OECDpublishing

Please cite this paper as:

Hampel, M. (2005), "Measuring Educational Productivity in Standards-Based Accountability Systems", *OECD Education Working Papers*, No. 4, OECD Publishing. http://dx.doi.org/10.1787/224417012465

OECD Education Working Papers
No. 4

Measuring Educational Productivity in Standards-Based Accountability Systems

Martin Hampel

Unclassified **EDU/WKP(2005)1**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development **16-Dec-2005**

English - Or. English

DIRECTORATE FOR EDUCATION

EDU/WKP(2005)1
Unclassified

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

(Education Working Paper No. 4)

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 post-doctoral 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.

JT00196110

English - Or. English

ECD Education Working Papers

This series is designed to make available to a wider readership selected studies on education prepared for use within the OECD. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language English or French with a summary in the other.

Comment on the series is welcome, and should be sent to either edu.contact@oecd.org or the Directorate for Education, 2, rue André Pascal, 75775 PARIS CEDEX 16, France.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its member countries.

---

www.oecd.org/edu/workingpapers

INTRODUCTORY NOTE

1. 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 kindergarten through 12th grade 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 debate in the education community, parents, educators, administrators, legislators and stakeholders require an objective way of ascertaining the progress of public schools throughout the United States.

2. The research presented in this paper relates data on school performance and educational spending in ways that allow for an assessment of the productivity of school systems. While the data being utilised relate to the United States only, the methodology being proposed also would lend itself to comparisons of aspects of the productivity of education systems internationally, utilising data from the OECD Programme for International Student Assessment (PISA).

3. The figures discussed in this paper are found in Annex 1 and Annex 2.

LITERATURE AND METHODOLOGY

4. 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 impossible to provide an exhaustive review of the literature, and any overview could not be comprehensive. However, Armor 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 2003). Armor worked as a graduate student on J. Coleman's classic study *Equality of Educational Opportunity* (Coleman *et. al.* 1966), which pioneered the identification of the relationship between socioeconomic background and student performance. Coleman's 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 *et al.* (2001), while Schweke (2004) also provides an additional overview.

5. 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 also focuses on identifying the position of individual entities, such as school districts, relative to these environmental variables, which is important from a methodology perspective.[3]

6. Under the No Child Left Behind Act of 2001 (NCLB), states and school districts in the United States now have more flexibility in how they use federal education funds. Accordingly, Standard & Poor's School Evaluation Services (SES) has introduced 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 Standard & Poor's 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 enrolment of economically disadvantaged students.

7. While the NCLB establishes the goal of educational proficiency in reading, math, and science, such proficiency is usually measured by cut-off 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.

8. 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.

9. The RoSI approach introduced in this methodology paper can be extended beyond the framework of analyzing United States district level data into an international context, and could be used as a model for undertaking similar comparisons, using national education systems as the entities being compared. While a straightforward extension of the analysis to the international level seems directly feasible, introducing modified indicator definitions might add useful information. However, data for these modified indicators are typically not currently available at the United States district level.

10. For the performance variable, utilizing test scores that are converted to average grade-year equivalents might be an alternative view of relative performance, allowing for the expression of the difference between actual educational attainment and expected performance. Since most U.S. standardized test results are reported as proficiency rates rather than as scaled scores with information on how to link them to grade level equivalents, this approach does not seem feasible at the U.S. district level. Using OECD PISA data might be a way around this limitation at the international level, but this approach would likely prevent the analysis of U.S. states.

Getting more out of test data

11. To assist states and districts that do not currently have value-added assessment systems but that 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.

12. While looking at performance over poverty is a worthwhile approach in its own right and provides interesting and actionable insights, the input variable – poverty – cannot be controlled directly by education decision makers. 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 terms of educational performance does a certain level of spending achieve?

Defining a Return on Spending Index

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

1. Choosing an appropriate performance indicator;

2. Selecting the appropriate corresponding spending variable;

3. Defining a RoSI. This indicator will provide a general productivity measure as a proxy for average educational return, given a certain spending level; and

4. Performing a 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]

CHOOSING AN APPROPRIATE PERFORMANCE INDICATOR

14. 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.

15. Figure 1 provides an example of a scatter plot showing the New York grade 8 Mathematics Test Proficiency Rate versus the enrolment levels 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.

16. 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 three years (2001 to 2003), plus a corresponding graduation rate and retention rate indicator.[6] The corresponding Error Band plot is shown in Figure 2.

17. 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 which partially have a higher average, results in an increase of the average MPM Index value compared to the grade 8 mathematics test results.

18. 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 one year are used.[7] For this report, data come from the 2001-2002 fiscal year.

SELECTING THE APPROPRIATE CORRESPONDING SPENDING VARIABLE

19. 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.

20. 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 RoSI Error Band analysis introduced below will explicitly take the proportional enrolment of economically disadvantaged students into account.

21. 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 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]

22. 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]

23. Similarly to the potential modification of the performance variable definition in the international context, a different spending measure might be valuable as well, such as a longer term average spending indicator or the cumulative spending amount over the typical educational life of the student.[10]

DEFINING A RETURN ON SPENDING INDEX

24. 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 indicator, yielding the average amount of money spent per unit of achievement measured.[11]

25. The structure of such a measure with respect to the enrolment of economically disadvantaged students is usually dominated by the performance variable, whose relationship to the enrolment of economically disadvantaged students is typically much stronger than that of the spending distribution. Therefore the PCI could be inverted to create a RoSI, which is defined as a performance indicator divided by a spending variable and which can be interpreted as a productivity indicator.[12]

26. The additional benefit of the RoSI methodology lies in its more intuitive meaning as a measure of productivity. Larger values are often viewed more favourably 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.[13]

PERFORMING A COMPARATIVE RETURN ANALYSIS

27. 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 behaviour 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, when combined with additional criteria such as a minimum performance level, signifies whether an entity is using resources efficiently.[14]

28. Figure 3 shows the overlay of the performance Error Band of Figure 2 with a RoSI Error Band in relationship to the enrolment 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.[15] 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.

29. This figure shows clearly that 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 relative width of the RoSI Error Band is greater than the relative width of the performance measure Error Band.[16]

30. One possible follow-up analysis would be to look at the Proficiency Rate value and the RoSI value for each district separately.

31. 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.[17] 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 enrolment 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 behaviour, which consists of proficiency above the statistical 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).

32. 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.

33. 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 Index 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 enrolment of economically disadvantaged students.

34. Since the MPM Index is defined as a three-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 development of the productivity of each district over time similar to a multiyear analysis of the performance Error Band.

35. The RoSI approach presented in this paper expands the Error Band analysis of a performance measure in relationship to the enrolment 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 decision makers.[18]

36. 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.

37. Another important extension of the RoSI methodology would be the inclusion of an indicator that measures the inequality of performance and productivity across the demographic spectrum, *i.e.* a Slope Score representing a measure of the steepness of the regression line and thereby providing an indicating of the underlying inequality. Within the United States, different state tests using distinct designs can usually not be directly compared, and therefore slope scores may be less directly meaningful without adjustments.

38. 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.

REFERENCES

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

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

Gazzerro, P. and M. Hampel (2004), "Identifying Outperforming and Underperforming Schools: Introducing the SES 'Error Band' Method", SES Research Methods Paper No. 1, Standard & Poor's, New York, N.Y.

Hampel, M. (2005), "Introducing the SES Risk-Adjusted Performance Index", SES Research Methods Paper No. 2, Standard & Poor's, New York, N.Y.

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

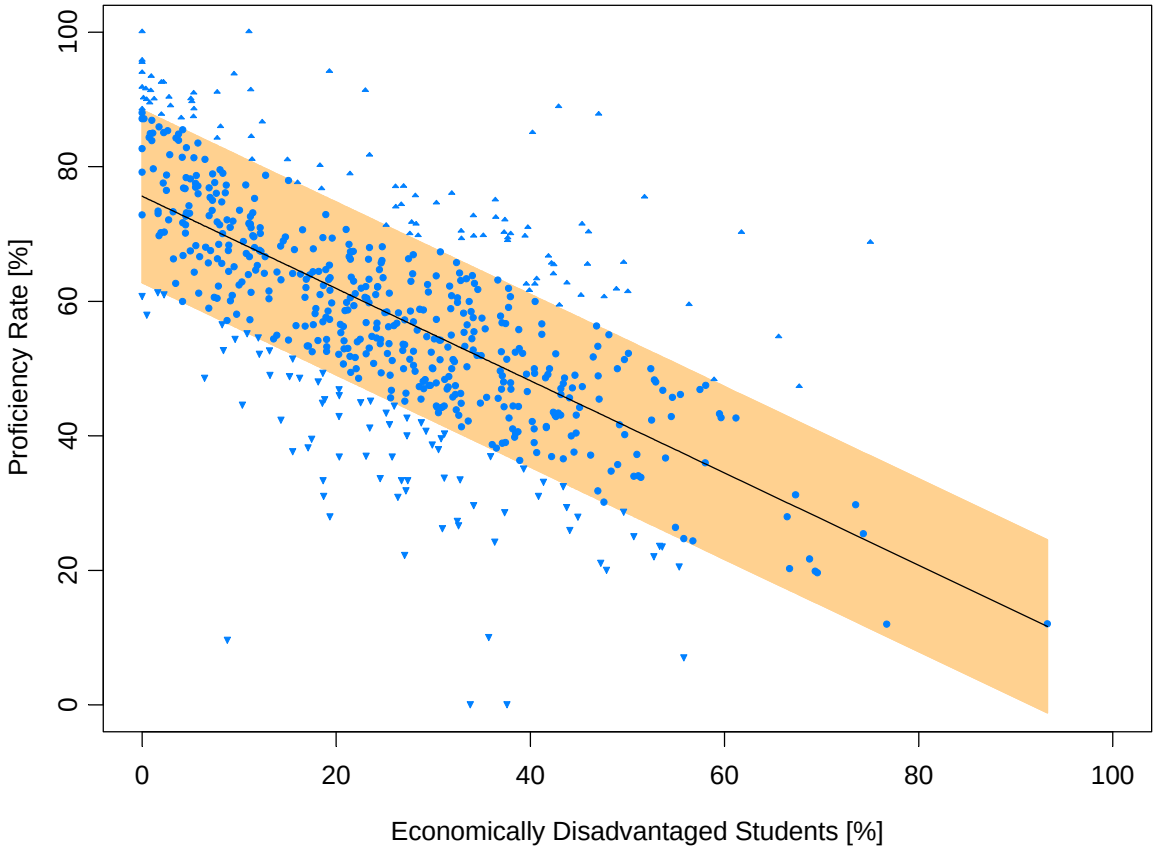
Monk, D.H., H.J. Walberg and M.C. Wang (2001), *Improving Educational Productivity. Research in Educational Productivity,* Information Age, Greenwich, C.T.

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

Standard & Poor's (2004), *Resource Adequacy Study for the New York State Commission on Education Reform* [internal publication], Standard & Poor's, New York, N.Y.

ANNEX 1: FIGURES

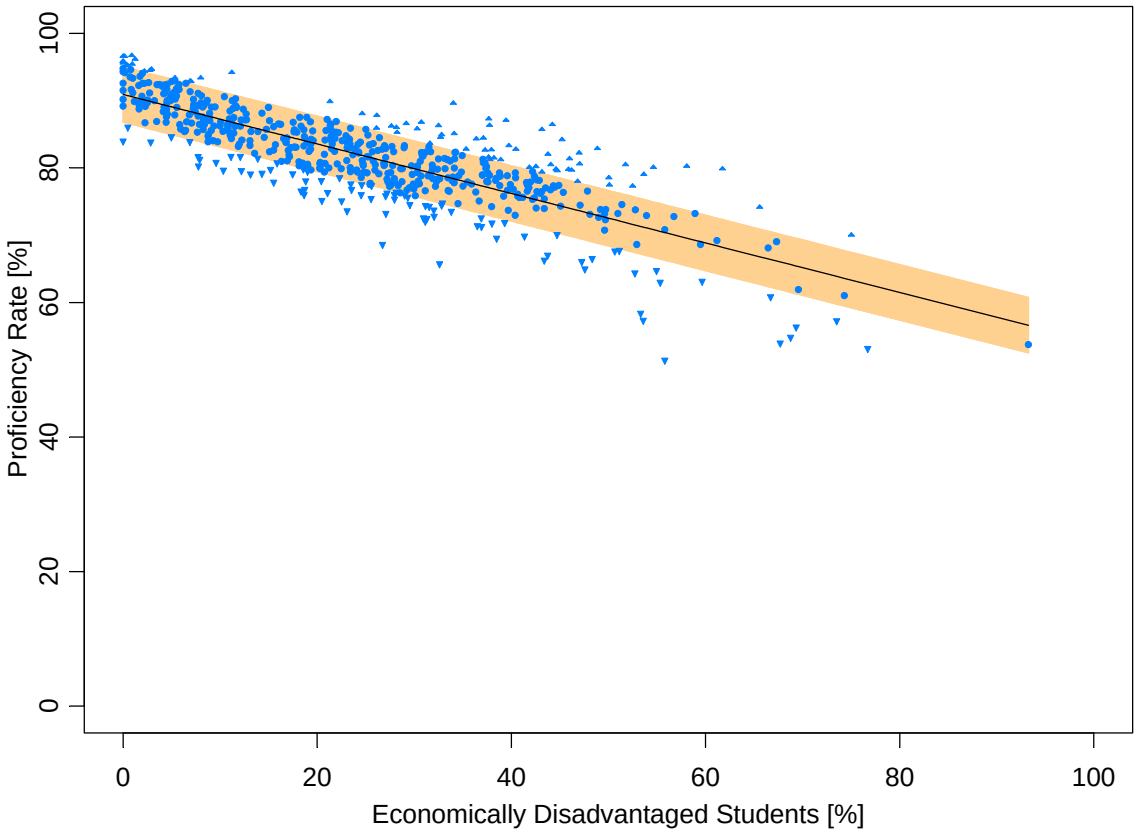
Figure 1. Scatter plot of New York State grade 8 mathematics test Proficiency Rate versus enrolment of economically disadvantaged students for 2002 (using available data for 635 school districts)



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.

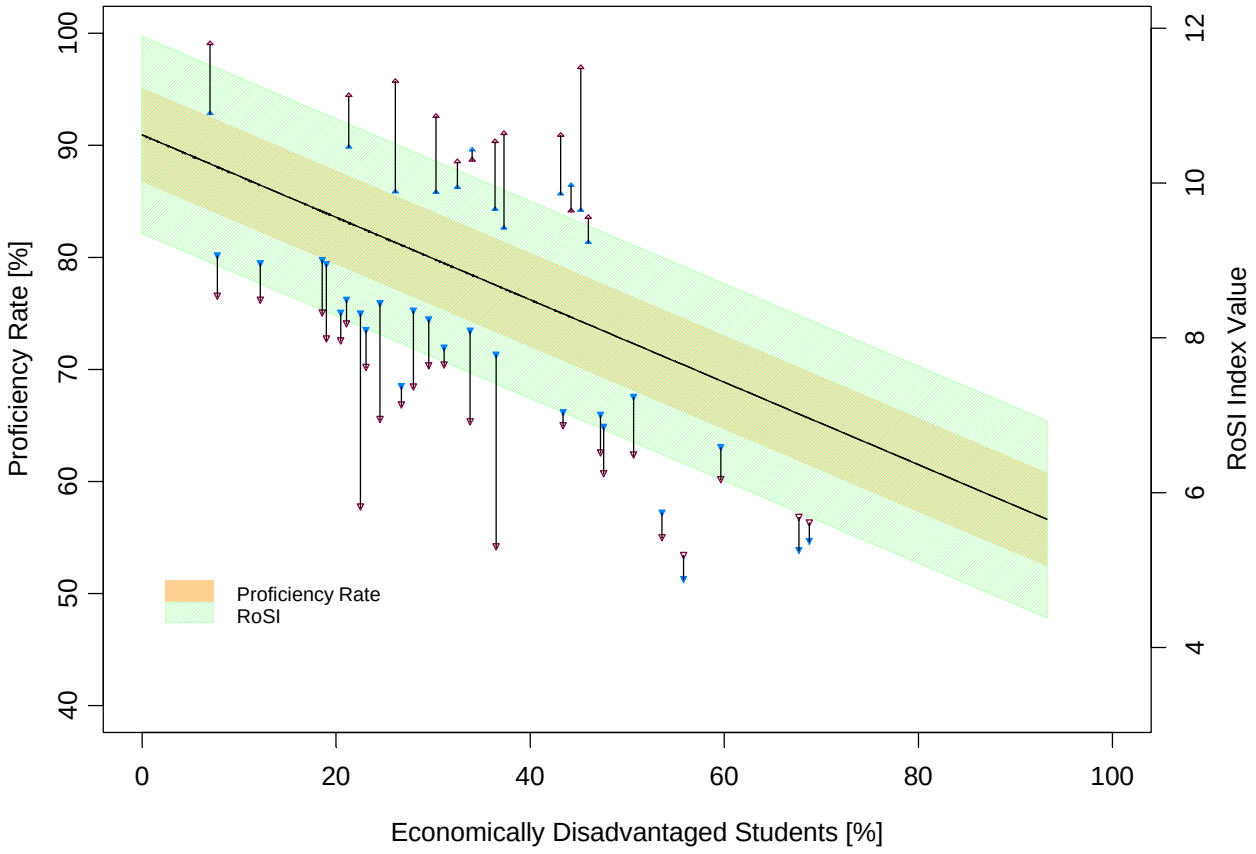
Figure 2. Scatter plot of the Multiple Performance Measures Index (MPM) for New York State (using available data for 581 school districts), expressed as a Proficiency Rate



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.

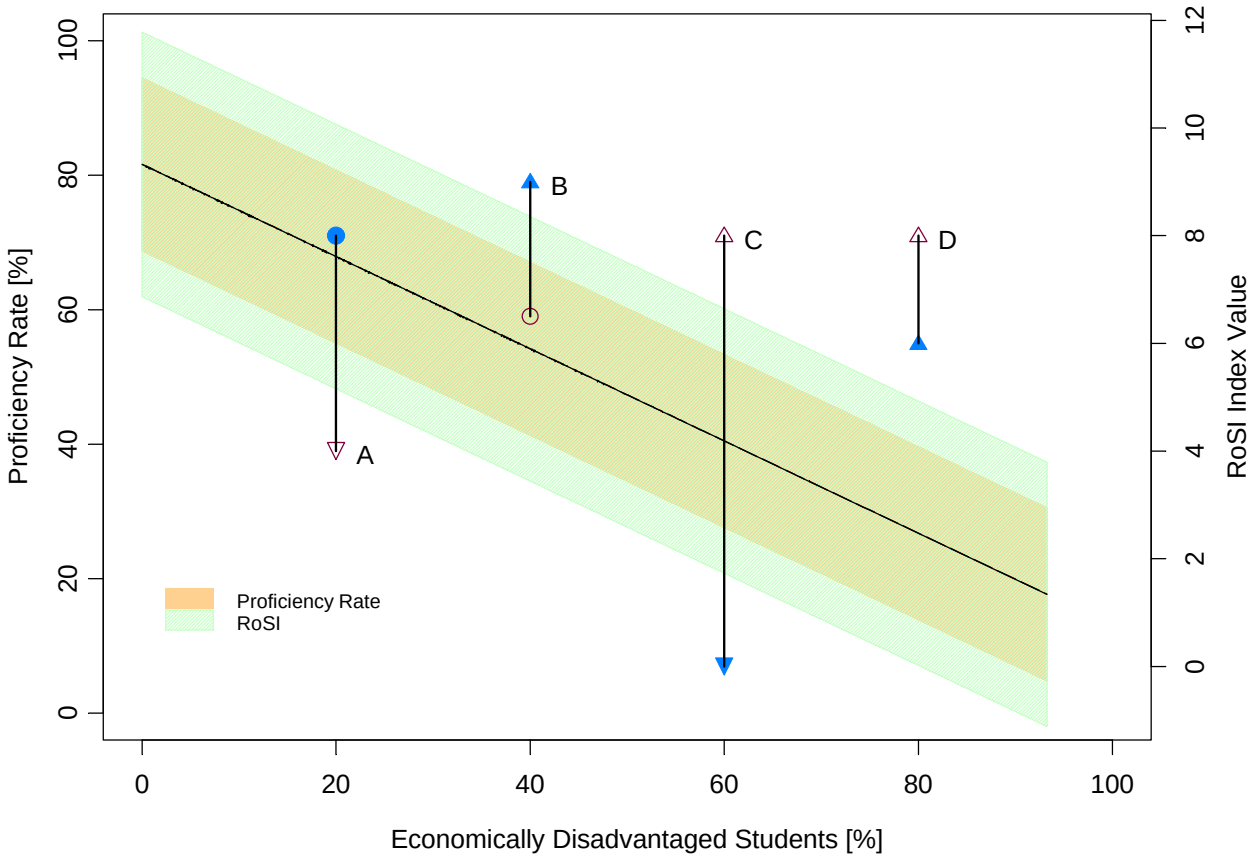
Figure 3. Extended plot of scatter plot in Figure 2, with the addition of a RoSI Error Band using the scale at the right-hand side of the plot, rescaled such that the linear regression lines overlap



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.

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



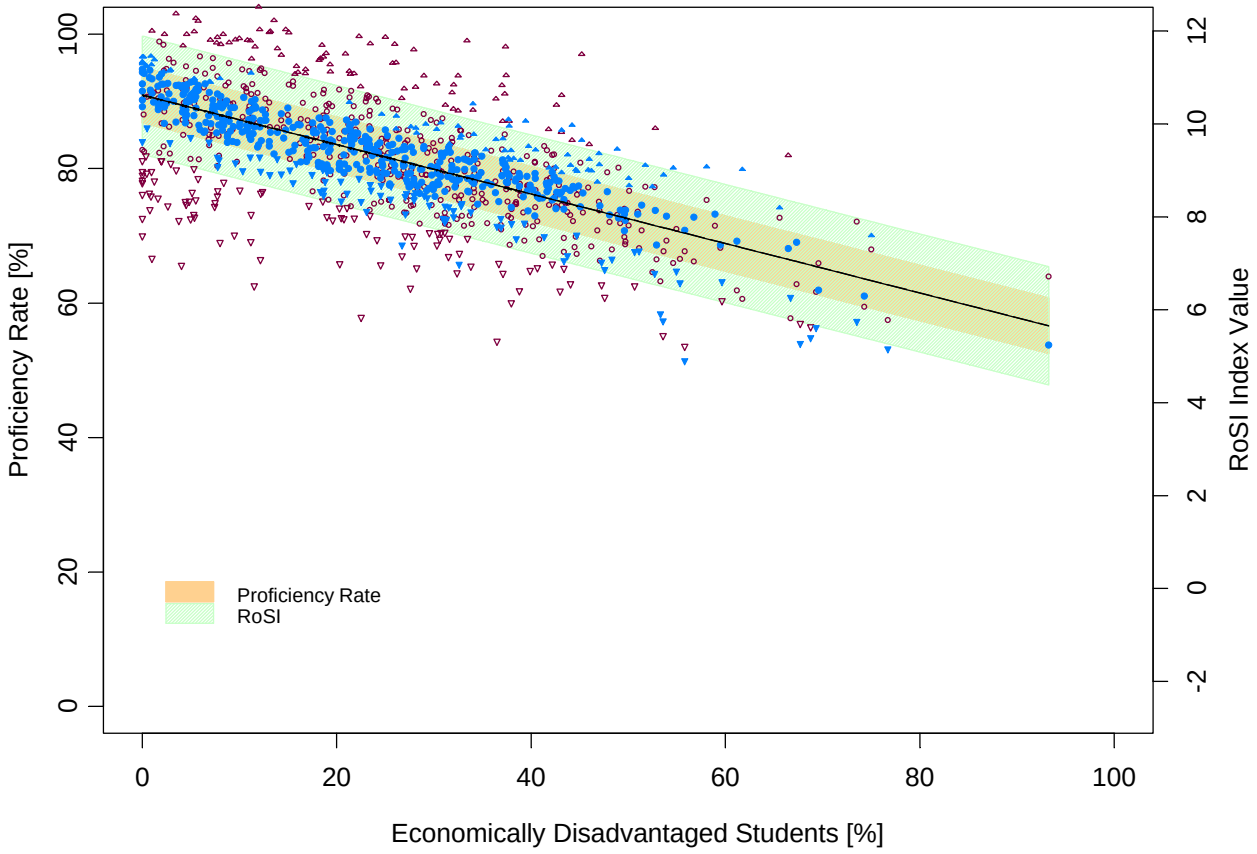
Note: Proficiency Rates above, within, and below the Proficiency Rate band are indicated by upward facing triangles (▲), circles (●), and downward facing triangles (▼), respectively. RoSI values of districts that lie above, within, or below the RoSI band are indicated by open upward facing triangles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

Source: Author.

ANNEX 2: DATA POINTS UTILIZED FOR PROFICIENCY RATE AND ROSI ERROR BAND ANALYSIS

Figure 5 contains essentially the same two Error Bands as Figure 3, but 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 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

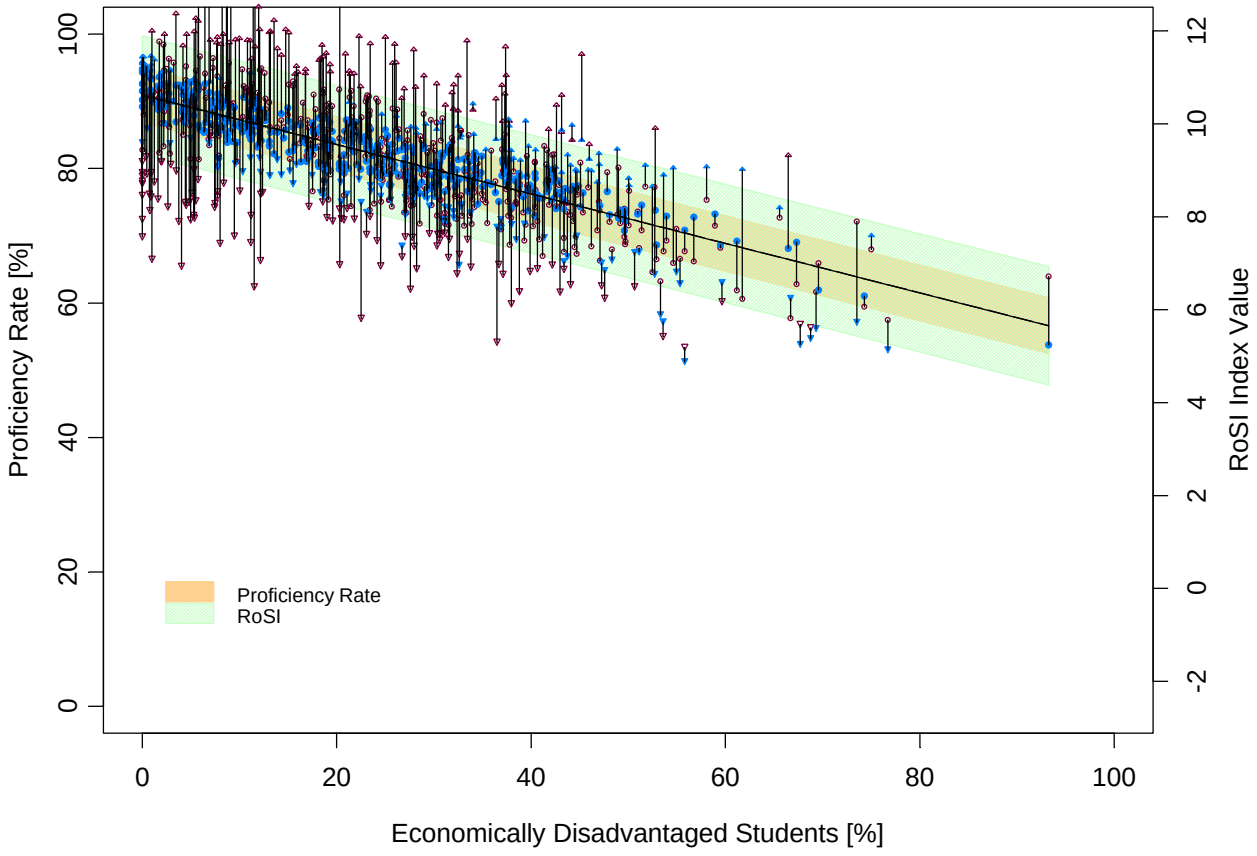


Note: Proficiency Rates above, within, and below the Proficiency Rate band are indicated by upward facing triangles (▲), circles (●), and downward facing triangles (▼), respectively. RoSI values of districts that lie above, within, or below the band are indicated by open upward facing triangles (△), 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 shows the same information as Figure 5, but 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. Thus 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.

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.

OECD EDUCATION WORKING PAPERS

No. 1 TEACHER DEMAND AND SUPPLY: IMPROVING TEACHING QUALITY AND ADDRESSING TEACHER SHORTAGES (2002), Paulo Santiago

No. 2 TEACHER EDUCATION AND THE TEACHING CAREER IN AN ERA OF LIFELONG LEARNING (2002), John Coolahan

No. 3 TOWARDS AN UNDERSTANDING OF THE MECHANISMS THAT LINK QUALIFICATIONS AND LIFELONG LEARNING, Friederike Behringer and Mike Coles

OECD Education Working Papers can be found on the OECD Education Web site at: www.oecd.org/edu/workingpapers. If you wish to be informed about the release of new OECD Education working papers, please: (i) Go to the OECD Web site: www.oecd.org; (ii) Click on "My OECD"; (iii) Sign up and create an account with "My OECD"; (iv) Select "Education" as one of your favourite themes; and (v) Pick "OECD Education Working Papers" as one of the newsletters you would like to receive. For further information please write to: edu.contact@oecd.org.

NOTES

[1] U.S. Department of Education (no date), "How No Child Left Behind Ensures Schools Get Results", see http://www.ed.gov/nclb/accountability/results/getting_results.pdf.

[2] An example of the discussion over the impact of educational spending is the exchange between Hanushek and Hedges, references to which can be found in Armor (2003).

[3] Klitgaard and Hall (1973) take a somewhat similar approach, attempting to identify effective schools based on the analysis of residuals.

[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 a performance zone in documents addressed to the general public.

[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 note 12.

[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 2001-2002 fiscal year; the same data are used in this paper.

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

[8] Further details about the aspects mentioned in this section can be found in the report *Resource Adequacy Study for the New York State Commission on Education Reform* (2004).

[9] The 2002 financial data were the latest data publicly available at the time of the publication of the report *Resource Adequacy Study for the New York State Commission on Education Reform*. Since the year 2002 falls in the middle of the three-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.

[10] As in the case of the performance variable, there are currently not sufficient easily publicly accessible data to produce such a long ranging aggregation for states within the United States.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] In this context, relative width refers to the width of the band expressed relative to the typical scale of the variable. An appropriate proxy for the scale is the sample mean of the data, and therefore the relative width can be expressed as the width of the band divided by the sample mean. Due to the different scales, the graphical representation can display either band as broader, depending on the data set under consideration.

[17] 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.

[18] 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.