.

[27] These differences are much greater across students rather than across schools.

[28] A comparison table of all the AFC schools to the rest of the city and all the Chancellor's schools to the rest of the city is available from the authors.

Table 2. Comparisons across best schools (unweighted)

	(1) AFC only N = 25	(2) Overlap AFC+ Chancellor's N = 67	(3) Chancellor's only N = 66	(4) The rest N = 444
Student characteristics				
Average reading and math *z*-score	0.36	0.69	0.44	–0.14
Average reading and math gain	–0.02	–0.07	–0.01	0.02
Average *z*-score for poor students	0.20	0.45	0.33	–0.18
Average *z*-score for non-poor students	0.66	0.92	0.66	0.24
Average *z*-score for Black students	0.20	0.32	0.10	–0.21
Average *z*-score for Hispanic students	0.15	0.41	0.20	–0.22
Average *z*-score for Asian and other students	0.62	0.85	0.68	0.23
Percent free lunch eligible	55.90	41.15	57.24	82.13
Percent Black	25.66	13.32	8.46	43.04
Percent Hispanic	29.66	17.24	24.19	40.00
Percent Asian and others	13.22	23.25	28.95	7.21
Percent LEP	7.50	10.80	16.69	13.85
Percent recent immigrants	5.73	8.72	11.16	6.13
Percent special education	3.75	3.91	4.38	5.14
Percent resource room	6.80	6.52	6.65	6.24
School characteristics				
Total expenditure per pupil (in dollars)	9,506	8,932	8,933	10,074
Non-classroom teacher expenditure (in dollars)	5,517	5,073	4,978	5,849
Pupil-teacher ratio	14.20	15.37	15.07	13.30
Percent teachers licensed/permanently assigned	78.44	90.76	92.70	78.88
Percent teachers with master's degree	75.66	85.10	86.28	74.85
Enrollment	734	722	768	824
Manhattan	0.28	0.21	0.09	0.16
Bronx	0.12	0.01	0.02	0.21
Brooklyn	0.32	0.34	0.30	0.37
Queens	0.20	0.39	0.42	0.21
Staten Island	0.08	0.04	0.17	0.05

NOTE: Again, N is smaller for subgroup performance variables. N ranges from 21 to 25 in column (1), 61 to 67 in column (2), 55 to 66 in column (3), and 299 to 444 in column (4).

SOURCE: Authors' calculations based upon data provided by the New York City Department of Education. The list of SURR schools is from the New York State Education Department website. The lists of NCLB and Chancellor's schools are from the New York City Department of Education website. The list of AFC schools is from Hemphill (1999, 2002).

and 25 are not. Of the 133 Chancellor's schools in our sample, about half (66) are not on the AFC's list. Overlapping schools perform at the highest level (0.69) and the AFC-only schools perform at the lowest (0.36). Interestingly, the AFC-only schools have the highest spending of the three groups (about $9,500 per pupil vs. $8,900 in the others) and the lowest teacher quality, while the Chancellor's-only schools have the highest teacher quality. The proportions of at-risk students vary across the three groups, as does geographical location (with, notably, a low representation of Chancellor's-only schools in Manhattan).

B. Is Bad in the Eye of the Beholder? NCLB Schools vs. SURR Schools

Table 3 presents descriptive statistics for NCLB schools that are not SURR schools (column [1]), schools that are both NCLB and SURR schools (column [2]), SURR schools that are not NCLB schools (column [3]), and the rest of the city schools (column [4]).[29] The worst schools have lower average *z*-scores (–0.24 to –0.59) than the other schools have (0.13). Non-poor students do better than average, as do Asian students (except for those in SURR schools). Poor and Black students do worse than

[29] A comparison table of all NCLB schools to the rest of the city and all the SURR schools to the rest of the city is available from the authors upon request.

average, but Hispanics in SURR or NCLB + SURR schools do not. Not surprisingly, the worst schools have more poor (89 percent to 93 percent), Black (39 percent to 51 percent), Hispanic (48 percent to 58 percent), and LEP (16 percent to 18 percent) students than the rest of the schools (69 percent, 32 percent, 31 percent, and 13 percent, respectively). They also have slightly more students in special education and students in resource room in 2 out of 3 cases. They do have fewer immigrant students (3 percent to 6 percent vs. 8 percent in the rest of the schools). The worst schools spend more per pupil ($10,000 to $12,000) than the other schools (under $10,000) and they have less desirable teacher characteristics. They are larger than the other schools (837 to 966 vs. 784 students, on average). More worst schools than other schools are located in the Bronx and fewer worst schools are in Brooklyn and Queens.

There are differences across the worst schools. Eight of the 130 NCLB schools in our sample are also SURR schools; the other 122 are not. Roughly the other half of the SURR schools (10) are not NCLB schools. All of the worst schools have lower performance levels than the other schools have, and the NCLB schools have the highest performance levels of the worst schools. The SURR schools are the highest spenders among the worst schools, with $12,000 per pupil on average vs. $11,000 in the overlapping schools and $10,000 in the NCLB schools. The latter also have more advantageous teacher characteristics and the lowest proportion of at-risk students in most categories. The overlapping schools are largest (966 students on average) and the SURR schools are smallest (837 students). The distribution of worst schools across boroughs is very different in each group, with the NCLB schools being the most evenly distributed.

All of the worst schools have lower performance levels than the other schools have, and the NCLB schools have the highest performance levels of the worst schools.

VI. An Economic Approach Based on Efficiency

A. Education Production Functions

This section describes one quantitative technique that we have developed to rank schools according to their performance and resources, education production functions (EPFs).[30] This method provides a measure of efficiency that is used to identify the best and worst schools. These lists of schools can then be compared to the others, such as AFC schools, Chancellor's schools, NCLB schools, and SURR schools. This technique adjusts for features of the environment and resource availability. It relies on much stronger theoretical underpinnings than the methodologies used to put together the aforementioned lists, and it uses more data, in addition to having a different conceptual base.

EPFs have their roots in economic input-output theory, according to which a school (much like a firm) combines inputs to produce maximum educational output. Accordingly, this method takes into account the inputs that produce education (students and resources, primarily) and thus controls for differences in these inputs across schools. More specifically, it is generally agreed that schools should not be held accountable for resources that are not under their control and school efficiency should not reflect the level of the inputs, but rather the work the schools are doing with these resources.[31]

An EPF is a regression-based technique with a measure of output as the dependent variable and inputs as the independent variables. Theory dictates the choice of inputs and functional form, which may call for the use of nonlinearity and other options.[32] Levels, changes, or

[30] Three other research-based methods that can be used are adjusted-performance measures (APMs), cost functions, and Data Envelopment Analysis (DEA). For more on comparisons of classifications across the four analytical methods, see Rubenstein, Stiefel, Schwartz, and Bel Hadj Amor (2003). The *New York Times* published a version of school-level performance measures based on regression equations for several years. See, for example, Josh Barbanel (1999a, 1999b, 1999c, 2000), who used test scores as outcome measures with statistical controls for income and sometimes English proficiency. Each school was compared to other schools with a similar mix of students.

[31] See Levin (1975) for an early study using this framework.

[32] For more information on EPFs, see Hanushek (1986, 1996), Ferguson and Ladd (1996), and Schwartz and Zabel (2003).

Table 3. Comparisons across worst schools (unweighted)

	(1) NCLB only N = 122	(2) Overlap NCLB + SURR N = 8	(3) SURR only N = 10	(4) The rest N = 462
Student characteristics				
Average reading and math z-score	–0.24	–0.59	–0.58	0.13
Average reading and math gain	0.00	0.03	0.06	0.00
Average z-score for poor students	–0.25	–0.59	–0.59	0.04
Average z-score for non-poor students	0.11	–0.09	0.39	0.48
Average z-score for Black students	–0.25	–0.59	–0.62	–0.04
Average z-score for Hispanic students	–0.31	–0.55	–0.52	–0.01
Average z-score for Asian and other students	0.18	–0.45	–0.66	0.47
Percent free lunch eligible	89.27	93.36	89.94	68.96
Percent Black	44.94	38.84	50.51	32.26
Percent Hispanic	48.17	57.96	47.58	31.25
Percent Asian and others	4.01	1.65	1.26	14.03
Percent LEP	15.78	17.98	17.82	12.80
Percent recent immigrants	5.66	3.48	2.69	7.45
Percent special education	5.54	4.85	5.49	4.67
Percent resource room	6.38	6.48	5.71	6.34
School characteristics				
Total expenditure per pupil (in dollars)	10,208	10,559	11,836	9,633
Non-classroom teacher expenditure (in dollars)	5,985	6,314	7,074	5,523
Pupil-teacher ratio	13.13	12.01	11.66	14.00
Percent teachers licensed/permanently assigned	76.68	70.85	66.17	83.54
Percent teachers with master's degree	73.96	66.30	70.59	78.48
Enrollment	860	966	837	784
Manhattan	0.18	0.00	0.20	0.16
Bronx	0.30	0.75	0.60	0.11
Brooklyn	0.30	0.00	0.20	0.38
Queens	0.19	0.13	0.00	0.28
Staten Island	0.03	0.13	0.00	0.07

NOTE: For the subgroup performance variables, N ranges from 76 to 122 in column (1), 4 to 8 in column (2), 3 to 10 in column (3), and 364 to 462 in column (4).

SOURCE: Authors' calculations based upon data provided by the New York City Department of Education. The list of SURR schools is from the New York State Education Department website. The lists of NCLB and Chancellor's schools are from the New York City Department of Education website. The list of AFC schools is from Hemphill (1999, 2002).

gains may be used, with the change or gain as the dependent variable or the level as the dependent variable and an option to include a measure of prior performance as an independent variable.

Ideally, an EPF is estimated with a panel of data, rather than a cross-section, so that the efficiency measure is the coefficient on a school fixed effect. In other words, a series of dummy variables (fixed effects), one for each school, are included in the model and the coefficients on these variables measure the difference in performance between each school and the reference school. The larger the fixed effect coefficient, the greater the efficiency. Fixed effects reduce omitted variable bias by controlling for time-invariant factors specific to each school. It is important to note, however, that fixed effects reduce and do not eliminate omitted variable bias, such that each measure of efficiency still includes some other school factors.[33] A typical EPF may look as follows:

[33] It is possible to "purge" the fixed effects of some time-invariant characteristics, such as location, by running a second regression, where the fixed effects are the dependent variable (see Schwartz and Zabel 2003).

$$Y_{gst} = \beta_0 + \beta_1 Y_{g-1,s,t-1} + \beta_{2i}\sum_i student\_characteristics_{jgst} + \beta_{3j}\sum_j school\_characteristics_{jgst} + efficiency_s + e_{gst}$$

where Y is a measure of output, e is an error term with the usual properties, and g indicates grade, s school, and t time.

How do the lists of best and worst schools, Advocates for Children, Chancellor's, NCLB, and SURR schools, compare to the EPF lists?[34] The EPFs are estimated using a balanced panel of 602 elementary schools for the years 1995–96 through 2000–01. More specifically, they are computed using fifth-grade reading performance; enrollment and student characteristics are for the fifth grade as well.[35] The other variables are measured at the school level. There are 158 schools that are Chancellor's and/or Advocates for Children and 140 schools that are NCLB and/or SURR. For comparison purposes, we divide the New York City schools into three comparison groups: the 160 schools that are ranked lowest; the 160 ranked highest, according to the EPFs; and the schools that are in between.

There is some agreement between best and efficient schools. Indeed, table 4 indicates that few of the best schools are highly inefficient (top panel, row C). Yet, being one of the best schools in the city does not necessarily imply being one of the most efficient, and the extent of the overlap between "bestness" and efficiency varies by subgroup. More specifically, setting aside the few least efficient schools, over two-thirds of the remaining overlapping best schools are highly efficient (row A) while about one-third are not (row B); a little under half of the Advocates for Children schools are highly efficient (row A), while the majority of the Chancellor's schools are not (row B).

While there is some agreement between worst and inefficient schools, some of the worst schools are not* highly *inefficient.

Similarly, while there is some agreement between worst and inefficient schools, some of the worst schools are not *highly* inefficient. Indeed, table 4 indicates that few of the worst schools are highly efficient (bottom panel, row D).[36] While most overlapping and SURR schools are highly inefficient (6 out of 8, and 7 out of 10, respectively [row F]), there are fewer highly inefficient NCLB schools (37 [row F]) than there are non-highly inefficient NCLB schools (59 [row E]).

B. The Best Schools

Why were some of the city's most efficient schools not "good enough" for the Advocates for Children and Chancellor's lists? In table 5, the most efficient best schools (column [1]) are compared to the rest of the most efficient schools (column [2]). Almost half of the 160 most efficient schools (71) are among the best New York City schools, leaving 89 schools that are highly efficient and yet did not meet the Advocates for Children nor the Chancellor's criteria. How do these schools differ? The schools that were left out of the best lists do not perform as well as the other schools (0.02 vs. 0.61), although they have improved (the average gain is 0.05 vs. –0.04). And while it costs them $120 more per pupil to achieve this performance, they obtain it with more disadvantaged student populations (significantly more poor, Black, Hispanic, and LEP students) and lower teacher quality in larger schools.[37] Performance levels in the non-best schools are lower than in the other schools for all subgroups of students.[38] Thus, the efficient schools

[34] Information on Adjusted Performance Measures is available from the authors.

[35] Descriptive statistics for the fifth grade are available upon request.

[36] In fact, none of the SURR schools (whether or not they are NCLB) are highly efficient (row D).

[37] A version of table 5 that breaks down the best schools into AFC-only, overlapping, and Chancellor's-only schools is available from the authors. Overall conclusions remain consistent, except that the AFC schools have higher spending and lower teacher quality than all other highly efficient schools (best or not) and they achieve the highest gain.

[38] Only 10 of the city's best schools are among the least efficient schools. These are schools that perform better than the other least efficient schools on both performance measures and achieve this performance with more advantaged student populations (the difference in the representation of Black students is striking, 11 percent in the best schools vs. 55 percent in the other schools). They are relatively small schools (417 vs. 659 students on average) that spend less than the rest ($10,551 per pupil vs. $11,312), yet have higher teacher quality (results available upon request, including for the Advocates For Children/Chancellor's breakdown).

Table 4. Cross tabs of best and worst schools and efficiency

	The best				
	AFC only	Overlap AFC + Chancellor's	Chancellor's only	The rest	Total
A Most efficient					
Frequency	10	45	16	89	160
Percent	1.66	7.48	2.66	14.78	26.58
Row percent	6.25	28.13	10.00	55.63	
Column percent	40.00	67.16	24.24	20.05	
B In between					
Frequency	13	21	43	205	282
Percent	2.16	3.49	7.14	34.05	46.84
Row percent	4.61	7.45	15.25	72.70	
Column percent	52.00	31.34	65.15	46.17	
C Least efficient					
Frequency	2	1	7	150	160
Percent	0.33	0.17	1.16	24.92	26.58
Row percent	1.25	0.63	4.38	93.75	
Column percent	8.00	1.49	10.61	33.78	
Total					
Frequency	25	67	66	444	602
Percent	4.15	11.13	10.96	73.75	100.00

	The worst				
	NCLB only	Overlap NCLB + SURR	SURR only	The rest	Total
D Most efficient					
Frequency	26	0	0	134	160
Percent	4.32	0.00	0.00	22.26	26.58
Row percent	16.25	0.00	0.00	83.75	
Column percent	21.31	0.00	0.00	29.00	
E In between					
Frequency	59	2	3	218	282
Percent	9.80	0.33	0.50	36.21	46.84
Row percent	20.92	0.71	1.06	77.30	
Column percent	48.36	25.00	30.00	47.19	
F Least efficient					
Frequency	37	6	7	110	160
Percent	6.15	1.00	1.16	18.27	26.58
Row percent	23.13	3.75	4.38	68.75	
Column percent	30.33	75.00	70.00	23.81	
Total					
Frequency	122	8	10	462	602
Percent	20.27	1.33	1.66	76.74	100.00

SOURCE: Authors' calculations based upon data provided by the New York City Department of Education. The list of SURR schools is from the New York State Education Department website. The lists of NCLB and Chancellor's schools are from the New York City Department of Education website. The list of AFC schools is from Hemphill (1999, 2002).

Table 5. Most efficient schools: The best versus the rest (unweighted)

	(1) The best N = 71	(2) The rest N = 89
Student characteristics		
Average reading and math z-score	0.61	0.02
Average reading and math gain	–0.04	0.05
Average *z*-score for poor students	0.38	–0.03
Average *z*-score for non-poor students	0.91	0.39
Average *z*-score for Black students	0.25	–0.03
Average *z*-score for Hispanic students	0.33	–0.07
Average *z*-score for Asian and other students	0.82	0.35
Percent free lunch eligible	49.69	80.54
Percent Black	16.35	29.50
Percent Hispanic	26.53	54.36
Percent Asian and others	23.22	8.73
Percent LEP	12.53	20.29
Percent recent immigrants	8.84	7.83
Percent special education	4.01	4.15
Percent resource room	6.40	5.73
School characteristics		
Total expenditure per pupil (in dollars)	8,717	8,837
Non-classroom teacher expenditure (in dollars)	4,893	4,976
Pupil-teacher ratio	15.41	14.50
Percent teachers licensed/permanently assigned	88.25	79.73
Percent teachers with master's degree	83.00	75.45
Enrollment	792	1,057

NOTE: For the subgroup performance variables, N ranges from 67 to 71 in column (1) and 67 to 89 in column (2).

SOURCE: Authors' calculations based upon data provided by the New York City Department of Education.

that do not make the Advocates for Children or Chancellor's lists do well with their clientele, but on an absolute level, not as well as schools with an easier clientele.

C. The Worst Schools

Just as some highly efficient schools were not good enough to make the AFC or Chancellor's lists, some highly inefficient schools were not considered "bad enough" to be included in the NCLB or SURR lists. What distinguishes these schools from the other inefficient schools? In order to address this question, the least efficient, worst schools (table 6, column [1]) are compared to the other least efficient schools (column [2]). About a third of the 160 least efficient schools are among the worst New York City schools. These schools have lower performance in levels and in gains than the other least efficient schools (–0.43 and –0.04 vs. –0.25 and –0.01, respectively). The lower performance level holds for most subgroups. These schools educate more poor (90 percent), Black (63 percent), and Hispanic (33 percent) students than the other inefficient schools do (80 percent, 48 percent, and 31 percent, respectively) with higher spending, per pupil (by about $200) and lower teacher quality. [39,40]

Thus, interestingly, whenever two groups of schools are compared, the "worst" of the two sets tends to have lower

[39] A version of table 6 that breaks down the least efficient worst schools into NCLB-only, overlapping, and SURR-only schools is available from the authors upon request. Once again, conclusions are about the same. Notably, the overlapping and SURR schools achieve higher gains and the Advocates for Children schools much lower gains than the rest of the City's least efficient schools.

[40] Twenty-six of the 160 most efficient schools in the sample are among the City's worst schools, and they are all NCLB-only schools. Compared to the rest of the highly efficient schools, they have lower performance in levels (–0.18 vs. 0.37, and this is true of subgroups as well) although not in gains; more poor, Hispanic, and LEP students (92 percent, 75 percent, and 27 percent vs. 62 percent, 36 percent, and 15 percent, respectively, in the other schools); and lower teacher quality in spite of higher spending ($9,126 vs. $8,717). Notably, they have a lower share of Black students than the other schools (19 percent vs. 25 percent). Results are available from the authors.

Table 6. Least efficient schools: The worst versus the rest (unweighted)

	(1) The worst N = 71	(2) The rest N = 89
Student characteristics		
Average reading and math *z*-score	–0.43	–0.25
Average reading and math gain	–0.04	–0.01
Average *z*-score for poor students	–0.44	–0.30
Average *z*-score for non-poor students	0.04	0.10
Average *z*-score for Black students	–0.45	–0.33
Average *z*-score for Hispanic students	–0.46	–0.32
Average *z*-score for Asian and other students	–0.05	0.15
Percent free lunch eligible	89.70	80.24
Percent Black	62.60	47.63
Percent Hispanic	32.70	30.74
Percent Asian and others	2.49	5.21
Percent LEP	9.45	10.29
Percent recent immigrants	3.17	4.42
Percent special education	5.56	5.46
Percent resource room	6.28	6.72
School characteristics		
Total expenditure per pupil (in dollars)	11,400	11,202
Non-classroom teacher expenditure (in dollars)	6,841	6,637
Pupil-teacher ratio	11.89	12.10
Percent teachers licensed/permanently assigned	73.82	76.19
Percent teachers with master's degree	71.48	71.74
Enrollment	706	615

NOTE: For the subgroup performance variables, N ranges from 24 to 50 in column (1) and 59 to 110 in column (2).
SOURCE: Authors' calculations based upon data provided by the New York City Department of Education.

performance, more disadvantaged student populations, higher spending, and lower teacher quality.

VII. Discussion and Implications for Policymakers

Several factors explain the differences we see in these lists. First, the various lists of best and worst schools were put together in different years (2003 for the NCLB and Chancellor's schools, 1999 and 2002 for the Advocates for Children schools) while the most recent year of data for the efficiency measures is 2000. In addition, each list of best and worst schools is put together for a specific year, while the EPFs require the use of several years of data. This can cause differences, even though averages of variables over those years are not significantly different from their values for each year.

Second, the efficiency measures control explicitly for exogenous factors and resources, which the other methodologies used to identify best and worst schools do not, at least not explicitly. In addition, the other methodologies take different factors into account. Most significantly, the Advocates for Children list takes into account an array of variables other than performance, including school atmosphere (stress, competition, safety); the number, quality and teaching methods of the teachers; the condition of the school building; and ethnic diversity.

Indeed, beyond such technical discrepancies as the year or number of years of data lies a more fundamental source of differences among the lists. All lists, except for the research-based ones, are based on the *performance* of schools, while the research-based ones aim at capturing school *efficiency*—they take into account clientele and resources. It seems clear that these two concepts are distinct, even though the other lists attempt to take into consideration a number of factors that affect school efficiency. Still, comparing subgroups of schools based on these factors is not equivalent to systematically taking into account factors that, as theory dictates, raise or lower the efficiency of a school.

Efficiency in public goods is in the public interest, but it is not necessarily in the interest of each individual or, as

it pertains to education, of each parent. A perfect illustration of this point manifested itself in March in an East Harlem, NY, school, which proved successful enough to prompt the City to suggest that it enroll more students. This suggestion was vehemently opposed by the students' parents, as well as the teachers and the principal, who were satisfied by the children's performance and did not want to jeopardize it by attempting to provide this opportunity to other children (Gootman 2004).[41]

Efficiency is a public concern. Yet because of the way the New York City school system is organized and funded, typical pressure for efficiency from taxpayers and competition between local governments does not really apply, although there are now demands from various levels of government to raise performance. The four lists of best and worst schools this paper discusses before going into the research-based one represent three levels of government—federal (NCLB), state (SURR), and city (Chancellor's)—as well as the not-for-profit sector (Advocates for Children), and while economists and policy planners advocate the importance of efficiency, none of these methods takes it into account. While they may still be used by parents looking to choose schools for their children, it is surprising that no public entity has made an effort to publicize other numbers, such as measures of efficiency. Who then can promote efficiency? Systems are being put in place to identify the best and worst schools and provide support for the improvement of the schools that need it, but there is a need for a mechanism that can assess and promote efficiency in public schools and districts. While efficiency may not yet be well enough defined and assessed to be a solid basis for accountability systems, there may be things that state policymakers can begin to do to reach that stage: data collection, training, research, policy evaluation, assessing funding requirements, etc. (Camphire 2004).

Perhaps the absence of such a mechanism is primarily a concern for large cities. Indeed, in a small school system, pressure from the voters to lower property taxes may act as an incentive for efficiency. In a large school system such as New York City, school funding comes from a large pool of money, and there may be more of a disconnection between the sources of funding, the funding itself, and its uses and users. As such, small school districts may provide a good model for the search for efficiency.

We find that efficiency groupings differ from the best or worst groupings; there is some overlap, but it is not complete. Once a satisfactory way to measure efficiency is found, it would be helpful for policymakers who are deciding whether to punish or assist schools to know if low-performing schools are also inefficient or if high-performing schools are efficient. Low-performing inefficient schools might require reorganization, while low-performing, efficient schools might benefit from increased resources. On the other hand, high-performing schools may be in need of intervention. High-performing *efficient* schools could be left alone, but high-performing inefficient schools could be required to choose between doing more for their students or operating with fewer resources. This is one way policymakers could take advantage of the two criteria—performance and efficiency—at their disposal to evaluate and improve schools.

[41] Gootman, Elissa. (2004, March 3). Many at Successful Middle School Oppose Its Expansion. *The New York Times*.

References

Barbanel, J. (1999a, March 21). Testing, Testing: How the City's Third-Graders Performed in 1998. *The New York Times*, p. 8.

Barbanel, J. (1999b, March 28). Report Cards Point to Struggles for Middle and High Schools. *The New York Times*, p. 10.

Barbanel, J. (1999c, April 11). How the County's Third-Graders Are Reading. *The New York Times*, p. 12.

Barbanel, J. (2000, March 19). The Report Card for Middle and High Schools. *The New York Times*, p. 12.

Bessent, A.M., and Bessent, E.W. (1980). Determining the Comparative Efficiency of Schools Through Data Envelopment Analysis, *Educational Administration Quarterly*, *16*(2): 57–75.

Camphire, G. (2004). Delivering Efficiency, Focus on Bolstering School Finance and Accountability. *Southwest Educational Development Laboratory Letter, 16*(1): 9–11.

Charnes, A., Cooper, W.W., Lewin, A.Y., and Seiford, L.M. (1994). The DEA Process, Usages and Interpretations. In A. Charnes, W.W. Cooper, A.Y. Lewin, and L.M. Seiford, (Eds.), *Data Envelopment Analysis: Theory, Methodology and Application* (pp. 425–435). Boston: Kluwer Academic.

Charnes, A., Cooper, W.W., and Rhodes, E. (1978). Measuring the Efficiency of Decision-Making Units, *European Journal of Operational Research*, *2:* 429–444.

Clotfelter, C.T., and Ladd, H.F. (1996). Recognizing and Rewarding Success in Public Schools. In Ladd, H.F. (Ed.), *Holding Schools Accountable, Performance-Based Reform in Education* (pp. 23–63). Washington, DC: The Brookings Institution.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F.D., and York, R. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

Dill, D.D., and Soo, M. (2003, November). *Is There a Global Definition of Academic Quality?: A Cross-National Analysis of University Ranking Systems.* Paper prepared for the Annual Conference of the Association of Public Policy and Management, Washington, DC.

Downes, T.A., and Pogue, T.F. (1994). Adjusting School Aid Formulas for the Higher Cost of Educating Disadvantaged Students. *The National Tax Journal*, *47*(1): 89–110.

Duncombe, W., and Yinger, J. (1997). Why Is It So Hard to Help Central City Schools? *Journal of Policy Analysis and Management, 16*(1): 85–113.

Duncombe, W., and Yinger, J. (1998). School Finance Reform: Aid Formulas and Equity Objectives. *The National Tax Journal, 51*(2): 239–262.

Duncombe, W., and Yinger, J. (1999). Performance Standards and Educational Cost Indexes: You Can't Have One Without the Other. In H.F. Ladd, R. Chalk, and J.S. Hansen (Eds.), *Equity and Adequacy in Education Finance* (pp. 260–297).Washington, DC: National Academy Press.

Edmonds, R.R. (1979). Effective Schools for the Urban Poor. *Educational Leadership*, *37*(1): 15–24.

Edmonds, R.R. (1982). Programs of School Improvement: An Overview. *Educational Leadership, 40*(3): 4–11.

Färe, R., and Lovell K. (1978). Measuring the Technical Efficiency of Production. *Journal of Economic Theory*, *19*: 150–162.

Farrell, M.J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, *120*(3): 253–281.

Ferguson, R.F., and Ladd, H.F. (1996). How and Why Money Matters: An Analysis of Alabama Schools. In H.F. Ladd (Ed.), *Holding Schools Accountable* (pp. 265–298). Washington, DC: The Brookings Institution.

Fiske, E.B. (2002). *The Fiske Guide to Colleges 2003.* Naperville, IL: Sourcebooks.

Gootman, E. (2004, March 3). Many at Successful Middle School Oppose Its Expansion. *The New York Times,* p. 6.

Gormley, W.T., Jr., and Weimer, D.L. (1999). *Organizational Report Cards*. Cambridge, MA: Harvard University Press.

Hanushek, E.A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature, 24*(3): 1141–1177.

Hanushek, E.A. (1996). *Assessing the Effects of School Resources on Student Performance: An Update* (Working Paper No. 424). Rochester Center for Economic Research.

Hanushek, E.A., and Raymond, M.E. (2002). The Confusing World of Educational Accountability. *National Tax Journal*, *54*(2): 365–384.

Hemphill, C. (1999). *Public Middle Schools: New York City's Best.* New York, NY: Soho Press.

Hemphill, C. (2002). *New York City's Best Public Elementary Schools: A Parents' Guide.* New York, NY: Teachers College Press.

Kadamus, J.A. (2000, January 26). *Status Report on Schools Under Registration Review (SURR).* Memo to Board of Regents. Available: http://www.emsc.nysed.gov/nyc/regrev.html#SURRList/RgntJa00.pdf.

Kane, T.J., and Staiger, D.O. (2002). The Promise and Pitfalls of Using Imprecise School Accountability Measures. *Journal of Economic Perspectives*, *16*(4): 91–114.

Ladd, H.F. (1999). The Dallas School Accountability and Incentive Program: An Evaluation of Its Impacts on Student Outcomes. *Economics of Education Review, 18*(1): 1–16.

Ladd, H.F. (2002). School-Based Educational Accountability Systems: The Promise and the Pitfalls. *National Tax Journal*, *54*(2): 385–400.

Ladd, H.F., and Walsh, R.P. (2002). Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right. *Economics of Education Review, 21*(1): 1–17.

Levin, H.M. (1975). Cost-Effectiveness in Evaluation Research. In M. Guttentag and E.L. Struening (Eds.), *Handbook of Evaluation Research* (vol. 2) (pp. 89–122). Beverly Hills, CA: Sage.

Linn, R.L., and Haug, C. (2002). Stability of School-Building Accountability Scores and Gains. *Educational Evaluation and Policy Analysis*, *24*(1): 29–36.

The New York State Education Department Office of Elementary, Middle, Secondary and Continuing Education. (1998, September). *The Registration Review Process.* Retrieved May 20, 2004, from http://www.emsc.nysed.gov/nyc/regrev.html#SURRList/rrproca.pdf.

Reschovsky, A., and Imazeki, J. (2001). Achieving Educational Adequacy Through School Finance Reform. *The National Tax Journal*, *26*(2): 373–396.

Roderick, M., Jacob, B.A., and Bryk, A.S. (2002). The Impact of High-Stakes Testing in Chicago on Student Achievement in Promotional Gate Grades. *Educational Evaluation and Policy Analysis, 24* (4): 333–357.

Rubenstein, R.H., Stiefel, L., and Schwartz, A.E. (2003). Better Than Raw: A Guide to Measuring Organizational Performance With Adjusted Performance Measures. *Public Administration Review*, *63*(5): 607–615.

Schwartz, A.E., and Zabel, J. (2005). The Good, the Bad, and the Ugly: Measuring School Efficiency Using School Production Functions. In L. Stiefel, A.E. Schwartz, R. Rubenstein, and J. Zabel (Eds.), *Measuring School Performance and Efficiency: Implications for Practice and Research* (pp. 37–66). Larchmont, NY: Eye on Education.

Schwartz, A.E., Stiefel, L., and Bel Hadj Amor, H. (2005). Measuring School Performance Using Cost Functions. In L. Stiefel, A.E. Schwartz, R. Rubenstein, and J. Zabel (Eds.), *Measuring School Performance and Efficiency: Implications for Practice and Research* (pp. 69–72). Larchmont, NY: Eye on Education.

Stiefel, L., Rubenstein, R.H., and Schwartz, A.E. (1999). Measuring School Efficiency Using School-Level Data: Theory and Practice. In A. Odden and M.E. Goertz (Eds.), *School-Based Financing* (pp. 67–87). Thousand Oaks, CA: Corwin Press.

Stiefel, L., Schwartz, A.E., Bel Hadj Amor, H., and Kim, D.Y. (2005). Adjusted Measures of School Performance: A Cross-State Perspective. In L. Stiefel, A.E. Schwartz, R. Rubenstein, and J. Zabel (Eds.), *Measuring School Performance and Efficiency: Implications for Practice and Research* (pp. 17–36). Larchmont, NY: Eye on Education.

Yan, E. (2003, February 14). 208 Schools Make the Grade: Some Educators Grumble List Is Artificial Benchmark. *New York Newsday.* Available: http://www.nynewsday.com/news/education/nyc-skul0215,0,466799.story?coll=nyc%2Deducation%2Dutility.