

Education Working Paper Archive

**What Do Cost Functions Tell Us About the Cost of an Adequate
Education?**

January 3, 2008

Robert M. Costrell

*Department of Education Reform
University of Arkansas at Fayetteville*

Eric Hanushek

Stanford University

Susanna Loeb

Stanford University

Abstract:

Econometric cost functions have begun to appear in education adequacy cases with greater frequency. Cost functions are superficially attractive because they give the impression of objectivity, holding out the promise of scientifically estimating the cost of achieving specified levels of performance from actual data on spending. By contrast, the opinions of education stakeholders form the basis of the most common approach to estimating the cost of adequacy, the professional-judgment method. The problem is that education cost functions do not in fact tell us the cost of achieving any specified level of performance. Instead, they provide estimates of average spending for districts of given characteristics and current performance. It is a huge and unwarranted stretch to go from this interpretation of regression results to the claim that they provide estimates of the minimum cost of achieving current performance levels, and it is even more problematic to extrapolate the cost of achieving at higher levels. In this paper we review the cost function technique and provide evidence that draws into question the usefulness of the cost function approach for estimating the cost of an adequate education.

The authors wish to acknowledge the support of the Missouri Show-Me Institute. The usual disclaimers apply.

Introduction

Econometric cost functions have begun to appear in education adequacy cases with greater frequency. While previously considered too technical for courts to understand, recent litigation in Missouri featured separate cost function estimates commissioned by each of two plaintiffs. A prior Texas court case presented results from two dueling cost studies commissioned by the opposing sides.¹ This increased use of the cost-function methodology likely reflects growing skepticism about other methods typically used to estimate the cost of providing an adequate education. In particular, the “professional judgment” (PJ) method has begun to lose favor.² In this approach, panels of educators design prototype schools that they believe will provide adequate educational opportunities and then the consultants hired to conduct the study attach costs to these prototypes. Even a sympathetic trial judge in Massachusetts concluded that the PJ study submitted there was “something of a wish list.”³ Hence, although PJ studies are invariably included, recent finance cases have attempted to bolster these with econometric cost functions.

Cost functions are superficially attractive because they appear objective, holding out the promise of scientifically estimating the cost of achieving specified levels of performance from actual data on spending instead of relying on opinions, as do professional-judgment estimates. In keeping with this perception, a group of education finance specialists began arguing that econometric cost functions are the most scientifically valid method to determine the cost of adequacy. To make this argument, they asserted that the methods for estimating cost functions in the private sector – where competition tends to drive out inefficient producers – could be readily adapted to public education. They prepared estimates for legislative committees and courts in states such as New York, Texas, Kansas, and Missouri, and published their work in academic journals. The problem, we shall argue, is that education cost functions do not in fact tell us the cost of achieving any specified level of performance, as claimed.

This is not to say that cost functions tell us nothing. They do provide estimates of average spending for districts of given characteristics and indicate how spending varies by these characteristics in the specific state. For example, they may tell us that in state X, per pupil spending averages Y thousand dollars for districts with a certain percent of free-lunch or reduced-price lunch eligible (FRL) students or of black students and that the average rises or falls by Z dollars as these percentages change. Regression equations provide a useful summary of such patterns. By extension, including measures of performance (e.g., test scores) as a variable permits summarizing what the average spending is for districts with given demographics and performance levels.

However, it is a huge and unwarranted stretch to go from this modest interpretation of regression results to the far more extravagant claim that these provide estimates of the cost of achieving any given performance level for districts of given demographics. There are two key heroic assumptions that are required: (1) the estimates of *average spending* among comparable districts

¹ Imazeki and Reschovsky (2004b), Gronberg, Jansen, Taylor, and Booker (2004).

² The alternative methods are discussed in Ladd, Chalk, and Hansen (1999). Particular attention to the use of cost functions can be found in Gronberg, Jansen, Taylor, and Booker (2004), Duncombe (2006), and Baker (2006a). Critiques can be found in Hanushek (2006, 2007).

³ Costrell (2007), p. 291.

can be adjusted so that they reflect the *minimum efficient cost*⁴ to generate current performance levels, and; (2) the estimated variation in average spending across districts with different performance levels can be used to extrapolate the costs of raising performance to levels not currently observed by comparable districts.

As we will show in this paper, the method typically used to convert average spending figures into estimates of efficient cost accomplishes nothing of the sort. For that reason, there is no foundation for interpreting spending variations across districts with different demographics as the required spending premiums for demographic groups. Finally, the estimated relationship between “cost” and performance is highly unreliable – it is typically estimated with huge imprecision, wide sensitivity to model specification, and by methods that often fail to eliminate statistical bias. As a result, the cost estimates for raising performance to target levels have no scientific basis.

None of this should be surprising. The recent push for experiments in education research is just one of many indications of the difficulty of estimating the effects of resources on student learning. Why would we need experiments if we could just use average district spending and average student test scores, as do cost functions, to estimate the effect of resources on achievement? Decades of research have repeatedly failed to find a systematic empirical relationship between average spending and performance. It would be quite noteworthy if a handful of recent spending equations were to suddenly have found a relationship that had eluded decades of previous investigation. This simply is not the case. The deeper reasons for this and the consequences thereof are the subject of this paper.

The Basic Problem: The Cloud

The logic behind regression-based estimates of the cost of adequacy is seemingly compelling. Why shouldn't we be able to use data on district spending and student test-performance to estimate the costs of achieving a given outcome goal?

The dimensions of the difficulty with this are easiest to see by looking at the simple relationship between spending and performance. Figure 1 shows a plot of spending and performance in 2006 for the 522 districts of Missouri. The vast majority of districts lie in a solid cloud of spending between \$5,000 and \$8,000 per student and getting average achievement on the Missouri Assessment Program (MAP) tests between roughly 700 and 800. At virtually any spending level in the range of \$6,000-8,000 there are some districts below 700 points and some over 800. This blob of data illustrates the two dimensions of the difficulty referred to above: (i) average spending differs greatly from minimum spending at any given performance level; and (ii) there is no apparent association between average performance and average spending in this group.

There is a smaller number of districts spending over \$9,000 but still no obvious pattern of being high or low on the math tests. Additionally, the size of the circles indicates the student

⁴ To an economist, this is a doubly redundant phrase, since “cost” implies efficiency, which in turn implies minimum spending necessary to achieve a given outcome. Since this usage may not be universal, we use this phrase for clarity and emphasis.

populations. Some large districts are above average in performance, while others are below average. The two large and high spending districts that stand out are Kansas City and St. Louis. Both are noticeably below average in student performance.

Taking all the districts together, the line in the picture shows that the simple relationship between spending and achievement is essentially flat. Even if, on average, there is a small relationship between average spending and average achievement, either positive or negative, the relationship is very weak. That is the fundamental challenge. How can one project the spending necessary to improve student performance to any level when the available data show little tendency toward higher achievement when given extra funds?

Districts of course differ in a variety of dimensions other than spending, leading to a considerable amount of analytical effort to control for other factors in order to uncover any systematic influence of spending. The basic question is whether other factors that might affect performance, such as poverty levels, can be used to sort districts out of the cloud of Figure 1 such that a pattern with spending emerges. Extensive efforts to do this, beginning with the “Coleman Report” (Coleman et al. (1966)), have been quite unsuccessful. These efforts, generally labeled estimation of production functions, have concentrated specifically on different backgrounds of students and have attempted to standardize for family inputs that are outside the control of schools.⁵

The Cost Function Approach

The estimation of cost functions approaches this problem in a slightly different manner than most research exploring the relationship between spending and achievement. It focuses on how achievement levels determine spending, as opposed to how spending determines achievement. When put in terms of the determinants of spending, other things logically enter the analysis. First, districts might differ meaningfully in the prices that they face for inputs, particularly teachers. The price for teachers and other college graduates can be quite different for one district than for another because of the labor markets in which they compete. If districts must pay higher prices to obtain the same quality of resources, then omitting price differences could bias the estimated relationship between achievement and spending. Second, cost functions, similar to production functions, must account for possible variation in resource needs arising from students who have fewer resources at home and thus may require more resources at school, on average, to achieve the same level of performance. Again, if need differences are omitted from cost functions, the estimated relationship between achievement and spending may be biased. Third, districts may differ in the efficiency with which they use their funds. Two districts with similar spending, similar prices and similar needs might achieve quite different outcomes, based on the efficiency with which they use their dollars to produce the outcome in question. To isolate cost, these estimates must address differences in efficiency.

The underlying premise of the cost function estimation is that correcting for price differences, the demands of different student bodies, and the efficiency of district spending will yield a clear relationship between achievement and the spending that is required to achieve each level of

⁵ Hanushek (2003)

performance. This relationship then permits identifying the spending required to achieve any chosen level of student achievement.

Do these corrections work?

To answer this question, we trace through some specific cost function analyses. We focus on those submitted in the Missouri court case because the data were readily available for purposes of replication and analysis.⁶ However, the issues identified here apply to the entire genre of cost functions based on the “efficiency control” approach.⁷

Figure 2 provides a similar picture to that previously presented.⁸ The performance measure, for 2005, is a composite of each district’s performance on the state assessments -- specifically the percent of students in the top two categories (out of five) on the math and communications arts exams across three grades. Unlike Figure 1, this figure places spending, to be determined by achievement and school factors, on the vertical axis. Figure 2 again shows there is a wide range of spending observed at any given level of performance.⁹ As a result, the line fitted through these data exhibits a very weak relationship.¹⁰

What a cost function tries to do is to go beyond this simple (weak) relationship to estimate for a district of given characteristics the minimum expenditure required to meet some target performance level. This can be logically broken down into three steps in constructing the cost estimates:

- 1) Control for district characteristics, so that “likes” can be compared with “likes.” As mentioned, one reason for the wide range of spending is that districts differ in characteristics, such as demography, school size, input prices, and variables thought to affect efficiency. The variation in spending among districts with comparable scores is partially related to these differences. Cost functions statistically control for demographics and other district characteristics with the conventional technique of multiple regression, discussed in the next section.

⁶ Baker (2006b), Duncombe (2007). Baker was retained by the main group of plaintiff districts, the Committee for Educational Equality, and Duncombe was retained separately by the City of St. Louis. For the defense, Costrell was retained by the Attorney General of Missouri and Hanushek by the Defendant Intervenors (Shock, Sinquefeld & Smith).

⁷ In addition to some of the studies cited above, a partial list would also include, Duncombe and Yinger (1997, 2000, 2005, 2007), Imazeki (2006), Imazeki and Reschovsky (2004a, 2004b), and Reschovsky and Imazeki (2003). Imazeki and Reschovsky, in their various publications about costs in Texas alternately use an efficiency index derived from a data envelope analysis (DEW), including a Hefindahl index, or ignore the issue.

⁸ Figures 2-5 are based on the Duncombe data and analyses. The Baker data and analyses are very similar, and the corresponding figures are available upon request. Both studies pool data across several years, although these diagrams only depict one year.

⁹ Similarly, there is a very wide horizontal range: at any given spending level, performance varies widely.

¹⁰ In fact, if these data suggest any relationship at all, it is U-shaped, rather than linear, which means a negative relationship between performance and spending over the lower ranges of performance and a positive relationship over the higher ranges. The linear relationship has an R^2 of only 4%, the portion of the variation in spending accounted for by variations in performance. A quadratic relationship, depicting the U-shaped curve, provides a much better fit, with an R^2 of 30%.

- 2) Purge inefficiency from the estimates of spending. This is the key step in converting a spending function to a “cost” function. It does so by standardizing the values of the “efficiency controls” used in step (1). If successful, this procedure would identify the minimum expenditure required to perform at the current level.
- 3) Estimate the cost of raising performance to the target level. This involves using the estimated relationship between cost and performance to predict the cost associated with increasing performance to a set goal. It requires a reliable estimate of the relationship between cost and performance from step (1).

As we shall see below, the cost function methodology does not succeed in this agenda. In order to understand the issues more fully, we provide a detailed discussion of these steps: (1) controlling for district characteristics, (2) purging inefficiency from average spending, and (3) estimating the additional cost for additional performance.

The Econometrics of Spending Equations: Controlling for District Characteristics

The basic technique of linear regression is illustrated by the line through the data in Figure 2. Each point on the line represents the best estimate for *average* spending among districts with any given test score. Dots above the line (blue) denote districts spending above the estimated average and dots below the line (red) denote districts spending below the estimated average, for any given test score.

The technique of *multiple* linear regression is conceptually identical, except that it adds more variables with which to predict spending. The additional variables cannot be depicted graphically in two dimensions, but the idea of adding variables to an equation is straightforward:

$$(1) \quad \text{spending}_{it} = \beta_0 + \beta_1 \cdot \text{performance}_{it} \\ + \beta_2 \cdot (\text{teacher_salaries})_{it} + \beta_3 \cdot (\%FRL)_{it} + \dots + \beta_n \cdot (\text{prop_val})_{it} + \dots + u_{it}$$

Spending in district i and year t is specified to depend on student performance, teacher salaries (as the key input price), percent FRL, other demographic and school variables (such as school size), and a set of “efficiency controls” such as property values (discussed in the next section). The unexplained component, u_{it} , is the error term representing factors not captured by the measured attributes. It can be positive or negative, but has an average value of zero. The regression estimates the coefficients $\beta_0, \beta_1, \beta_2$, etc., to provide the best fit to the data, minimizing the variation in the estimated error term.¹¹

The key point here is that the resulting equation is a spending equation which gives an estimate of average spending for a district of given performance and other characteristics. There is nothing controversial about this statement – the cost function practitioners would agree, since

¹¹ In the interest of simplicity, the text omits a number of technical details. For example, these equations are often estimated in logarithmic form for the dependent variable and some independent variables. Also, typically the estimation uses instrumental variables for the performance variable (and perhaps others, such as teacher salaries), as will be discussed in a later section.

this is only the first step in their estimation of the cost, or *minimum* spending necessary to produce a given level of performance.

We will defer discussion of the key coefficient on performance, β_1 , to a later section, but some of the other coefficients are readily interpreted. The estimated coefficient β_3 represents the additional spending, on average, among districts with higher percentages of FRL students, holding other variables constant. In essence it indicates what districts with different levels of poverty are spending. It does *not* represent the extra *cost* required to achieve any given performance level for FRL students. All a positive β_3 coefficient in equation (1) would reflect is a tendency of either the state or the district to spend more heavily when there is a greater proportion of students in poverty, while any similar tendency to spend less on poor students would yield a negative coefficient. This interpretation of β_3 holds regardless of whether extra spending is required to increase performance or is effective at doing so.¹²

The distinction is quite important, because coefficients estimated from such equations are regularly adduced to specify cost premiums (or student weights) in school funding formulas.¹³ For example, in Missouri, the estimate of β_3 was taken to mean that a student receiving a subsidized lunch in an average district is over 50% “more expensive than a student not receiving a subsidized lunch to bring up to the same performance level,” an interpretation that goes beyond what is warranted from a spending equation.¹⁴

The interpretation of demographic coefficients is further illustrated by variables for race. As an example, the estimate by Baker (2006b) of the extra spending for black students in Missouri was 70 percent.¹⁵ The direct interpretation of this equation is that Missouri spends more on average in districts with higher concentrations of black students (controlling for FRL, etc.). This is consistent with Missouri’s history of mandated remedies in prior desegregation cases. But, since these estimates are drawn from *spending* equations (not *cost* equations), it is an over-interpretation to conclude that these coefficients represent the required extra *cost* for black students to achieve any given level of performance.

Consider next the control for teacher salaries. The idea here, drawn from the theory of competitive markets, is that if important input prices are beyond the producer’s control, they are an independent determinant of cost. For such input markets, the producer is said to be a “price-taker.” However, it is highly questionable whether such conditions are reasonably satisfied by teacher markets. While much of the variation in teacher salaries across districts is correlated with the wages of non-teaching college graduates in the region (labor market), within regions districts vary meaningfully in salaries they pay to teachers. This within region variation draws

¹² To be sure, it is uncontroversial that higher FRL is associated with lower district performance; but the statistical evidence that extra spending systematically raises performance over the observed range is highly controversial. Student-level data from Missouri indicates no relationship between spending and performance of African-American FRL students (Podgursky, Smith, and Springer (2007), Figure 10).

¹³ Duncombe and Yinger (2005).

¹⁴ Duncombe (2007), p. 24.

¹⁵ Because of the specific functional form, the estimate varies modestly depending on the percent of students that are black. The estimate given above is for the average district in the state, while for St. Louis, the figure is 85% (Baker (2006b)). The estimate in Duncombe (2007) also implies a very substantial premium, but because of the way that race entered the equation (interacted with FRL) the interpretation is less straight-forward.

into question the “price-taking” assumption.¹⁶ Consequently, district variation in teacher salaries likely includes discretionary variation, not simply variation in cost. This problem is recognized by some of the cost function practitioners, and their attempted solution is discussed in a later section (on instrumental variables).¹⁷ The point here is that input pricing illustrates the difficulty in adapting cost function estimation from competitive markets to the very different environment of public education.

To see the effect of all the controls in (1) taken together, consider each district’s *fitted* value for spending. This is the value for each district using the estimated β ’s in (1) and setting the error term to zero. It represents the estimated spending for the *average* district of that *specific* district’s characteristics. In the simple case of Figure 2, where performance was the only right-hand-side variable, the *fitted* value is represented by the line and the *actual* value is represented by the dots. The difference between *actual* spending and *fitted* spending is the distance from the dots to the line (also known as the *residual*).

How is this affected by the addition of all the explanatory variables in (1)?¹⁸ The answer is seen in Figure 3. The deviation of *actual* spending from *fitted* spending – the amount that each district differs from the regression line – is depicted on the vertical axis, plotted against the district’s performance. In effect, Figure 3 replicates Figure 2 except that instead of plotting actual spending on the vertical axis, it plots spending adjusted for performance and other district characteristics including student poverty, race, teacher salaries, etc.

For St. Louis, the effect of these controls is striking. In Figure 2, St. Louis was the highest spending among districts with comparable test scores. In Figure 3, St. Louis is among the lowest-spending of these districts, after *controlling* for district characteristics. St. Louis is a very large district that has high levels of FRL and of percent black students that go along with its high spending. Thus, after adjusting for these other factors, Figure 3 indicates that St. Louis spends a bit below (but quite close to) the average of what would be predicted based on Missouri spending patterns.

St. Louis is far from alone in spending below the estimated average of comparable districts: approximately half the districts in the state fall in the same category, as Figure 3 shows. This is true by definition of averages; since Lake Wobegon is not located in Missouri, about half the districts will be above average and about half below average. The same logic that holds for simple averages carries over to the regression methodology, which estimates average spending among comparable districts. The large number of deficits we saw in Figure 2, for simple regression, appears again in Figure 3 – by construction. To interpret these shortfalls from the average as an adequacy shortfall would be logically absurd, since it would mean there is *always* an adequacy shortfall among about half the districts, no matter how high or low spending is. To

¹⁶ The collective bargaining environment is a textbook case of the violation of the competitive price-taking assumption for inputs, as the impersonal forces of the market are replaced by relative bargaining power.

¹⁷ Some practitioners (including Baker (2006b)) use regional cost indices instead of teacher salaries. This avoids some of the difficulties discussed above, but may only weakly reflect prices faced by districts.

¹⁸ Duncombe’s equation includes performance (instrumented), teacher salaries (instrumented), % FRL, % FRL x % black, % SPED, indicator for K-12 district, a set of indicators for district size, property values, district income, state aid relative to income, % college educated, % age 65 or older, % housing units owner occupied, median housing price relative to average property values, and a series of year indicators.

be sure, these deviations are not quite the adequacy shortfalls implied by the cost function – that requires one further step – but, as we shall see, the nature of those shortfalls is very largely determined by the deviations shown in Figure 3, results which follow ineluctably from the logic of averages.

Although the statistical controls do not affect the fact that about half the districts spend above and below average, they do affect the size of the deviations. Comparing Figures 2 and 3, we see that the controls help account for some part of the spread in spending over the upper and lower ranges of test scores, but did not much reduce the estimated spread in the mid-range. The spread containing the bulk of these districts remains about \$2,000-\$3,000, as it is over much of the test score range. In short, using statistical controls for observable district characteristics helps to identify some spending patterns (e.g. by FRL and race), but still leaves unexplained a wide range of spending among districts of similar observed characteristics and performance.¹⁹

The Econometrics of Spending Equations: Controlling for District Efficiency

To convert the spending equation to a cost function one needs to identify the minimum expenditure necessary to achieve any given level of performance – the definition of efficient. As Duncombe (2007) points out, “Because data is available on spending, not costs, to estimate costs of education requires controlling for differences in school district efficiency (p.3).”

It is increasingly common to deal with this issue by including “efficiency controls” -- variables which are thought to affect efficiency -- among the explanatory variables in the spending equation (1).²⁰ Unfortunately, there is no line item in budgets for “waste, fraud, and abuse.” Moreover, if it were obvious what factors determined inefficiency in schools, local and state citizens and authorities would be likely to take actions to correct the inefficiency. Thus, the quest for a set of observed and measurable factors that convert the spending functions into cost functions by separating inefficiencies from required spending is obviously difficult.

As one example of using efficiency controls, Duncombe’s equation for Missouri includes seven “efficiency-related variables,” categorized as either “fiscal capacity” variables, such as per pupil property values, income, and state aid, or “monitoring variables,” such as percent of population aged 65 or more and percent of college-educated adults in the district. The argument here is that districts with greater “fiscal capacity” may experience less pressure to be efficient (or a greater inclination to spend on non-tested subjects), and that older or college-educated voters may exert

¹⁹ This variation could be the result of inadequate controls for true differences across districts. For example, the percent of students eligible for free or reduced-priced lunch is likely to be a poor measure of the variation in resources that students receive at home across districts, especially across relatively high-poverty districts. Yet, these coarse measures are often the only measures available to researchers or to those designing and implementing school funding formulas. However, as previous analyses of achievement show, even with exceptional measures of district characteristics, much of the variation in achievement for districts with the same spending is likely to remain.

²⁰ Other methods have also been used, which attempt to identify statistically the points at or near the bottom of figures comparable to Figure 3. These methods, stochastic frontier analysis and data envelopment methods, have been used by Duncombe and others in earlier papers, but recent work, including that presented in court, focuses on the method discussed in the text. See, for example, Grosskopf, Hayes, Taylor, and Weber (1997), Duncombe, Ruggiero, and Yinger (1996).

greater “monitoring” for efficiency. No analysis – within this paper or elsewhere – directly relates any of these variables to efficiency – that is just a maintained hypothesis. In a similar analysis for California districts, Imazeki (2006) includes “efficiency controls,” but focuses on local competition instead of fiscal capacity or monitoring as her measure of efficiency, using the Herfindahl index for the number of districts in the labor market.

These variables are simply added into the spending equation (1). At this point, the equation is still a spending equation – all that has been done here is to single out a subset of the explanatory variables. A district’s age, education, income, property values, the competition it faces, etc. may well affect spending patterns, over and above the student demographics and other variables, and the estimation of (1) sheds further light on those patterns. One may or may not choose to interpret these variables as controls for efficiency (and, if so, they are certainly imprecise controls), but either way (1) remains a spending equation.

The typical procedure used to convert (1) from a spending equation to a cost function is to standardize the level of efficiency across districts by setting the values of the efficiency variables at uniform levels, rather than the actual district-specific values, and setting the error term to zero as given by Equation (2).

$$(2) \quad \begin{aligned} ("cost" \text{ of achieving current performance})_{it} &= \beta_0 + \beta_1 \cdot performance_{it} \\ &+ \beta_2 \cdot (teacher_salaries)_{it} + \beta_3 \cdot (\%FRL)_{it} + \dots + \beta_n \cdot (ave_prop_val) + \dots \end{aligned}$$

It is common in these cost-function analyses to set the value of the “efficiency controls” (such as property values per pupil) at the statewide average. Setting the error term to zero, of course, is also choosing the average. What this means is that about half the districts will be found to spend more and half less than the estimated “cost” of achieving at their *current* performance levels. This result is depicted in Figure 4, which presents the difference between each district’s actual spending and the estimated “cost” of achieving its actual performance level.

How are these figures to be interpreted? Spending for a district can be higher than cost because that district may not be using its resources wisely for maximizing the test performance of students. It is, on the other hand, *logically impossible* for a district to spend less than the minimum necessary to achieve actual performance levels. It would be one thing to recognize that “cost” may be imperfectly estimated and there could be a few outliers. But the estimation technique here *systematically* determines that spending is less than “cost” for about half the districts.²¹

Let us be clear on the source of the problem. One might think that the problem is the use of average values for the efficiency variables, rather than values that imply something closer to maximum efficiency (minimum spending). This is a valid criticism, but in fact the problem lies deeper.

²¹ Cost function analysts acknowledge that they are only estimating “average efficiency,” a term that would seem to modify the definition of cost. However, they continue to state that the estimated cost figures represent what is “necessary” or “required” to achieve any given result, which effectively restores the original definition. Figures 4 and 5 use the “required” terminology, from Duncombe (2007).

The primary source of the problem is that the “efficiency controls” do little to explain the variations in spending, and are rarely convincing measures of the full range of efficiency. The deviations depicted in Figure 4 have netted out the estimated effect of these variables on efficiency, but are still quite large. The step that purportedly converts the spending equation (1) to the “cost” equation (2) has very little effect on this fundamental problem.

Taking St. Louis as an example, the set of seven “efficiency” variables from Duncombe (2007) taken together tends to raise St. Louis spending above districts with average values of those variables, so the calculated “cost” using those averages is a bit lower than the fitted value in (1). Consequently, St. Louis’ slight deficit depicted in Figure 3 becomes a slight surplus in Figure 4: St. Louis spends slightly more than is “required” to achieve its actual test scores. As this example illustrates, for most districts there is not much difference between Figures 3 and 4. The interpretation placed on Figure 4 by the cost function methodology, however, is totally different – cost vs. spending. This re-interpretation is not defensible.

In short, the method that purports to convert average spending to cost does nothing of the sort. The adjustment from the “efficiency controls” is minor, not surprising given that it would be difficult to argue that these variables do a good job of measuring true variation in efficiency. The major step is that the deviations depicted in Figure 3 – deviations from average spending of comparable districts – are simply redefined as deviations from “cost.” That is why the “cost” estimates carry the logically incoherent implication that half the districts spend less than is necessary to achieve what they have achieved.

Extrapolating from the “Cost Function” to a Different Performance Level

The third step in calculating the cost of adequacy is to apply the estimated cost function to a target performance level. This step is accomplished by simply replacing actual performance with target performance in the calculation:

$$(3) \quad \begin{aligned} ("cost" \text{ of achieving target performance})_{it} &= \beta_0 + \beta_1 \cdot (target_performance) \\ &+ \beta_2 \cdot (teacher_salaries)_{it} + \beta_3 \cdot (\%FRL)_{it} + \dots + \beta_n \cdot (ave_prop_val) + \dots \end{aligned}$$

For example, one of the targets considered in Missouri was to raise St. Louis from its current level of 16.7 to a level of 26.3.²² If we apply this target to all districts (not just St. Louis), the result is to raise the “cost” for those districts below 26.3 and to reduce it for those districts above, relative to their estimated cost of current performance, provided that the estimate of β_1 is positive. This obviously increases the estimated shortfall from required spending for the former and reduces it for the latter. This imposes a substantial “tilt” on Figure 4, pushing down the points on the left side of the diagram and pushing up the points on the right.

The result is Figure 5, depicting actual spending vs. “required” spending to achieve the performance target of 26.3. These estimates redistribute the shortfalls from higher-performing

²² Duncombe (2007) identifies this as the Missouri School Improvement Program (MSIP) standard for St. Louis in 2008. This target happened to be near the state average in 2005, of 25.6.

districts to lower-performing ones.²³ For example, St. Louis was depicted in Figure 4 as spending slightly *above* what was “required” to achieve its current level, but Figure 5 depicts St. Louis as \$1,541 *below* what is “required” to achieve the higher target. Districts with lower performance are adjusted even further, to yield estimated shortfalls of over \$4,000.

To assess whether the estimates of cost in Figure 5 are valid, we must directly assess the two key features of the cost estimates: (1) the methodology for estimating the “cost” of generating *current* outcomes – which we have already seen is fundamentally flawed – and (2) the estimated coefficient β_1 which is applied to that base, to generate the “cost” of *target* performance.²⁴ The estimate of β_1 is key to the whole exercise, so it is critically important that it is estimated accurately, with a high degree of confidence, and that it not be sensitive to arbitrary choices in model selection. Unfortunately, there are several reasons why this standard is not met.

Imprecision in Estimated Coefficients, and in Estimated “Cost”

The first problem is that the regression coefficients are often estimated with relatively wide confidence intervals, even assuming that the model is correctly specified and appropriately estimated (assumptions we revisit below). For example, Duncombe’s estimate of β_1 is that costs rise by 0.39 percent for every 1 percent increase in performance. However, the 95 percent confidence interval ranges from 0.07 percent to 0.71 percent, spanning a factor of 10. Similarly, in the study of California districts, the 95 percent confidence interval for Imazeki (2006) estimates range from 0.05 percent to 0.63 percent. Even if everything else is correct, one can have very little confidence in the adjustments that lead to estimates of needed costs, moving from Figure 4 to Figure 5.

The problem of wide confidence intervals applies to the other coefficients as well, which is a matter of some importance for the issue of demographic cost premiums. For example, Duncombe’s estimate of β_3 implies a premium for FRL students of 52 percent, but the 95 percent confidence interval is 27 percent to 80 percent. Similarly, there is an implied premium for students in special education of 49 percent, but the 95 percent confidence interval is 19 percent to 80 percent. Again, these are very wide confidence intervals, and even they assume everything else is estimated correctly.

The imprecision in all the estimated coefficients, along with the estimated variance in the unexplained component of (1), contribute to wide confidence intervals in the estimated “cost” of meeting performance targets. For St. Louis, the estimated “cost” of performing at a level of 26.3 in 2005 is \$11,597. However, the 95 percent confidence interval is from \$8,367 to \$16,074.

²³ The fact that the process starts from a logically flawed base can still be seen in Figure 5, by examining the large number of red dots to the right of the vertical line. These are districts that are found to spend less than “required” to meet the standard that they are already meeting.

²⁴ Duncombe (2006) and Baker (2006a) have argued that the upward tilt in diagrams such as this in Kansas (Duncombe) and other states (Baker) provide some evidence in support of the approach’s statistical validity (albeit a “fairly weak validity test” in Duncombe’s view). However, as our step-by-step derivation shows, the tilt simply reflects the estimated sign of β_1 . The point is that *any* positive estimate of β_1 , even if it is highly problematic (for reasons such as those discussed in the next section), will necessarily generate a positive tilt in a diagram such as Figure 5. Thus, a positive tilt is of no independent value in assessing the validity of the cost function estimates.

Since this interval contains the current level of \$10,056, one cannot conclude that spending is inadequate to achieve that target at conventional confidence levels, even if the rest of the analysis is solid. In addition to the problems identified above, however, there are special problems with estimating β_1 , to which we now turn.

Special Problems with Estimating β_1 , the Cost of Raising Performance

There is a long history of trying to estimate the relationship between average spending and average performance, and it is not an encouraging one. For decades, it has proven difficult to find a systematic relationship, and the problems that have plagued that research also pertain to the cost function estimates. For one thing, the control variables are imperfect, the choice of variables is arbitrary in some cases, and the estimates are often sensitive to that choice.

More importantly, it may be that spending affects performance, as opposed to the opposite that is assumed in the spending relationships that are estimated. Indeed, the whole theory of the court case is precisely that – that providing more resources leads to higher achievement. The implications of this are very serious for the estimation of the spending/cost relationships, because β_1 will now reflect both effects even though just the impacts of achievement on costs are desired. A related problem is the worry of omitted variables that comes from the possibility of a third factor such as parents' interest in education affecting both spending and achievement. Both of these problems give reason to believe that β_1 is likely to be estimated with bias.

Cost function analyses often try to use instrumental variables techniques to reduce bias. However, the requirements of this technique are difficult to fulfill and cost functions to date have not utilized convincing instruments.

Finally, the estimates differ dramatically depending on the specification, whether spending is modeled as a function of achievement or achievement is modeled as a function of spending.

(i) Sensitivity to Selection of Other Variables

The first problem is that the results are often highly sensitive to which variables are included in the model. For example, in both the Duncombe and Baker models for Missouri, the results are highly sensitive to the inclusion of race. If race is excluded from the model (as it surely would be, if it were to be used for an actual funding formula), the coefficient on performance, β_1 is no longer statistically significant, which is to say the 95 percent confidence interval includes zero.

Similarly, estimates in Baker (2006b) are highly sensitive to which “efficiency controls” are included in the estimating equation. His data set contains six such variables – similar to those used by Duncombe – though he selects only four of them. Among the 64 possible combinations of those six controls, the β_1 estimate is statistically indistinguishable from zero almost half the time, and in most of those cases the model's “fit” is better than the one chosen by Baker. One cannot have much confidence in any single estimate of β_1 if both the estimates and the confidence intervals are so highly sensitive to arbitrary choices in model specification.

These sensitivities are found in other states as well. Results provided in Duncombe (2006) show that the estimate of β_1 for Kansas loses its statistical significance if an interaction term is omitted (free lunch multiplied by pupil density). If the time period 2000-2004 is broken up into 2000-1 and 2003-4, the estimate for β_1 doubles between these subperiods, but for neither period is it statistically significant.

(ii) Endogeneity Bias, Omitted Variables Bias, and “Instrumental Variables”

A second problem is statistical bias due to mutual causation between spending and achievement (“endogeneity bias”) and/or omitted variables that are likely to affect, or at least be correlated with, both spending and performance. For example, some districts are more education-oriented than others, simply due to the gathering of like-minded citizens, with specific characteristics that are not captured by the observable variables. These districts may tend both to spend more and to have more highly performing children. If so, then the relationship between spending and performance will be biased upwards, since their statistical association will be picking up in part the effect on each of them of the unobserved degree of education-orientation.

The usual solution to this problem is a technique known as “instrumental variables.” Under this technique, “performance” is considered a “troublesome explanator” for spending and does not actually enter into the estimating equation (1).²⁵ Instead a proxy variable or set of variables is used, known as “instruments.” The idea is that instead of using variation in achievement that could be a result of a third variable that also affects spending and thus is subject to bias, this technique uses only the variation in achievement that comes from a known source that does not independently affect spending. The theory of this approach is compelling; however, in practice it is rarely well implemented. The problem is that this technique has some very stringent requirements, which are rarely met. In the context of cost function estimation, it is very difficult to identify variation in achievement that is the result of factors that do not independently influence spending. If these conditions are not met, the instrumental variable solution to the problem of bias can easily make the problem worse.²⁶

There are statistical tests that provide some defense against using invalid instruments, and at a minimum the cost functions should pass the relevant test. These tests are weak, since they have to assume that some of the instruments are valid, in order to test whether all of them are; yet, in the case of Missouri, the instruments failed these tests for both cost functions submitted to court. Thus, the adequacy estimates were not only methodologically flawed, but statistically invalid.

In addition to the problem of invalid instruments, which lead to biased estimates, there is an additional problem of weak instruments – proxy variables that are only weakly correlated with performance. This leads to an overstatement of the statistical significance of the performance coefficient. In other words, the claim that β_1 – the key coefficient in the whole exercise – is statistically distinguishable from zero, is often undermined by weak instruments. A final difficulty in the instrumental variables approach is that the choice of instruments may be

²⁵ When “teacher salaries” is used as an input price control (as in Duncombe (2007)), it is also treated as a “troublesome explanator,” and instrumented.

²⁶ Murray (2006)

somewhat arbitrary and the estimated performance coefficient may be quite sensitive to the choice of instruments.²⁷

(iii) Sensitivity to Specification as “Cost” vs. “Production” Function

Finally, cost estimates are extremely sensitive to whether spending is modeled as a function of achievement or achievement as a function of cost. There are two traditions looking at the relationship between student performance and spending: the production function approach and the cost function approach. The key difference between the two is whether the focus of attention is achievement or spending. Each approach standardizes for a variety of other factors such as economic disadvantage of families, district attributes such as population density, and other things, and then looks at the remaining correlation of spending and achievement. The difference is whether spending is on the left side of (1) and performance on the right (cost function) or whether these are reversed (production function).

The first thing to note is that these two approaches must necessarily be related. After all, they look at the relationship between the same basic elements of achievement and spending. Viewing them together provides an easy interpretation of the empirical evidence, but unfortunately this is seldom done. The one exception, where production function and cost function approaches are placed side by side, is Imazeki (2006). Imazeki’s analysis finds that achieving adequacy in California is estimated to require additional spending of \$1.7 billion if a cost function estimate is used and \$1.5 trillion if a production function estimate is used – clearly a striking difference.

Both the cost function and production function estimates show weak and imprecise relationships between average district spending and average student achievement, as illustrated in Figure 6 for eighth-grade math scores in the 522 districts of Missouri in 2006. After allowing for differences in the free and reduced price lunch populations, in the racial composition (percent black), and in the number of students, one can plot achievement against spending in a way that uses statistical methods to control for the other characteristics mentioned.

The figure shows that there is a slight upward slope of the spending line, but the dominant picture of Figure 6 is (once again) essentially a cloud – where districts with the same spending get wildly different achievement. The line has a statistically significant but relatively small positive slope of 0.0028 scale points per dollar ($t=3.1$).²⁸ The flatness of this line is important: spending more money given the current way it is spent yields very little achievement gain. Put another way, if one wishes to get a large change in achievement (as discussed below), it will cost a very large amount of money, even assuming that this linear relationship can be extended far away from the current spending: it costs \$357 per pupil to raise achievement one point.

²⁷ The Missouri cost functions suffered from both problems discussed in this paragraph, although the point was somewhat moot since the instruments chosen were invalid.

²⁸ It should be pointed out that Figures 6 and 7 (below) are not necessarily representative of all student outcome measures. If one took a different grade to look at these relationships or looked at reading instead of math, most alternatives actually give insignificant relationships between spending and achievement, and frequently they have the wrong sign. This might be expected since the regressions are drawing lines through these clouds of points with very little shape to the points that allow estimating such a relationship. A few districts performing at a slightly different point in the cloud can change the slope of the relationship.

Figure 6 is not very encouraging for the proponents of reaching adequate levels of performance through solely spending more money. If it requires tripling or quadrupling funds to get students to the adequate level, most reasonable people will immediately see that this is not a viable public policy.

But there is another way of looking at the data. By looking at how spending varies with achievement – the cost function approach that we have been discussing above – the picture looks far more manageable. Figure 7 turns the previous picture on its side and looks at the amount of spending as a function of achievement (after allowing for the same factors of free and reduced price lunch, race, and district size). Again the dominant feature is the cloud of districts that spend very different amounts to reach any given performance level. But now the line that goes through the points tells a very different story. It is flat once again, but this now indicates that one can move across very large achievement levels at modest cost. The regression coefficient indicates that each \$6.62 raises achievement one point.

These regression coefficients reflect the same data – they both have identical t-statistics of 3.13 – but they differ dramatically on the estimated cost of raising achievement: \$357 per point vs. \$6.62 per point, a factor of 54 (and of course this ignores the wide confidence intervals around each of these estimates). The ultimate reason that these estimates differ so much, even though they use the same data, is that the fit is not very tight. If the fit were perfect, the estimates would coincide: turning the diagram on its side would not only turn the dots on their side, but also the line. However, when the fit is so weak, each diagram will generate a flat curve, because they are each minimizing the variation in error terms measured vertically.

The cost function makes it appear that it is much more feasible to change achievement by simply spending more with the current schools and the current institutional arrangements. For example, in Missouri, the average score on math is 733 and proficiency is defined as 800, so there is a gap of 67 points. Under the “cost” function estimate, it costs $67 \times \$6.62 = \443 per student to close the gap. Under the production function estimate, the “cost” is \$23,919. When the estimates vary so wildly from two equally defensible ways of looking at the data – neither one of which finds a strong relationship – it is hard to place much credence on either estimate.

Conclusion

Determining the dollars necessary to provide an adequate education is not an easy task. The commonly employed technique of using professional judgment to design prototype schools is far from satisfying. Case studies of particularly successful schools may provide insights into effective approaches but are also unsatisfying because success is often the function of particularly dynamic leadership or teaching that may be difficult to replicate under current institutional arrangements. Regression-based approaches, often called “cost function analyses,” provide a superficially attractive alternative because they apply seemingly objective methods to data on district spending and achievement to determine the cost of reaching standards.

While on first blush the regression-based approaches are appealing, on further exploration, as discussed above, they are fraught with problems, revealing very little about the cost of improving student achievement. The issues facing regression-based models are of two over-arching types: technical problems that skilled analysts with sufficient data can correct in their models and conceptual problems that bring the overall approach into question.

Given sufficient data, a skilled analyst can estimate a regression-based model to produce informative estimates of the spending patterns, by district characteristics and outcomes. Even the most skilled analyst, however, will typically find “cost” estimates that are highly imprecise, sensitive to judgment calls in modeling, and subject to bias.

The underlying difficulty is that even after controlling for a host of variables (including labor market prices, student and school characteristics, among other variables) there is still a great deal of variation across districts in their outcomes for students, in districts with the same expenditures. There are a number of reasons for these differences that draw the regression-based approaches into question. In particular, we have little way of knowing how much of these differences are driven by unobserved cost or price differences, by mismanagement, or by a focus on goals other than the student achievement measures used in the cost functions.

Cost function analysts are aware of these problems. They use efficiency controls and instrumental variables approaches to adjust for these difficulties. However, as demonstrated above, in practice both approaches fall woefully short of convincing. We simply do not have good measures of efficiency. The proxies that have been used are, at best, weak measures of efficiency with substantial measurement error, and measurement error itself creates bias. Instrumental variables can, in theory, address the biases due to omitted variables and mutual causation, but, in practice no researchers have identified strong and valid instruments. Weak and invalid instruments have been shown repeatedly to overstate statistical significance and to increase bias rather than mitigate it.

The usual practice of identifying “cost” as the average spending among comparable districts always yields the logically impossible result that about half the districts spend less than is required to achieve what they have achieved. This problem has practical implications as well. If courts and policy-makers accept a methodology that defines *minimum* expenditures by *averages*, they will then have to raise the expenditures of those below the average, thereby raising the average again. This methodology is a recipe for perpetual findings of inadequacy under forever-recurring litigation.

The failure of regression-based approaches to identify the cost of adequacy is nowhere as clear as when comparing the results of spending as a function of achievement to those of achievement as a function of spending. Cost functions assume that spending changes as a function of achievement; but it makes just as much sense, if not more in the case of education, to assume that achievement changes as a function of spending. A comparison of these two approaches, however, produces vastly different estimates with vastly different implications for policy if interpreted as identifying the causal effect of spending on achievement. Of course, such an interpretation is not warranted.

The “cost function” approach simply does not identify the causal relationship between spending and achievement. This failure should not be surprising. We would not need randomized

experiments or detailed longitudinal data on student learning to estimate the effects of resources if this could be done so simply with district-level data on spending and average student achievement. However, while not surprising, the problems with regression-based approaches do highlight the difficulty of basing school finance decisions on currently available estimates of the cost of adequacy. All techniques for estimating the cost of adequacy are seriously flawed. None of them can provide a convincing cost figure.

At best, each method provides some limited information – the current distribution of spending and achievement, the cost of a variety of prototype schools, the activities and expenditures in some particularly successful schools. This information can be better than no information for what is ultimately a political decision of how much to spend, but it cannot provide a dollar figure that will guarantee student success or even the opportunity for student success. The most important lesson that emerges from the data – with its wide variation in achievement for comparable expenditures – is that how money is spent is crucial for determining student outcomes. Educational excellence requires a system with the knowledge, professional capacity, incentives and accountability that will lead schools to determine how to spend their funds most effectively to raise student achievement and reach the variety of goals we have for students.

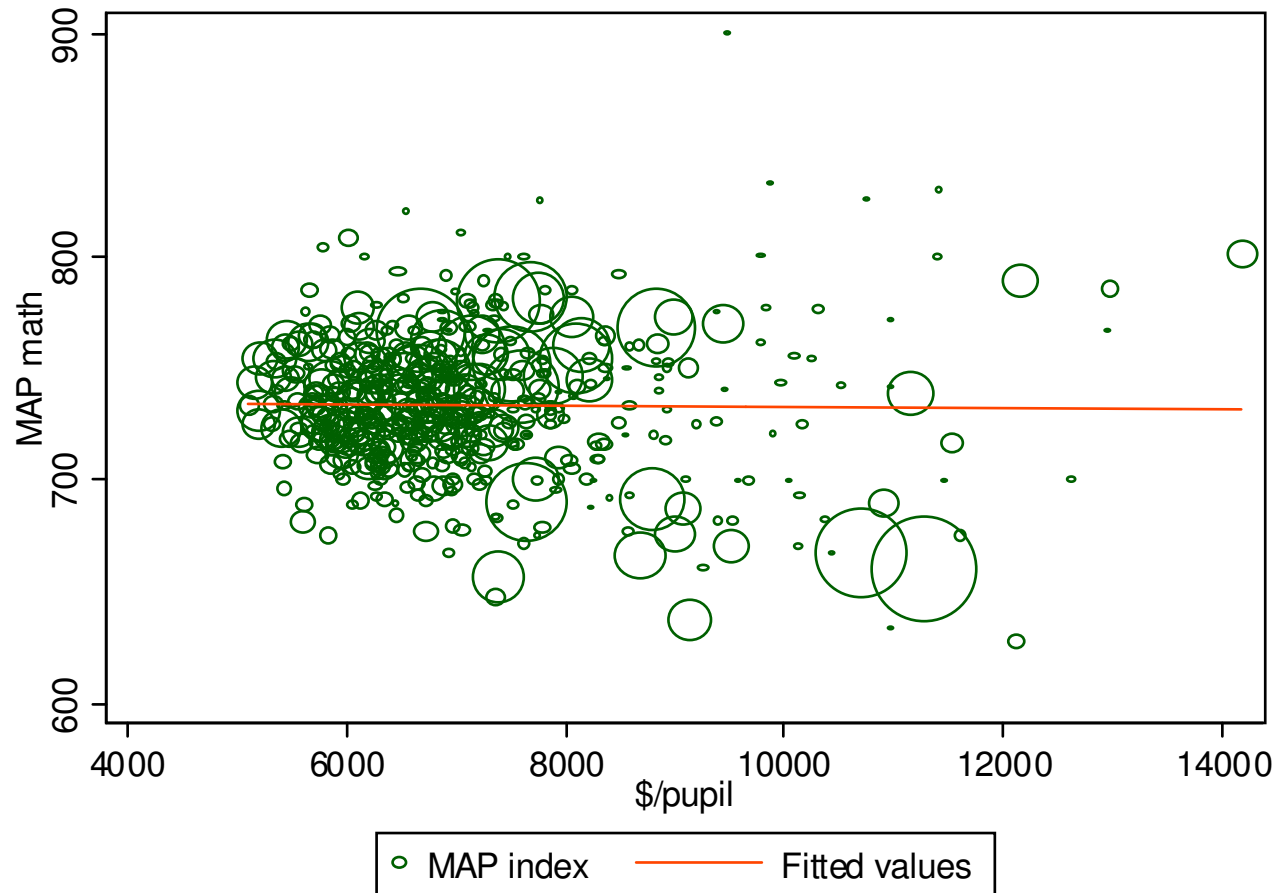


Figure 1. Missouri District Average 8th Grade Mathematics Scores and District Spending: 2006

Figure 2: Expenditures vs. Performance in Missouri, 2005

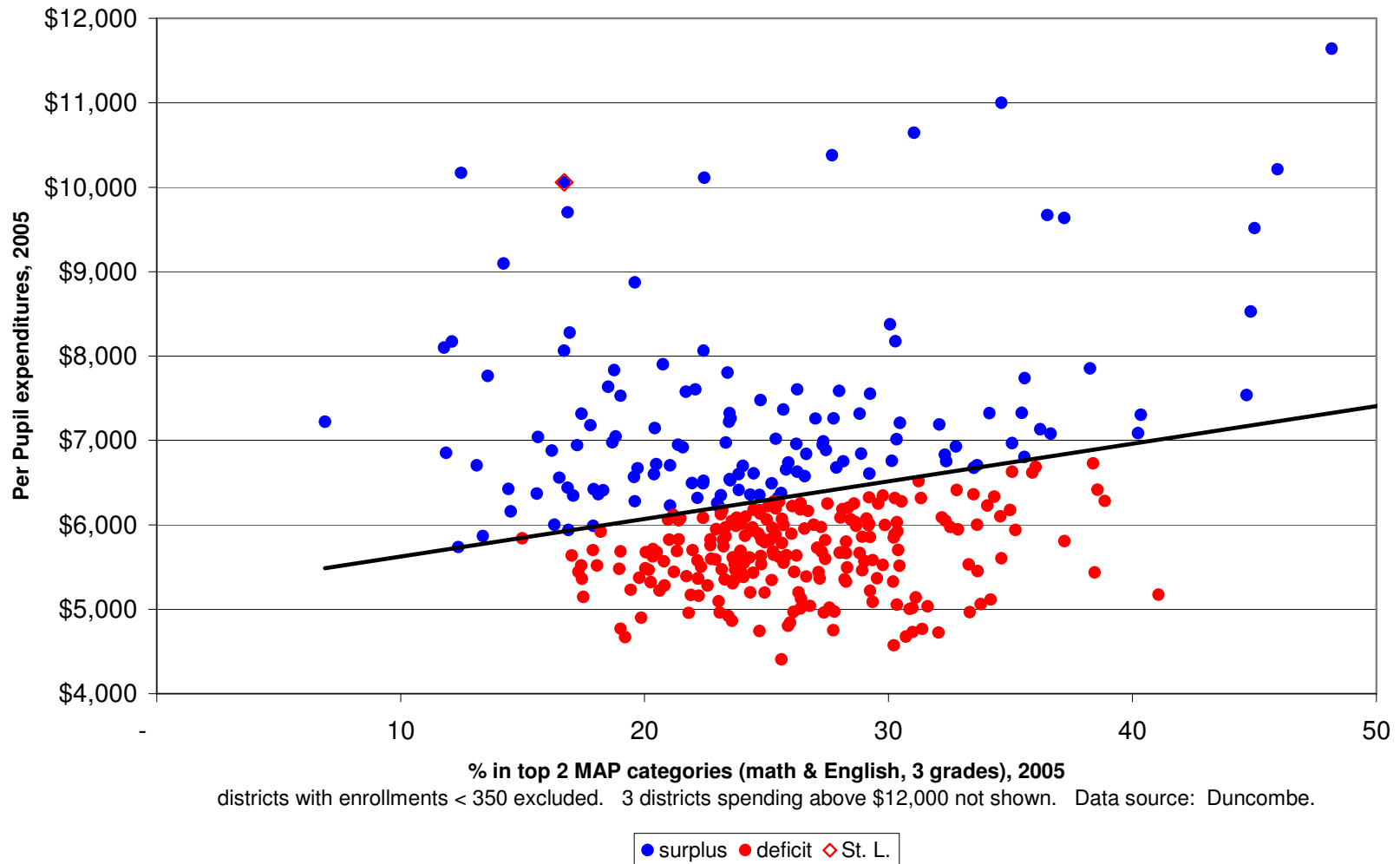
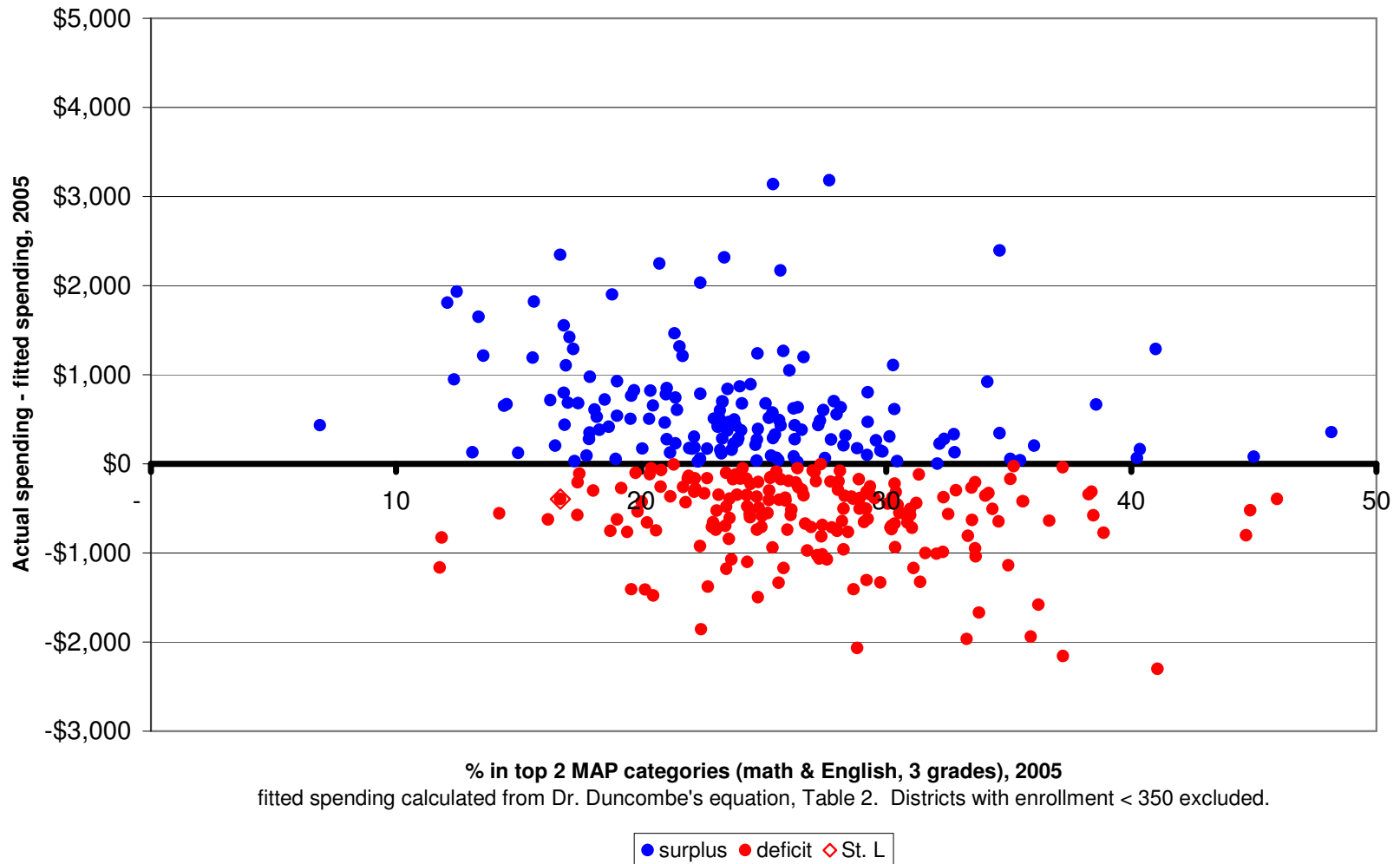
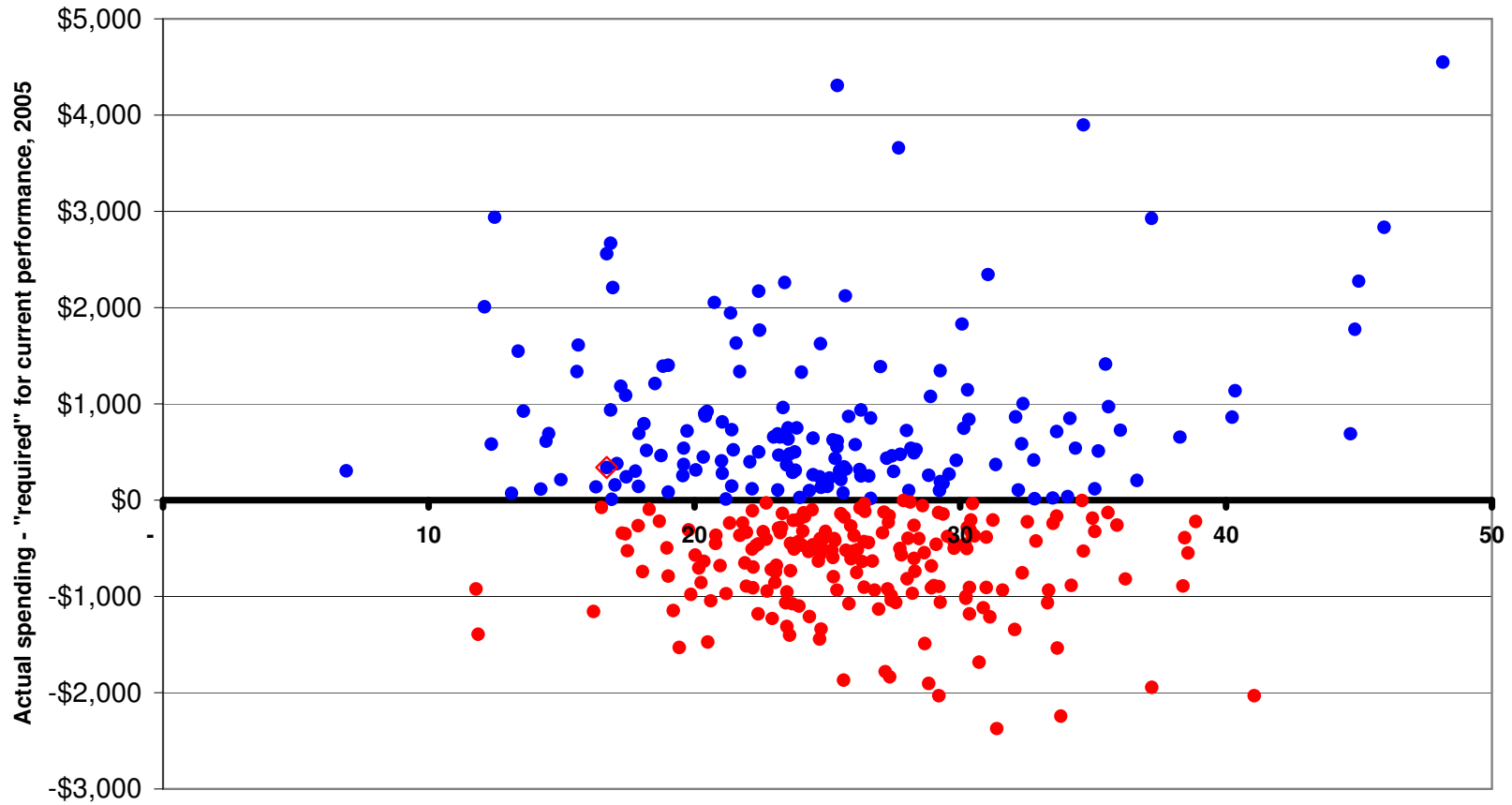


Figure 3: Actual vs. Fitted Spending, with Controls for District Characteristics



fitted spending calculated from Dr. Duncombe's equation, Table 2. Districts with enrollment < 350 excluded.

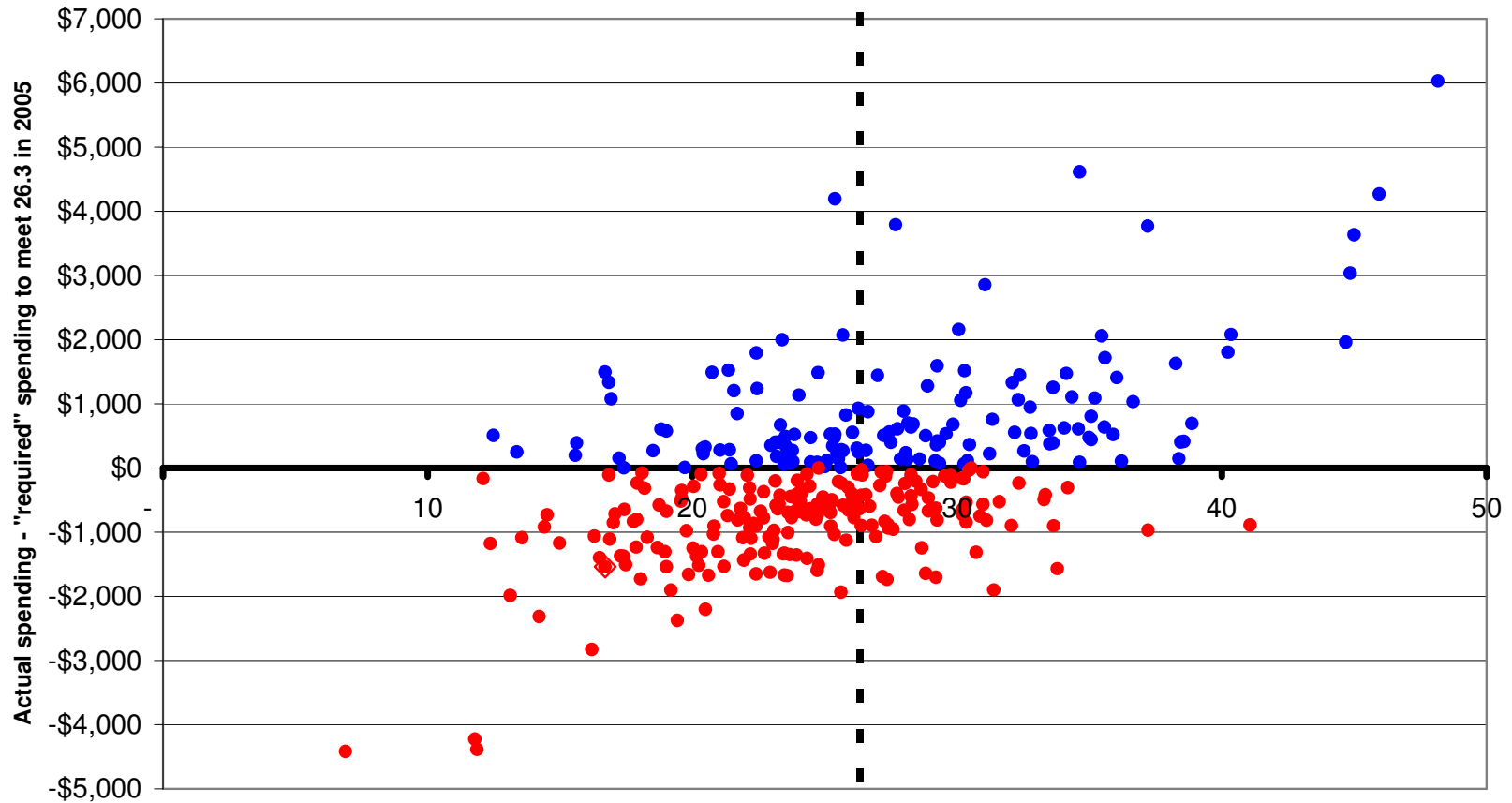
Figure 4: Actual vs. "Required" Spending to Achieve Current Performance



% in top 2 MAP categories (math & English, 3 grades), 2005
"required" spending calculated from Dr. Duncombe's equation, Table 2. Districts with enrollment < 350 excluded.

● surplus ● deficit ◇ St. L

Figure 5: Actual vs. "Required" Spending to Achieve Target Performance



% in top 2 MAP categories (math & English, 3 grades), 2005

"required" spending calculated from Dr. Duncombe's equation, Table 2. Districts with enrollment < 350 excluded.

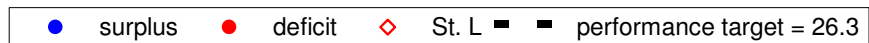


Figure 6. “Production Function” Relationship between 8th Grade Math and Spending (holding constant race, enrollment, and free or reduced lunch eligibility), Missouri

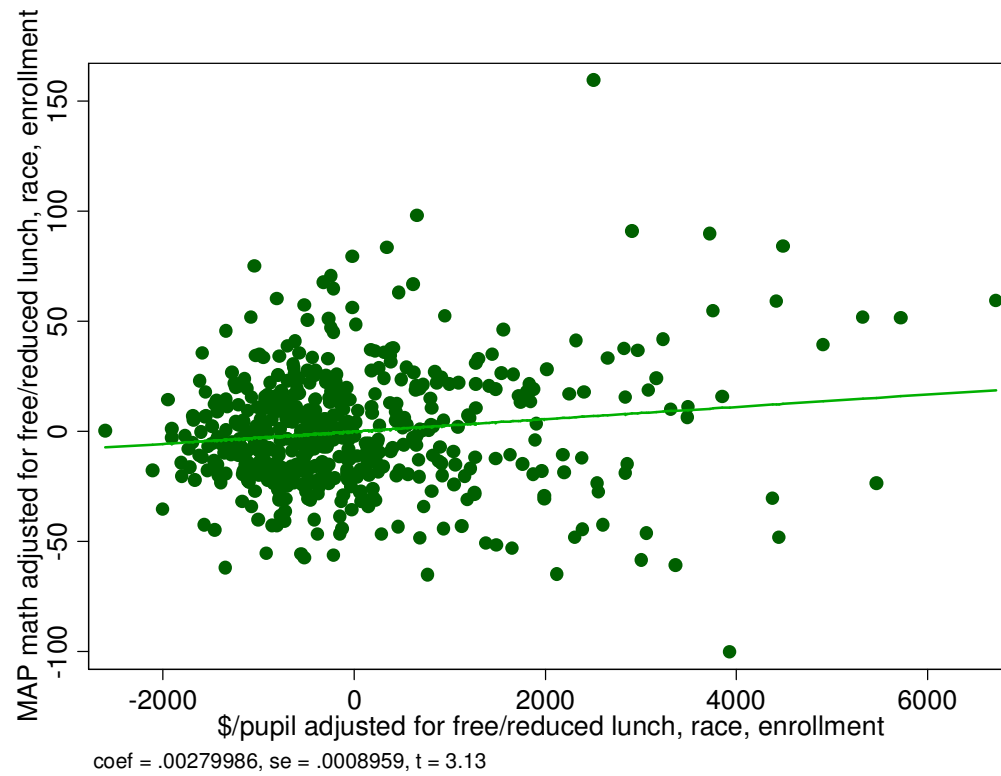
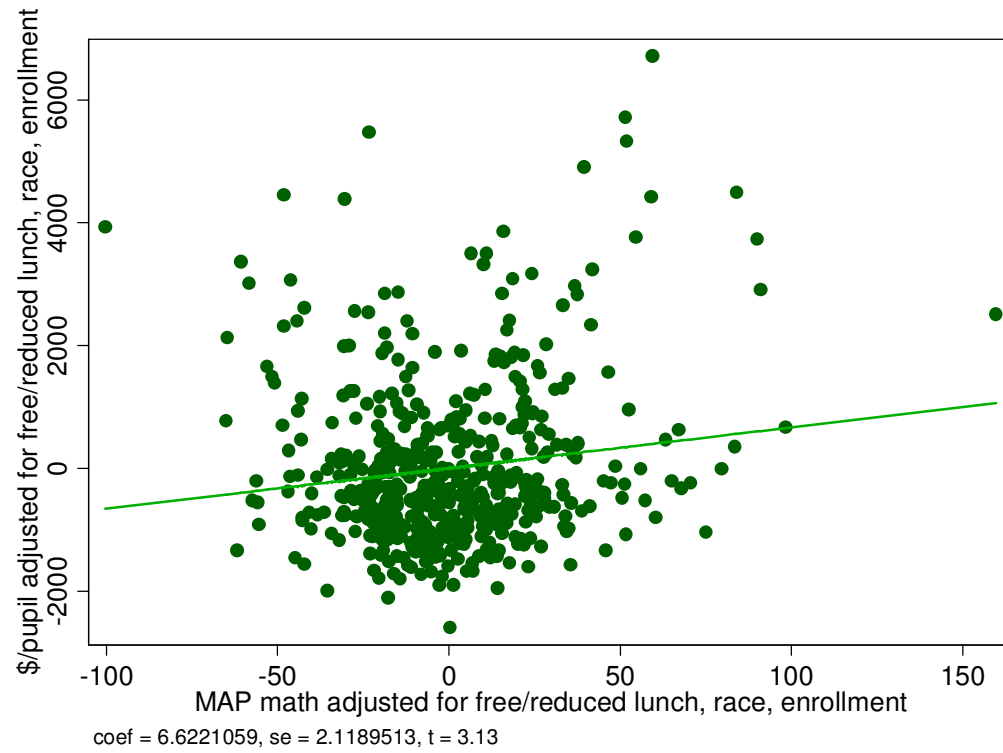


Figure 7. “Cost Function” Relationship between Spending And 8th Grade Math (holding constant race, enrollment, and free or reduced lunch eligibility), Missouri



References

- Baker, Bruce D. 2006a. "Evaluating the Reliability, Validity, and Usefulness of Education Cost Studies." *Journal of Education Finance* 32,no.2 (Fall):170-201.
- . 2006b. *Missouri's school finance fomula fails to guarantee equal or minimally adequate educational opportunity to Missouri schoolchildren*. Lawrence, KS: University of Kansas
- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Costrell, Robert M. 2007. "The Winning Defense in Massachusetts" In *School Money Trials: The Legal Pursuit of Educational Adequacy*, edited by Martin R. West and Paul E. Peterson. Washington: Brookings:278-304.
- Duncombe, William. 2006. "Responding to the Charge of Alchemy: Strategies for Evaluating the Reliability and Validity of Costing-Out Research." *Journal of Education Finance*.
- . 2007. *Estimating the Cost of Meeting Student Performance Standards in the St. Louis Public Schools* (January).
- Duncombe, William, John Ruggiero, and John Yinger. 1996. "Alternative approaches to measuring the cost of education." In *Holding schools accountable: Performance-based reform in education*, edited by Helen F. Ladd. Washington, DC: Brookings:327-356.
- Duncombe, William, and John Yinger. 1997. "Why is it so hard to help central city schools?" *Journal of Policy Analysis and Management* 16,no.1:85-113.
- . 2000. "Financing higher student performance standards: the case of New York State." *Economics of Education Review* 19,no.4 (October):363-386.
- . 2005. "How much more does a disadvantaged student cost?" *Economics of Education Review* 24,no.5 (October):513-532.
- . 2007. "Does School District Consolidation Cut Costs?" *Education Finance and Policy* 3, no. 4 (Fall): 341-375.
- Gronberg, Timothy J., Dennis W. Jansen, Lori L. Taylor, and Kevin Booker. 2004. *School Outcomes and School Costs: The Cost Function Approach*. Texas A&M University

- Grosskopf, Shawna, Kathy J. Hayes, Lori L. Taylor, and William L. Weber. 1997. "Budget-constrained frontier measures of fiscal equality and efficiency in schooling." *Review of Economics and Statistics* 74,no.1 (February):116-124.
- Hanushek, Eric A. 2003. "The failure of input-based schooling policies." *Economic Journal* 113,no.485 (February):F64-F98.
- . 2006. "Science Violated: Spending Projections and the "Costing Out" of an Adequate Education." In *Courting Failure: How School Finance Lawsuits Exploit Judges' Good Intentions and Harm Our Children*, edited by Eric A. Hanushek. Stanford: Education Next Books:257-311.
- . 2007. "The alchemy of 'costing out' an adequate education." In *School Money Trials: The Legal Pursuit of Educational Adequacy*, edited by Martin R. West and Paul E. Peterson. Washington: Brookings:77-101.
- Imazeki, Jennifer. 2006. "Assessing the costs of K-12 education in California public schools." Stanford, CA, Stanford University (December 1).
- Imazeki, Jennifer, and Andrew Reschovsky. 2004a. *Estimating the Costs of Meeting the Texas Educational Accountability Standards*. (July 9).
- . 2004b. "School finance reform in Texas: A never-ending story?" In *Helping children left behind: State aid and the pursuit of educational equity*, edited by John Yinger. Cambridge, MA: MIT Press:251-281.
- Ladd, Helen F., Rosemary Chalk, and Janet S. Hansen, eds. 1999. *Equity and Adequacy in Education Finance: Issues and Perspectives*. Washington: National Academies Press.
- Murray, Michael P. 2006. "Avoiding Invalid Instruments and Coping with Weak Instruments." *Journal of Economic Perspectives*(Fall):111-132.
- Podgursky, Michael, James Smith, and Matthew G. Springer. 2007. "A New Defendant at the Table: An Overview of Missouri School Finance and Recent Litigation." In *Show-Me Institute Conference on Education Finance*. University of Missouri-Columbia.
- Reschovsky, Andrew, and Jennifer Imazeki. 2003. "Let No Child Be Left Behind: Determining the Cost of Improving Student Performance." *Public Finance Review* 31(May):263-290.