

Cost First, Efficiency Second

A Cost-Adjusted Production Approach to Evaluating School District Efficiency

Bruce D. Baker

University of Miami educpolicy@gmail.com

ABSTRACT

When a school district's outcomes fall short, is it because the district lacks resources relative to the cost of reaching common outcomes, or because it uses the resources it has poorly? Conditional on resources relative to cost, how far does each district fall short of the best attainable outcome, and what distinguishes those that fall furthest short? For four decades, researchers have tried to estimate the relative efficiency of schools and districts directly, fitting production or cost models by ordinary least squares, data envelopment analysis or stochastic frontier analysis (SFA) and reading efficiency off the residual. Production models generally fail because they cannot fully account for differences in the cost of producing outcomes; what they call inefficiency is largely unmeasured need. Cost efficiency is the better framing, but direct estimation of a cost frontier asks one model to price cost factors, handle endogenous outcomes and separate inefficiency from noise all at once. This paper takes a sequential approach. I first estimate an education cost function, the National Education Cost Model (NECM), compare four alternative specifications of student need and racial composition, and select among them on instrument strength, out-of-sample prediction, the strength of the funding gap–outcome gap relationship and residual bias by racial composition. I then evaluate efficiency in a cost-adjusted production framing: a stochastic frontier relating each district's outcome gap to its spending relative to NECM predicted cost. I show that the residual of the funding gap–outcome gap relationship is the output-space image of the cost function error, which is why it carries information about efficiency, and why it does so only when cost is correctly priced. That residual combines three things that are hard to separate: measurement error, omitted variables bias and real differences in efficiency. Under race-neutral cost models, residuals fall by 0.43 to 0.49 SD for each unit increase in Black enrollment share; under the preferred Black $\times$ density model the bias largely disappears (+0.08). Applied to 128,713 district-years in 38 states (2009–2024), the frontier yields efficiency rankings nearly identical to a simple residual ranking (rank correlation 0.997). Relative to their own state's averages, the least efficient districts are smaller, spend more, have similar funding relative to cost and are more heavily staffed in every category, about 1 SD above the state mean in total staff per pupil. Their excess administrative spending is real but small, about \$80 to \$100 per pupil.

Keywords: education cost function; efficiency; stochastic frontier; school finance; staffing

JEL codes: I21, I22, H75, C26

Suggested citation: Baker, B.D. (2026, October). Cost first, efficiency second: A cost-adjusted production approach to evaluating school district efficiency. School Finance 101 Working Paper.

Version: October 2026. Available at <https://schoolfinance101.com/recent-papers-articles/>.

Cost First, Efficiency Second

1. Introduction

The debate over school funding has shifted from whether money matters to whether districts waste what they have. A decade of quasi-experimental evidence shows that sustained increases in school spending improve achievement and attainment (Jackson, Johnson, & Persico, 2016; Lafortune, Rothstein, & Schanzenbach, 2018; Jackson & Mackevicius, 2024). Opponents of additional funding now argue more often that districts are inefficient: too many administrators, too much non-instructional spending, too little to show for it. Those claims usually rest on comparisons of spending and test scores that ignore what it costs a given district, with its students and its setting, to reach a common outcome.

This paper asks a question those comparisons cannot answer. Conditional on its resources relative to the cost of reaching common outcomes, how far does each district fall short of the best outcome attainable, and what distinguishes the districts that fall furthest short? The answer matters for policy. If the districts with the weakest outcomes for their resources are those with bloated central offices, the case for redirecting administrative spending is strong. If they are districts that look like their peers on funding relative to cost but differ in scale, staffing or student need the cost model does not fully price, the case is weaker, and the more productive response is closer study of those districts rather than a general claim of waste.

Researchers have long tried to do better by estimating efficiency directly. The production function tradition regresses outcomes on spending or purchased inputs and student characteristics, and treats a district's shortfall from the predicted or best-practice outcome as inefficiency (Hanushek, 1986; Ruggiero, 1996; Grosskopf, Hayes, Taylor, & Weber, 1997; Chakraborty, Biswas, & Lewis, 2001). The cost function tradition reverses the model, regressing spending on outcomes, input prices, student needs and scale, and treats spending above the predicted or frontier cost as inefficiency (Duncombe, Ruggiero, & Yinger, 1996; Gronberg, Jansen, & Taylor, 2011; Gronberg,

Jansen, Karakaplan, & Taylor, 2015). Both have been estimated with ordinary least squares, data envelopment analysis (DEA) and stochastic frontier analysis (SFA).

Production models generally fail for a simple reason. They cannot fully account for differences in cost. A district serving high concentrations of poverty, students with disabilities or English learners, in a small or sparse setting, in a high-wage labor market, needs more resources to reach the same outcome. When those differences are only partly captured by control variables, the shortfall attributed to inefficiency is largely unmeasured need. In the data used here, raw per-pupil spending is essentially uncorrelated with district outcomes (enrollment-weighted $r = -0.08$ in 2022–2024). Spending relative to NECM predicted cost is strongly correlated with them ($r = 0.60$). A production model built on raw spending would see little return to resources and would rank high-need, high-spending districts as wasteful.

Cost efficiency is the better framing, because it starts from what outcomes cost. But estimating a cost frontier directly asks a single model to do three hard things at once: price the cost factors, deal with outcomes that districts choose jointly with spending, and separate inefficiency from noise. The cost function literature has never settled whether efficiency controls can separate necessary cost from waste (Costrell, Hanushek, & Loeb, 2008; Duncombe & Yinger, 2011), and the one-sided error of a cost frontier is defined on a spending residual that also absorbs every mispriced cost factor.

This paper argues for a sequential alternative. Estimate the cost function first, with the purpose of pricing costs well, not of measuring efficiency. Then evaluate efficiency in a cost-adjusted production framing: compare each district's outcomes with what similar resources, measured relative to cost, achieve elsewhere. The cost model does the work production models cannot, putting districts on a common footing of resources relative to need. The efficiency model then works in outcome space, where shortfalls are interpretable and where the decomposition of inefficiency from noise is applied to a measure districts and policymakers care about.

The analyses that follow illustrate this approach with the National Education Cost Model (NECM; Baker, Weber, & Srikanth, 2021; Baker, 2025b). Section 2 reviews the production and

cost approaches to efficiency. Section 3 summarizes the NECM method and four alternative specifications, following Baker (2025b), and the criteria used to choose among them. Section 4 shows formally why the residual of the funding gap–outcome gap relationship carries information about relative efficiency, and why that information is only as good as the cost model. It organizes the residual into three components, measurement error, omitted variables bias and real differences in efficiency, that the rest of the paper works to separate. Section 5 reports the alternative cost models, the gap relationship and a stochastic frontier built on the preferred model. Section 6 compares the staffing and fiscal allocations of more and less efficient districts. Section 7 compares the approach with alternatives, including panel frontiers and structural models, states what the estimates can and cannot be used for, and discusses limits.

2. Estimating efficiency directly: production and cost approaches

2.1 Production models

The production approach models outcomes as a function of school resources and student and community characteristics. Early education production function research found weak and inconsistent relationships between spending and outcomes (Hanushek, 1986), a result later attributed in part to the endogeneity of spending: compensatory aid flows to districts with greater need, so spending and unmeasured need move together. Efficiency studies built on this framework extend it by treating the shortfall from a best-practice outcome as inefficiency, using DEA (Ruggiero, 1996; Grosskopf et al., 1997, 2001) or stochastic frontiers (Chakraborty et al., 2001). Reviews of the efficiency literature document the range of methods and the sensitivity of results to them (De Witte & López-Torres, 2017).

The central weakness is that every cost factor omitted from, or mis-specified in, a production model passes into the efficiency residual. Ruggiero (1996) addressed this by conditioning DEA comparisons on environmental factors; others include student poverty, disability or English learner shares as nondiscretionary inputs. But the list of factors is never complete, their functional form is rarely tested, and their coefficients in a production model do not have cost interpretations that can

be validated. The result is efficiency rankings that track student need.

2.2 Cost models

Education cost functions model spending as a function of outcomes, input prices, student needs and scale (Duncombe et al., 1996; Duncombe & Yinger, 2005; Imazeki & Reschovsky, 2004). Because observed spending includes spending that does not raise measured outcomes, cost functions handle efficiency in one of two ways. The first includes efficiency controls, such as fiscal capacity, public monitoring and competition, and holds them at fixed values when predicting cost.¹ The second estimates a stochastic cost frontier directly, treating spending above the frontier as inefficiency (Gronberg et al., 2011, 2015). The two are not mutually exclusive. The most recent work by Taylor and colleagues, in the *Southern Economic Journal* and in the Kansas cost study, adopts a two-stage approach with instrumented outcomes and indirect efficiency controls but still estimates the cost function as a stochastic frontier (Gronberg, Jansen, & Taylor, 2017; Taylor, Willis, Berg-Jacobson, Jaquet, & Caparas, 2018).²

The direct cost frontier is the most natural way to measure cost efficiency, but it is demanding. Outcomes are endogenous to spending and require instruments; combining instrumental variables with a one-sided error is not straightforward. The one-sided error is identified from the skewness of a spending residual that also contains every mispriced cost factor. And the resulting efficiency score is a spending ratio, which tells a policymaker how much a district overspends for its outcomes

¹Indirect efficiency controls are intended to address an omitted variables problem. Efficiency is unobserved, and districts that are less efficient spend more for given outcomes. If efficiency is correlated with outcomes or with the cost factors in the model, leaving it out biases the estimated cost coefficients, for example by attributing to student need spending that reflects slack. The Duncombe–Yinger approach includes variables that, in theory, predict efficiency: fiscal capacity, which relaxes the budget pressure to economize, and measures of public monitoring and competition. The cost coefficients are estimated conditional on these controls, which are then held at fixed values when predicting cost (Duncombe & Yinger, 2005, 2011). The controls are a remedy for bias in the cost coefficients. They are not designed to measure the efficiency of individual districts, and the cost function residual still combines the three components described in Section 4.1: measurement error, any remaining omitted costs and real differences in efficiency.

²Combining indirect efficiency controls with a stochastic frontier asks the cost model to do both jobs: the controls address omitted variables bias in the cost coefficients, and the one-sided error is asked to capture whatever district-level inefficiency the controls leave. The approach in this paper keeps the indirect controls in the cost model, where they serve the first purpose, and moves the second step out of the cost model and into outcome space, where the frontier is estimated conditional on resources relative to the predicted cost.

but not how far its outcomes fall short of what its resources could produce.

2.3 Frontier methods and their limits

The stochastic frontier separates a one-sided inefficiency term from symmetric noise (Aigner, Lovell, & Schmidt, 1977; Meeusen & van den Broeck, 1977), with unit-level efficiency recovered from the conditional distribution of inefficiency given the composed residual (Jondrow, Lovell, Materov, & Schmidt, 1982; Battese & Coelli, 1988). Identification rests on skewness of the composed residual, and when skewness is weak, estimated inefficiency is small and imprecise (Waldman, 1982; Simar & Wilson, 2010). Greene (2005) showed that in panel settings unobserved heterogeneity is easily mistaken for inefficiency. Kumbhakar and Tsionas (2006) develop input-oriented efficiency in production frontiers, the measure used here.

The closest recent application in education is Horrace, Rothbart and Yang (2022), who estimate a true fixed effects stochastic frontier for New York City public middle schools from 2014 to 2016. Their panel model separates persistent from transient inefficiency in mathematics and English language arts test score gains, and they use multivariate inference to identify subsets of efficient schools, which they propose as decision rules for allocating resources and incentives. They find about 58 percent of schools efficient in mathematics gains and 16 percent in English language arts. Their setting, schools within a single district sharing assessments and accounting rules, avoids the cost and comparability problems that dominate national district-level work, and those are the problems the sequential approach here is designed for. Their persistent–transient decomposition and their inference for efficient subsets are natural extensions of the three-year averages and band classifications used here.

Two lessons follow for education. First, whatever the estimator, efficiency is a residual, and a residual is only as clean as the model of cost that precedes it. Second, in national work, error common to a state, in linked assessments or in reported spending, will appear as state-level differences in efficiency (Reardon, Kalogrides, & Ho, 2021). The approach here addresses the first by estimating and validating cost before measuring efficiency, and the second by estimating the efficiency

frontier with and without state effects.

3. Step one: the National Education Cost Model

3.1 Method

The NECM is a district-level education cost function estimated by two-stage least squares on a national panel of district-years (Baker, Weber, & Srikanth, 2021; Baker, Di Carlo, & Weber, 2024; Baker, 2025b). It follows the cost function tradition of Duncombe and Yinger (2005, 2011); its theoretical basis is set out in Baker (2025a). The model relates log current spending per pupil E for district i in year t to outcomes, input prices, student needs, grade composition, scale, density and efficiency controls:

$$\ln E_{it} = \theta_0 + \theta_1 O_{it} + \theta_2 W_{it} + N_{it}'\delta + G_{it}'\lambda + S_{it}'\kappa + \theta_3 \ln D_{it} + Z_{it}'\phi + \tau t + e_{it}$$

The terms are:

- O , the outcome index, in standard deviations relative to the national mean, from Stanford Education Data Archive (SEDA) nationally linked reading and mathematics assessments (Reardon et al., 2024; Fahle et al., 2024).
- W , an extended Education Comparable Wage Index for regional labor costs (Taylor & Fowler, 2006).
- N , student need: the Census child poverty rate, the share of students with disabilities (centered on the state mean), the English learner share and, in some specifications, the Black enrollment share and its interaction with population density.
- G , grade composition: the shares of enrollment in pre-kindergarten and in grades 9–12.
- S , scale: indicators for enrollment under 100, 101–300, 301–600, 601–1,200, 1,201–1,500 and 1,501–2,000, against districts over 2,000.
- D , population density.
- Z , efficiency controls: the share of the population aged 5–17 (public monitoring), the ratio

of median housing value to the national median (fiscal capacity) and a Herfindahl index of enrollment concentration (competition).

- t , a linear time trend.

Outcomes are treated as endogenous because districts choose spending and outcomes jointly. O is instrumented with characteristics of surrounding districts: their income-to-poverty ratio and their racial composition. When the Black share enters the cost model, the Hispanic share of surrounding districts serves as the racial composition instrument; when it does not, the combined Black and Hispanic share does. Standard errors are robust and clustered by district.

3.2 From cost function to predicted cost

Predicted cost C is the spending the estimated model implies for a district to reach a common outcome target, with three adjustments. The outcome index is set to the target, here the enrollment-weighted national mean for 2013–2015 (a second target, the Massachusetts average, is also produced). The efficiency controls Z are held at their enrollment-weighted national means, so that spending associated with high fiscal capacity, weak monitoring or limited competition is not counted as cost. Each district's need, grade, scale, density and wage measures are set to their actual values.

Two gap measures follow. The **funding gap** compares actual spending S with predicted cost C , as a ratio (S/C) or a difference ($S - C$). The **outcome gap** compares actual outcomes with the target ($O - O^*$). A district with a funding gap of zero spends what the model says it needs to reach national-average outcomes.

For policy use, the NECM also translates predicted cost into a simplified formula. Predicted cost is regressed on the cost factors with an exponential mean (Poisson pseudo-likelihood), yielding multiplicative weights for poverty, disability, English learners, grade composition and scale, analogous to the weights in a state aid formula. The dollar funding gaps used in the validity tests in Section 5.2 are computed against this simulated cost.

3.3 Four alternative specifications

Baker (2025b) asks whether, and how, the racial composition of districts should enter the cost model. The question matters for efficiency analysis. If Black enrollment share is associated with higher costs, through concentrated poverty, segregation and the conditions that accompany it, but the model omits it, then the outcome shortfalls of predominantly Black districts will be read as inefficiency. Four specifications differ only in how need enters (Table 1).

Table 1. Alternative NECM specifications.

Specification	Poverty	Black share	Interaction	Racial composition instrument
Race-neutral	Child poverty rate	—	—	Neighbors' % Black or Hispanic
Black	Child poverty rate	% Black	—	Neighbors' % Hispanic
Poverty × density	Child poverty rate	—	Poverty × ln density	Neighbors' % Black or Hispanic
Black × density	Child poverty rate	% Black	% Black × ln density	Neighbors' % Hispanic

Notes: All include the outcome index (instrumented), labor cost index, disability and English learner shares, grade shares, enrollment size categories, ln density, efficiency controls and a time trend. A parallel set substitutes a composite socioeconomic status index for the child poverty rate.

The density interactions test whether the cost associated with poverty or with Black enrollment differs between dense urban and sparse rural settings. The poverty × density model asks whether a race-neutral urban poverty concentration measure can stand in for racial composition.

3.4 Selection criteria

The specifications are compared on four criteria (Baker, 2025b):

1. Instrument strength and validity: the Kleibergen–Paap F statistic for excluded instruments and Hansen's J test of overidentifying restrictions.
2. Out-of-sample prediction: models are fit to a random 80 percent of district-years and used to predict spending in the remaining 20 percent, evaluated by the correlation of predicted with

actual spending and by mean percentage error.

3. Validity of the gap relationship: the share of variance in outcome gaps explained by funding gaps, with year fixed effects and enrollment weights.
4. Residual bias: whether the residuals of the gap relationship are systematically related to Black enrollment share, Hispanic share or poverty.

The third and fourth criteria are the bridge to efficiency analysis. A cost model that prices need well should produce funding gaps that predict outcome gaps strongly, with residuals unrelated to the demographic characteristics the model is meant to price. Section 4 shows why.

3.5 Identifying assumptions

The cost model's key parameter is θ_1 , the cost of a higher outcome, and everything downstream depends on it: predicted cost, the funding gap and the slope against which efficiency is judged. Four assumptions identify it.

Instrument relevance. Outcomes are instrumented with two characteristics of surrounding districts: their income-to-poverty ratio and their racial composition. Both predict own outcomes conditional on the other regressors. In the Black $\times$ density model, the first-stage coefficient on neighbors' income-to-poverty ratio has $t = 19.7$ and that on neighbors' Hispanic share $t = -3.3$. The Kleibergen–Paap rank statistic rejects underidentification ($LM = 390$, $p < 0.001$) and the cluster-robust first-stage F is 204, well above conventional weak-instrument thresholds (Kleibergen & Paap, 2006). The Anderson–Rubin test, which is robust to weak instruments, rejects a zero outcome coefficient ($F = 291$, $p < 0.001$; Anderson & Rubin, 1949).

Instrument exclusion. The exclusion restriction is that, conditional on a district's own need, wages, scale, density and efficiency controls, the characteristics of its neighbors affect what it achieves but not what it costs the district to achieve it. The rationale is sorting: families and teachers sort across regional housing and labor markets, so neighbors' income and composition carry information about the conditions under which a district produces outcomes. Comparison-district instruments of this kind are standard in the cost function literature (Duncombe & Yinger, 2005).

The restriction fails if neighbors' characteristics proxy for regional cost conditions the wage index misses, such as amenities that affect teacher recruitment, or for competitive pressure the Herfindahl index misses. Either would put neighbor composition into the cost equation directly. Hansen's J test is consistent with a violation by at least one instrument in three of the four updated specifications (Section 5.1; Hansen, 1982). The test cannot say which instrument, or whether the disagreement is large enough to matter for predicted cost. The out-of-sample, gap and residual bias tests in Section 3.4 are the external checks on whether the resulting θ_1 produces useful cost predictions.

Measurement error in index variables. Several regressors are themselves estimates: the outcome index (linked test scores), the comparable wage index, the housing value ratio and the Herfindahl index. Error in the outcome index is the most consequential, because it would attenuate θ_1 . Instrumenting outcomes corrects for classical error in the index, provided the instruments are uncorrelated with that error. Linking error shared by all districts in a state is the exception: it is common to a district and its neighbors and is not removed by instrumenting, which is one reason the efficiency step is estimated with state intercepts. Error in the other indices biases their own coefficients and, through correlation, others, but it enters predicted cost the same way it entered estimation. The preferred models use a single, directly measured poverty rate rather than a composite socioeconomic index, partly to avoid compounding errors across components; the composite-index versions are reported only as sensitivity checks.

Specification. Two omissions deserve note. The share of children in single-parent households, available from American Community Survey tabulations for school districts, is not in the need vector. It is strongly correlated with poverty and Black enrollment share, so part of its cost is likely carried by those variables, but its omission could contribute to the residual poverty bias reported in Section 5.2. No district-level measure of teacher unionization or collective bargaining is available nationally. Evidence on unions and efficiency is mixed: Hoxby (1996) finds that unionization raises spending and inputs while lowering their productivity, Lovenheim (2009) finds little effect of union certification on pay, spending or dropout rates in three Midwestern states, and Strunk (2011) finds that more restrictive contracts in California are associated with lower student performance.

State bargaining law, which varies mainly between states, is absorbed by the state intercepts in the efficiency step; within-state variation in contract restrictiveness is not, and passes into estimated inefficiency. Finally, the cost model uses a linear time trend, while the efficiency frontier uses year fixed effects. A linear trend will misfit years with unusual national spending shocks, notably the federal pandemic relief spending of 2021–2024, and year effects are the more flexible alternative.

4. Step two: from the cost function to cost-adjusted efficiency

4.1 What the gap residual contains

The residual of the funding gap–outcome gap relationship is the cost function’s error, translated into outcome units. Write the cost model compactly, with X collecting the priced cost factors (W , N , G , S , D and the trend):

$$\ln S_{it} = \theta_0 + \theta_1 O_{it} + X_{it}'\beta + Z_{it}'\phi + e_{it}$$

Predicted cost sets outcomes to the target O^* and the efficiency controls to their national means:

$$\ln C_{it} = \theta_0 + \theta_1 O^* + X_{it}'\beta + \bar{Z}'\phi$$

Subtracting gives the log funding gap, and rearranging gives the outcome gap:

$$O_{it} - O^* = \frac{1}{\theta_1} \ln \left(\frac{S_{it}}{C_{it}} \right) - \frac{1}{\theta_1} [(Z_{it} - \bar{Z})'\phi + e_{it}]$$

Three implications follow.

First, if every district were equally efficient and cost were priced correctly, outcome gaps would line up along a single line with slope $1/\theta_1$. With the published NECM estimate of $\theta_1 = 1.783$, the implied slope is 0.56 SD per log point of spending relative to cost; the updated estimate reported in Section 5 ($\theta_1 = 1.287$) implies 0.78. These are close to the slopes the frontier estimates directly (0.69 pooled, 0.91 within states), which is a consistency check on the two steps.

Second, deviations from that line come from two sources. One is the efficiency controls: a district with high fiscal capacity or limited competition tends to spend more than its priced costs require, and that spending is counted as resources, not cost. If it does not raise outcomes, the district falls below the line. The other is the cost function error e , the part of spending not explained by outcomes and priced costs. That error, and so the residual, combines three things that are hard to sort out:

1. **Measurement error.** Error in measured outcomes, from test linking and small samples, and in measured spending, from differences in how districts and states report expenditures. Some is district-specific and transient; some is shared by every district in a state.
2. **Omitted variables bias.** Costs the model does not price, or prices wrongly: needs left out of the model, a functional form that misses how costs rise at small scale or high concentrations of poverty, and cost conditions that vary by state. Omitted costs correlated with district characteristics produce systematic residuals.
3. **Real differences in efficiency.** Spending in excess of what the observed outcome requires, given correctly priced costs. This is what the analysis aims to measure.

A district whose outcomes fall below what its funding gap predicts is one that spends more than the cost model says its observed outcomes require. That is the definition of cost inefficiency, the third component, but only once the first two are set aside. The residual is informative about efficiency because it inherits the cost function's error, and ambiguous for the same reason.

Third, the second component is only as small as the cost model is good. Any cost factor correlated with a district characteristic and omitted from the model will produce residuals correlated with that characteristic, and those residuals will look like inefficiency. This is why residual bias by racial composition (criterion 4 in Section 3.4) is a model selection test and not merely a fairness diagnostic. A race-neutral model that leaves the costs associated with segregation unpriced will systematically label predominantly Black districts inefficient.

Why not use e directly? The cost function residual is defined relative to instrumented outcomes and is in spending units; its one-sided component, in a direct cost frontier, must be identified jointly

with the instrumented outcome coefficient and every cost factor. Separating the steps lets the cost model be judged on whether it prices cost well, against criteria external to efficiency, and lets the efficiency model be estimated where the decomposition is simpler and the result is easier to interpret: in outcome space, conditional on a validated measure of resources relative to need.

4.2 Sorting out the three components

No single test separates measurement error, omitted variables bias and real inefficiency. Each leaves a different signature, though, and the design of the analysis is aimed at narrowing the first two so that more of what remains is the third (Table 2).

Table 2. Three components of the funding gap–outcome gap residual.

Component	Signature in the data	Evidence in this paper	How the analysis addresses it
1a. Measurement error common to a state	Whole states shifted above or below the line; attenuated pooled slopes	State intercepts reproduce pooled state efficiency rankings ($r = 0.96$) and absorb most between-state variance (19.5% to 1.1%); within-state slope steeper (0.91 vs 0.69)	State fixed effects frontier; state-standardized resource profiles
1b. District-level measurement error and noise	Symmetric, transient deviations	Noise large relative to inefficiency (σ_v 0.24 vs E[u] 0.12 SD)	Stochastic frontier noise term; three-year averages; bands rather than point estimates
2. Omitted variables bias	Residuals correlated with characteristics the cost model should price	Race-neutral residual penalty for Black enrollment (−0.49) removed by Black $\times$ density model (+0.08); residual poverty gradient (−1.10); least efficient districts smaller than state peers despite scale pricing; no state effects in NECM	Alternative cost specifications and residual bias tests; state intercepts absorb state-level omitted costs; within-state comparisons
3. Real differences in efficiency	Persistent shortfalls among districts with similar resources relative to cost, in the same state	Least efficient districts funded similarly relative to cost, with more staff of every kind and lower instruction shares than state peers	What remains after the above; never isolated cleanly

The components are not equally important everywhere. Between states, the first, and possibly the second, appear to drive much of the variation: the pattern of state intercepts is what one would

expect from cross-state differences in test linking and spending reports, and because the NECM omits state fixed effects by design, state-level cost conditions it does not price also pass into the funding gap. Within states, the first matters less, and the harder problem is telling the second from the third: whether a district that falls short is mispriced or genuinely inefficient.

4.3 The cost-adjusted production frontier

The efficiency step estimates a stochastic production frontier in which the input is spending relative to NECM predicted cost:

$$y_{it} = \alpha_t + \gamma_s + \beta \ln \left(\frac{S_{it}}{C_{it}} \right) + v_{it} - u_{it}, \quad v_{it} \sim N(0, \sigma_v^2), \quad u_{it} \geq 0$$

Here y is the outcome gap, α_t are year effects, v is symmetric noise and u is inefficiency, the shortfall of a district's outcome from the frontier. In the baseline model $\gamma_s = 0$. In the state fixed effects model γ_s is a full set of state intercepts, a “true fixed effects” frontier in the sense of Greene (2005) applied at the state level, so that measurement error common to a state is not attributed to its districts. The log form captures diminishing returns. Inefficiency is exponential, with half-normal as an alternative. Models are estimated by enrollment-weighted pseudo-maximum likelihood, with standard errors clustered by district.

The NECM used in this step is cost model 1 of the 2024 release, the Black $\times$ density specification that Baker (2025b) identifies as preferred. The outcome index for the frontier is limited to SEDA reading and mathematics in grades 6–8, so efficiency here means efficiency in producing middle-grades achievement in two subjects.

4.4 Input-oriented technical efficiency

Because y is measured in standard deviations and can be negative, the conventional output-oriented measure $\exp(-u)$ is not meaningful. Efficiency is instead expressed in input terms. Given the log frontier, a district with inefficiency u could reach its current outcome with resources relative to cost reduced by the factor $\exp(-u/\beta)$. Technical efficiency (TE) is the conditional expectation of that

factor given the composed residual $\varepsilon = v - u$:

$$TE_{it} = E \left[\exp \left(-\frac{u_{it}}{\beta} \right) \mid \varepsilon_{it} \right]$$

A TE of 0.85 means a frontier-efficient district would need 85 percent of the district's current resources, relative to cost, to achieve the same outcome. This returns efficiency to the cost framing: TE is a cost-efficiency ratio, estimated from outcome shortfalls rather than spending excesses. Annual estimates are noisy, so TE is averaged over trailing three-year windows ($t-2$ to t), beginning in 2011.

With a constant inefficiency distribution, u given ε is a monotone function of the composed residual. The frontier therefore ranks districts almost exactly as the standardized OLS residuals of the gap relationship would. What it adds is a scale with a resource interpretation and a metric that is comparable across years.

4.5 Efficiency bands and state-standardized profiles

District-years are grouped into four bands of three-year TE: most efficient (≥ 0.90 , the top decile), typical (0.80–0.90), less efficient (0.70–0.80) and least efficient (< 0.70). The 0.80 and 0.70 cuts lie at natural breaks identified by one-dimensional k-means clustering. Under state fixed effects, bands hold the same shares, with cut points of 0.939, 0.880 and 0.816.

To compare resource allocation across bands, Census F-33 functional spending (U.S. Census Bureau, n.d.) and NCES Common Core of Data staffing (National Center for Education Statistics, n.d.) are linked for fiscal years 2018–2024 (56,005 district-years). Staffing is measured as full-time-equivalent staff per 100 pupils in ten categories; spending as shares of current elementary–secondary spending in nine functions. Because low-efficiency districts are concentrated in a few states with distinctive staffing patterns, each measure x is also expressed relative to the district's own state and year:

$$z_{ist} = \frac{x_{ist} - \mu_{st}}{\sigma_{st}}$$

where μ and σ are the enrollment-weighted mean and standard deviation across linked districts in state s and year t (264 state-years with at least five districts). Band means of z are enrollment-weighted.

5. Results

5.1 Alternative cost models

All four specifications produce cost factors of the expected sign and similar magnitude (Table 3). Estimates are from an updated NECM panel of 176,443 district-years in 11,061 districts, using the current SEDA release; they differ in level from the published estimates in Baker (2025b), which used 147,186 district-years. The outcome coefficient ranges from 1.29 to 1.63: raising the outcome target by 0.1 SD raises predicted cost by about 13 to 16 percent. The child poverty coefficient is 1.66 to 1.71 in the models with racial composition or the poverty interaction, and 2.95 in the race-neutral model, which loads the costs associated with Black enrollment onto poverty. Black share raises cost in both specifications in which it enters, and more so in denser settings (Black $\times$ ln density, 0.099). Scale diseconomies are large and stable: districts under 100 pupils face costs 55 to 59 log points above districts over 2,000.

The Black $\times$ density model performs best on most selection criteria (Table 4). It has the strongest first stage (Kleibergen–Paap $F = 204$), the highest out-of-sample correlation with actual spending (0.48) and the smallest root mean squared error in sample. Its prediction errors are nearly unrelated to Black enrollment share in the holdout sample ($r = -0.01$), whereas the race-neutral and poverty $\times$ density models under-predict spending in districts with larger Black shares ($r = -0.25$ and -0.21). It does less well on the overidentification test. In the updated estimates, Hansen’s J rejects at the 5 percent level in three of four specifications, including Black $\times$ density ($p = 0.002$); only the poverty $\times$ density model passes ($p = 0.56$). In the published estimates the Black $\times$ density model passed comfortably ($p = 0.795$; Baker, 2025b). With two excluded instruments, the J test asks whether they imply the same outcome coefficient, and with 176,000 observations small differences are detectable. The result is a reason to treat the precise level of θ_1 with caution, not to prefer a model that fails the

Table 3. Alternative NECM cost function estimates.

Variable	Race-neutral	Black	Poverty × density	Black × density
Outcome index (instrumented)	1.596 (0.108)	1.345 (0.084)	1.632 (0.115)	1.287 (0.078)
Labor cost index	0.549 (0.028)	0.494 (0.025)	0.621 (0.029)	0.515 (0.024)
Child poverty rate	2.949 (0.214)	1.713 (0.131)	1.705 (0.173)	1.661 (0.124)
Poverty × ln density			0.294 (0.040)	
% Black		0.896 (0.052)		0.341 (0.056)
% Black × ln density				0.099 (0.010)
% Disability (state-centered)	2.663 (0.143)	2.575 (0.122)	2.682 (0.147)	2.503 (0.116)
% English learners	1.270 (0.105)	1.351 (0.098)	1.227 (0.108)	1.269 (0.091)
% Pre-K	0.479 (0.135)	0.268 (0.115)	0.406 (0.135)	0.227 (0.111)
% Grades 9–12	0.686 (0.048)	0.636 (0.041)	0.696 (0.049)	0.616 (0.039)
Enrollment < 100	0.550 (0.066)	0.593 (0.062)	0.569 (0.066)	0.581 (0.061)
Enrollment 101–300	0.281 (0.019)	0.298 (0.017)	0.281 (0.019)	0.289 (0.017)
Enrollment 301–600	0.194 (0.014)	0.211 (0.013)	0.194 (0.014)	0.205 (0.012)
Enrollment 601–1,200	0.100 (0.011)	0.120 (0.010)	0.097 (0.011)	0.120 (0.010)
Enrollment 1,201–1,500	0.053 (0.013)	0.075 (0.011)	0.054 (0.013)	0.079 (0.011)
Enrollment 1,501–2,000	0.051 (0.012)	0.070 (0.010)	0.054 (0.012)	0.075 (0.010)
ln population density	−0.025 (0.005)	−0.046 (0.005)	−0.078 (0.010)	−0.053 (0.005)
% Population aged 5–17	−0.088 (0.089)	−0.587 (0.105)	−0.082 (0.092)	−0.574 (0.102)
Housing value ratio	−0.443 (0.035)	−0.353 (0.027)	−0.422 (0.034)	−0.325 (0.025)
Herfindahl index	−0.129 (0.512)	−1.080 (0.445)	−0.549 (0.514)	−0.834 (0.432)
Time trend	0.034 (0.002)	0.030 (0.002)	0.032 (0.002)	0.028 (0.002)
Constant	8.569 (0.062)	8.807 (0.060)	8.672 (0.067)	8.789 (0.057)

Notes: Dependent variable: ln current spending per pupil. Two-stage least squares; robust standard errors clustered by district in parentheses. N = 176,443 district-years, 11,061 districts. All coefficients significant at $p < 0.01$ except % pre-K in the Black and Black × density models ($p < 0.05$), the Herfindahl index ($p < 0.05$ Black; $p < 0.10$ Black × density; not significant otherwise) and % aged 5–17 in the race-neutral and poverty × density models (not significant).

other criteria.

Table 4. Model selection diagnostics, updated estimates.

Criterion	Race-neutral	Black	Poverty × density	Black × density
Kleibergen–Paap F, excluded instruments	145.3	186.6	144.1	204.1
Hansen J, p-value	0.016	0.012	0.559	0.002
Root MSE (ln spending)	0.464	0.396	0.469	0.384
Out-of-sample r, predicted vs actual spending	0.385	0.450	0.384	0.481
Out-of-sample r, prediction error vs % Black	−0.246	−0.063	−0.212	−0.011
Funding gap → outcome gap R ²	0.408		0.389	0.437
Residual coefficient on % Black	−0.489		−0.434	+0.078
Residual coefficient on % Hispanic	−0.091		−0.090	−0.133
Residual coefficient on child poverty	−0.652		−0.713	−1.103

Notes: Out-of-sample statistics from models fit to a random 80 percent of district-years (141,113) and evaluated on the remaining 20 percent; correlations enrollment-weighted; prediction error = (predicted – actual) ÷ actual. Gap relationship: enrollment-weighted regression of the outcome index on the dollar funding gap (spending minus simulated cost) with year fixed effects, about 178,800 district-years. Residual coefficients from enrollment-weighted regressions of those residuals on % Black, % Hispanic and child poverty with year fixed effects; all $p < 0.01$. Gap and residual tests were not run for the Black-only specification in this estimation. A parallel set of models using a composite socioeconomic status index in place of child poverty yields outcome coefficients of 1.46 to 1.56 and the same ordering on instrument strength.

5.2 The gap relationship and residual bias

Funding gaps explain 39 to 44 percent of the variance in outcome gaps, and the Black × density model explains the most. Under this model, a \$1,000 increase in spending above simulated cost is associated with an outcome gap 0.041 SD higher; the race-neutral and poverty × density models imply 0.037 and 0.034 SD.

The residual bias tests target the second component of the residual, omitted variables bias, and show what goes wrong when cost is mispriced. Under the race-neutral model, holding poverty and Hispanic share constant, a district with an entirely Black enrollment has a residual 0.49 SD below one with no Black students. Read as efficiency, that would say predominantly Black districts get far less from their resources than other districts. Under the poverty × density model, which tries to capture concentrated urban poverty without race, the penalty is nearly as large (−0.43).

Under the Black $\times$ density model it reverses sign and shrinks to +0.08. The apparent inefficiency of predominantly Black districts under race-neutral models is, for the most part, an unpriced cost.

The tests also show what remains. Under every specification, residuals fall with child poverty, and under the Black $\times$ density model the poverty coefficient is larger (−1.10) because poverty no longer proxies for Black enrollment. Here the second and third components are in direct conflict. High-poverty districts may be less efficient, or the model may underprice poverty-related costs. The data cannot separate them. Any efficiency analysis built on this cost model should be read with that caveat, and the comparisons in Section 6, which are made within states and describe resource use rather than poverty, are less exposed to it.

5.3 The cost-adjusted frontier

The frontier built on the preferred NECM is concave and precisely estimated (Table 5). In the baseline model the coefficient on log spending relative to cost is 0.691 (SE 0.015): a 10 percent increase in resources relative to cost raises frontier outcomes by 0.066 SD. With state fixed effects the coefficient rises to 0.906 (SE 0.017), or 0.086 SD per 10 percent. The steeper within-state slope is what one would expect if error common to a state, in linked assessments or reported spending, attenuates the pooled estimate.

Table 5. Stochastic frontier estimates.

	Baseline (year effects)	State fixed effects
β , $\ln(\text{spending} \div \text{cost})$	0.691 (0.015)	0.906 (0.017)
Frontier gain from +10% resources (SD)	0.066	0.086
σ_v	0.239	0.221
$E[u]$ (SD)	0.119	0.089
Skewness of OLS residuals (weighted)	−0.13	+0.04
Median three-year TE	0.859	0.914
Share of variance in three-year TE between states	19.5%	1.1%

Notes: Exponential inefficiency; year fixed effects; N = 128,713 district-years, 8,088 districts, 38 states, 2009–2024; enrollment-weighted pseudo-likelihood. Robust standard errors clustered by district. Source: Baker (2026).

Figure 1 shows the frontier and the basis on which districts are classified, using 2022–2024 averages for 7,921 districts. The frontier rises steeply among districts well below their cost targets

and flattens among those well above them. The dashed lines mark where the TE cut points of 0.90, 0.80 and 0.70 fall in outcome terms. Because TE is a monotone function of the composed residual, band membership is almost entirely a matter of vertical distance from the frontier, and the boundaries run parallel to it: fewer than 1 percent of districts fall on the wrong side of a boundary, the exceptions arising because TE is averaged over three annual estimates while the figure plots three-year average positions. Read horizontally, the same distance is the input-oriented shortfall. The arrow for one illustrative district shows the resources, relative to cost, at which a frontier district would reach the same outcome; the estimated TE is less extreme than this raw distance because it shrinks the residual for noise. Because noise is large, a district must lie well above the estimated frontier, about 0.13 SD, to be classified most efficient, and the least efficient boundary lies about 0.65 SD below it. Most efficient districts appear at every level of funding, including many well below their cost targets; least efficient districts are concentrated near or somewhat below their targets.

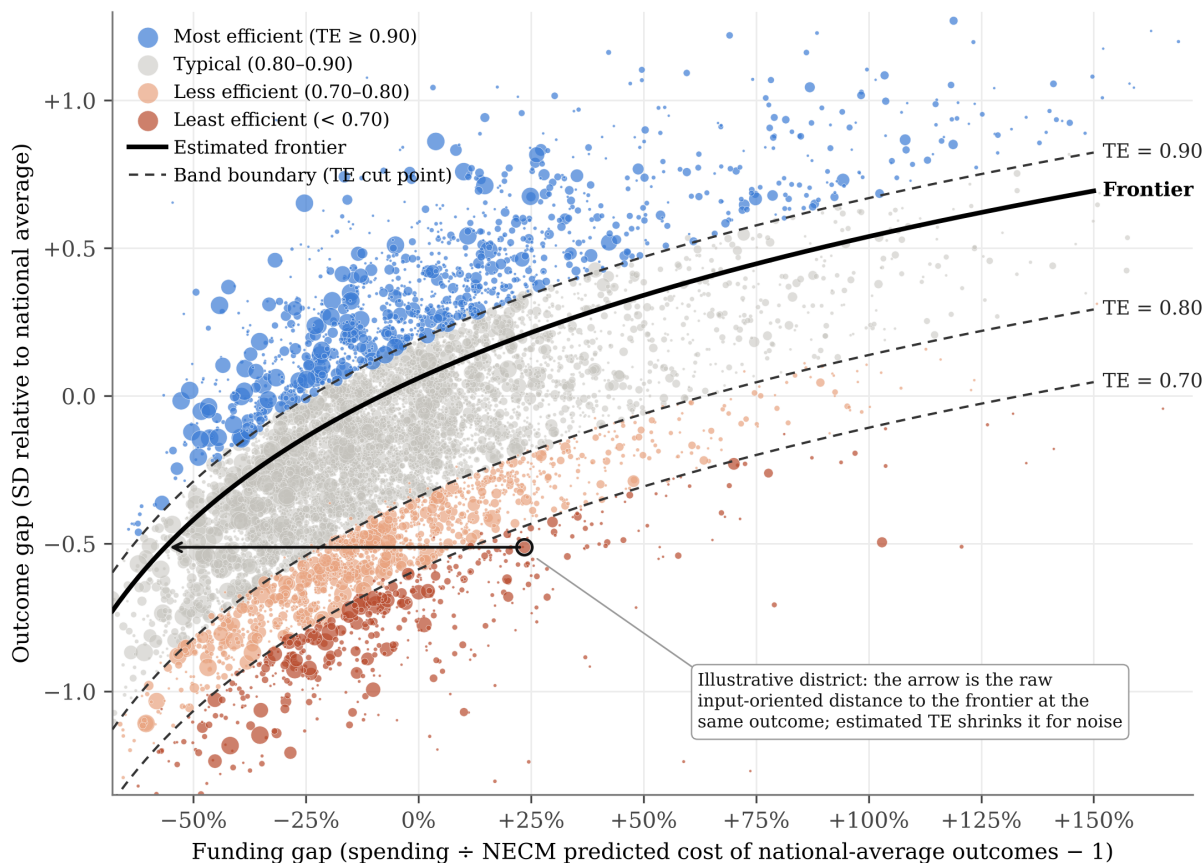
Three results bear on the sequential approach.

The frontier and the simple residual agree. The Spearman correlation between three-year TE and three-year standardized OLS residuals from the gap relationship is 0.997; 96 percent of district-years fall in the same band, and none moves between the top and bottom bands. Once cost is priced, the efficiency information is in the residual. The frontier adds an interpretable scale and a common metric across years, which reveals that the share of district-years in the two lowest bands doubled from about 11 percent in 2011 to 22 percent in 2024, within states as well as nationally.

District-level noise and measurement error, the first component, are large relative to inefficiency. σ_v is 0.24 against mean inefficiency of 0.12 SD in the baseline, so district TE estimates are shrunk toward the mean and band assignments are more reliable than point estimates.

State differences are mostly measurement. State intercepts in the fixed effects frontier range from +0.19 SD (North Carolina, Mississippi) to -0.45 SD (Rhode Island) and track baseline state mean TE almost exactly ($r = 0.96$). Including them reduces the between-state share of variance in TE from 19.5 to 1.1 percent, while district rankings within states are stable (rank correlation 0.88; no district-year moves between the most and least efficient bands).

Figure 1. Outcome frontier and efficiency bands, 2022–2024 averages, baseline model.



Notes: Each marker is a district ($n = 7,921$), sized by enrollment and colored by three-year technical efficiency as of 2024. Frontier: outcome gap = $0.06 + 0.691 \times \ln(\text{spending} \div \text{predicted cost})$, with 2022–2024 year effects averaged. Band boundaries are the outcome distances from the frontier that best separate adjacent TE bands (+0.13, -0.40 and -0.65 SD). Axes trimmed to show about 99 percent of districts. Source: Baker (2026) replication package.

In the terms of Section 4.2, between-state variation in pooled efficiency appears to be driven mainly by the first component, measurement error common to a state, and possibly by the second. Linked assessments and reported spending are least comparable across states. And because the NECM omits state fixed effects by design, state-level cost conditions it does not price, and state differences in what is counted as current spending, pass into the funding gap. The state fixed effects frontier removes both, but it also removes any real state-level differences in efficiency, which is why it brackets the answer rather than resolving it.

5.4 Efficiency is not productivity

Cost-adjusted efficiency answers a narrower question than critics of school spending usually ask. Across the 38 states, mean TE is unrelated to mean outcomes ($r = 0.00$) and strongly negatively related to spending relative to cost ($r = -0.70$); mean outcomes rise with spending ($r = 0.42$). Florida, North Carolina, Georgia and Texas rank among the most efficient states while spending 67 to 78 percent of NECM predicted cost, with outcomes 0.05 to 0.15 SD below the national average: they are efficient at producing relatively low outcomes. New Jersey spends about twice as much per pupil as Florida (\$19,124 against \$9,636), ranks 34th in efficiency and 4th in outcomes, 0.20 SD above Florida. Once state effects are included, New Jersey's mean TE (0.905) is essentially the same as Florida's (0.917). A finding that a district is efficient says nothing about whether its outcomes are adequate, and a finding that it is inefficient does not imply its outcomes could be sustained with less money.

6. Resource profiles of more and less efficient districts

6.1 Levels

The least efficient districts are not underfunded or overfunded relative to cost; they are higher-cost districts (Table 6). Under state fixed effects bands, per-pupil spending rises from \$12,682 in the most efficient band to \$16,701 in the least efficient, but NECM predicted cost rises in parallel, from \$13,641 to \$18,664. Median funding gaps are similar across bands (+1 to -8 percent). What

differs is outcomes: +0.38 SD in the most efficient band, −0.72 SD in the least. Instruction falls from 61.8 to 56.6 percent of current spending, general administration rises from 1.6 to 2.1 percent and operation and maintenance from 9.0 to 10.4 percent. Total staff rise from 12.6 to 15.2 per 100 pupils, and every staffing category rises with it.

Table 6. Profile of districts by efficiency band, state fixed effects model, FY2018–2024, enrollment-weighted.

	Most efficient	Typical	Less efficient	Least efficient	All
District-years	6,960	38,987	7,928	2,130	56,005
Mean three-year TE	0.946	0.916	0.858	0.772	0.909
Current spending per pupil	\$12,682	\$12,782	\$15,147	\$16,701	\$13,145
NECM predicted cost per pupil	\$13,641	\$15,934	\$17,712	\$18,664	\$15,932
Median funding gap	+1%	−8%	−7%	−5%	−7%
Outcome gap (SD)	+0.38	−0.12	−0.49	−0.72	−0.12
Instruction (% of current spending)	61.76	60.31	58.19	56.58	60.15
General administration (%)	1.62	1.73	1.88	2.09	1.75
School administration (%)	5.43	5.49	5.54	5.27	5.48
Operation & maintenance (%)	9.03	9.37	9.60	10.40	9.38
Classroom teachers per 100 pupils	6.13	6.38	6.38	6.74	6.36
Instructional aides per 100	1.53	1.59	1.84	2.06	1.62
School administrators per 100	0.35	0.38	0.41	0.45	0.38
LEA administrators per 100	0.12	0.15	0.16	0.20	0.15
LEA administrative support staff per 100	0.36	0.38	0.48	0.59	0.39
Total staff per 100	12.61	13.23	14.00	15.19	13.29

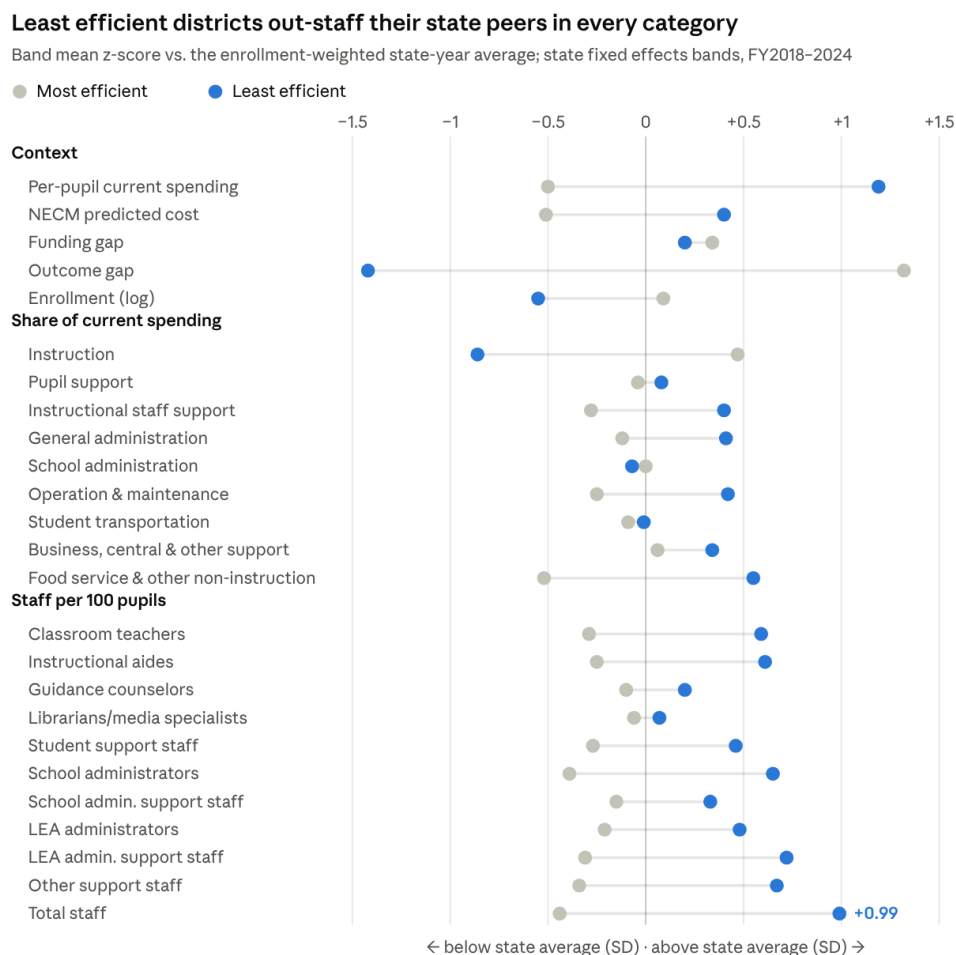
Notes: Bands hold the baseline shares (cut points 0.939, 0.880, 0.816). Spending shares from Census F-33; staffing from CCD FTE, zero counts treated as missing; measures winsorized at the 1st and 99th percentiles. Funding gap = spending ÷ predicted cost − 1.

6.2 Relative to state peers

National averages mix district and state differences. In the baseline model the least efficient districts are concentrated in New Jersey, Connecticut, Rhode Island, California and Illinois, whose staffing patterns differ from the national average for reasons unrelated to district efficiency. Figure 2 therefore expresses each measure in standard deviations from the district’s own state-year average.

Relative to their own states, the least efficient districts are more heavily staffed in every category. Under state fixed effects bands they are about 1 SD (+0.99) above their state average in total

Figure 2. Staffing and budget allocation relative to state averages, most and least efficient districts, state fixed effects bands.



Notes: Entries are enrollment-weighted band means of $z = (x - \mu_{st}) / \sigma_{st}$. Bands hold the baseline shares (cut points 0.939, 0.880, 0.816). Source: Baker (2026) replication package, Table A4; 56,005 linked district-years, 264 state-years, FY2018–2024.

staff per pupil, including classroom teachers (+0.59), instructional aides (+0.61), school administrators (+0.65) and LEA administrative support staff (+0.72). The most efficient districts are below their state average in every staffing category. Removing state effects from the frontier sharpens the contrast rather than creating it: under baseline bands, the least efficient districts are +0.56 SD in total staff and the most efficient –0.42 SD.

Budget shares move away from instruction (–0.86 SD), but not toward school administration (–0.07 SD), even though school administrator staffing is higher. The offsetting shares are general administration (+0.41), operation and maintenance (+0.42), instructional staff support (+0.40), business and central services (+0.34) and food service and other non-instruction (+0.55).

The pattern holds state by state. Comparing district-years with baseline TE of 0.90 or above against those below 0.80, in the 26 states with at least 20 district-years in each group, the less efficient group has more school administrators per pupil in all 26, a lower instruction share in 25, more classroom teachers per pupil in 23 and more LEA administrators in 19. Differences in administrative spending shares are less consistent: general administration is higher in 17 states and school administration in 13.

6.3 Scale, staffing and the three components

The resource profiles sit where the second and third components of the residual overlap.

Scale is a case of the second. The least efficient districts under state effects are smaller than their state peers (–0.55 SD in log enrollment; median enrollment 1,092 against 1,499 in the most efficient band). They spend well above their state average (+1.19 SD) and face somewhat higher predicted costs (+0.40 SD). The NECM prices scale with six enrollment categories, so part of small districts' higher spending and staffing is already counted as cost. That inefficiency still concentrates among smaller districts suggests the model does not fully capture diseconomies of small scale. The size categories may be too coarse, the cost of small size may depend on sparsity and isolation as well as enrollment, or scale may interact with other needs. To that extent, part of what the frontier labels inefficiency in small districts is an omitted cost.

The staffing differences may reflect the third component, real differences in how resources are used. The least efficient districts do not have a thinner classroom workforce; they have more of everything, with the largest excesses in central office and support staff, and a lower share of spending on instruction. Those are differences in allocation, not just in level. But staffing ratios rise mechanically as enrollment falls, and the state-standardized comparisons are not themselves cost-adjusted, so part of the staffing excess is the scale effect again. Separating the two requires comparing districts of similar size and cost within the same state.

In dollar terms the administrative difference is small. The general administration share in the least efficient band is about half a percentage point above the most efficient band, roughly \$80 per pupil. By the frontier estimate, \$80 to \$100 per pupil would be associated with an outcome difference well under 0.01 SD, against an outcome gap of more than 1 SD between the two bands. Policies that target administrative spending as the route to closing outcome gaps would address a small part of the difference.

7. Discussion and conclusions

7.1 What the sequential approach buys

The main argument of this paper is methodological. Efficiency in education is always a residual, and a residual is only as credible as the model of cost behind it. Estimating the cost function first, and judging it on whether it prices cost well, puts that model in the open where its assumptions can be tested. The alternative specifications in Section 5 show why that matters. Swapping a race-neutral cost model for the Black $\times$ density model changes the estimated residual penalty for an all-Black district from about -0.49 SD to $+0.08$ SD. A direct efficiency model built on the race-neutral specification, whether production or cost, would have reported that penalty as inefficiency.

Once cost is priced, the evaluation of efficiency is comparatively simple. The stochastic frontier and a ranking by standardized residuals of the gap relationship agree almost perfectly (rank correlation 0.997). The frontier's contributions are a scale with a resource interpretation, the share of current resources relative to cost a frontier district would need, and a metric that is comparable

across years. The heavy lifting is done in the cost model.

The approach also returns efficiency to a cost framing without asking the cost model to identify inefficiency. Input-oriented TE is a cost-efficiency ratio, but it is estimated from outcome shortfalls conditional on validated resources relative to need. And because the efficiency step is estimated separately, it can be estimated with and without state effects, bracketing the share of apparent inefficiency that is cross-state measurement error.

7.2 Comparison with alternative approaches

The sequential approach is one of several ways to study school efficiency, and its advantages are specific.

The main trade-off is between national coverage and causal or behavioral depth. Production function research has long emphasized how hard it is to recover the parameters of achievement production from observational data (Todd & Wolpin, 2003). Value-added and structural approaches answer questions this paper cannot, about the causal effect of particular schools or the response to policies not yet tried, but they do so for single districts or states with rich data. The sequential approach answers a narrower question for nearly every district in the country: given what it would cost this district to reach common outcomes, and what it actually has, how does it compare with districts like it? Its distinctive contributions are that the cost model can be tested before efficiency is measured, and that the efficiency results can be recomputed under alternative cost models to show how much of apparent inefficiency is a pricing choice.

7.3 What the estimates are for

The estimates are descriptive. They identify where outcomes fall short of what similar resources, relative to cost, achieve elsewhere, but they cannot say why, and they cannot recommend specific remedies such as expanding school choice or improving teacher retention. They have three uses.

1. **Screening.** Band classifications identify districts whose outcomes fall well short of those of comparably resourced peers in their own state, as candidates for closer study by state agencies or researchers, and districts that do well with little, as candidates for learning what they do

Table 7. Approaches to estimating school and district efficiency.

Approach	What it identifies	Strengths	Limits relative to this paper
Production frontier on raw spending or inputs (DEA, SFA)	Shortfall from best-practice outcomes given inputs and controls	Simple; no cost model needed	Unpriced need passes into inefficiency; rankings track student need
Direct stochastic cost frontier, including versions with instrumented outcomes and indirect efficiency controls (Taylor et al., 2018)	Spending above frontier cost for observed outcomes	Native cost-efficiency measure	Must combine endogenous outcomes with a one-sided error; mispriced cost passes into inefficiency unseen
Panel frontier with unit effects (e.g., Horrace et al., 2022)	Persistent and transient inefficiency	Separates stable from year-to-year shortfalls; formal inference on efficient sets	Unit effects absorb persistent inefficiency and persistent unpriced cost alike; within one district, comparability problems are small
School value-added, including lottery-based (Angrist et al., 2017)	Causal contribution of a school to achievement	Credible causal estimates of quality	Measures output, not output per dollar; needs admission lotteries or rich student data
Structural models of schools, families and teachers (e.g., Epple & Romano, 1998)	Behavioral parameters governing resource choices and sorting	Can simulate policies not yet adopted	Demanding data and modeling assumptions; rarely feasible at national district scale
Cost function, then cost-adjusted frontier (this paper)	Shortfall from best attainable outcome given resources relative to validated cost	Cost model tested on external criteria; interpretable resource-based scale; national coverage; state measurement error bracketed	Descriptive; inherits any cost model error; noise large relative to inefficiency

differently. Because noise is large relative to inefficiency, they should not be used to set funding or impose sanctions.

2. **Testing claims of waste.** Arguments against additional funding often assert that low-performing districts are wasteful. The approach provides a direct check. In these data, the least efficient districts are funded similarly relative to cost, their administrative spending excess is about \$80 to \$100 per pupil, and much apparent inefficiency under race-neutral cost models is unpriced cost. Claims of waste that ignore cost, or rest on national comparisons exposed to state measurement error, can be evaluated against these results.
3. **Separating adequacy from efficiency.** The frontier shows that efficient districts can produce low outcomes and inefficient districts high ones. Efficiency results therefore cannot be used to argue that outcomes are adequate or that resources could be cut without cost to students, and the cost model, not the frontier, remains the tool for setting funding targets.

7.4 Limits

Several limits apply.

- **Inefficiency remains a residual.** The residual combines measurement error, omitted variables bias and real differences in efficiency, and the frontier attributes whatever is not classified as noise to inefficiency. Any shortfall not explained by resources relative to NECM cost, including unpriced needs and underpriced scale, is counted as inefficiency. The residual bias tests show that the preferred model removes most of the bias associated with Black enrollment but leaves residuals that fall with child poverty. Whether that reflects lower efficiency in high-poverty districts or underpriced poverty costs cannot be resolved here. The omission of single-parent household share from the need vector may contribute to it.
- **The cost model's outcome coefficient is uncertain.** The updated estimates reject the overidentifying restrictions for most specifications, and the outcome coefficient differs between the published (1.78) and updated (1.29) Black $\times$ density estimates. The level of predicted cost, and so the level of funding gaps, depends on it. Efficiency rankings, which depend on

relative position, are less sensitive. Planned robustness checks include just-identified estimates with each instrument separately, limited-information maximum likelihood, and year effects in place of the linear trend.

- **Efficiency controls shape the result.** Holding fiscal capacity, monitoring and competition at national means removes their association with spending from cost, not from spending. High-capacity districts that spend above predicted cost are credited with positive funding gaps, and outcome shortfalls among them count toward inefficiency, which is the intended treatment but a choice nonetheless.
- **Noise is large.** With σ_v about twice mean inefficiency, district estimates are imprecise; bands are more reliable than point estimates.
- **Outcomes are narrow.** The frontier measures efficiency in producing grade 6–8 reading and mathematics achievement, while the cost model is estimated on a grades 3–8 index. Districts also spend on early learning, high school completion and non-academic outcomes.
- **Cross-state comparability is limited.** Eleven states with carried-forward outcome data are excluded, and state-level differences in testing and financial reporting pass into national comparisons. The state fixed effects results suggest most between-state variation in efficiency is of this kind.
- **The staffing comparisons are descriptive.** They are made within states but are not cost-adjusted, so part of the excess staffing in small districts reflects scale costs the NECM already prices.

7.5 Conclusions

Production models of school efficiency have generally failed because they cannot fully account for cost, and direct cost frontiers ask too much of one model. A sequential approach, estimating and validating the cost function first and then evaluating efficiency in a cost-adjusted production framing, separates the two problems. The residual of the funding gap–outcome gap relationship carries information about relative efficiency because it is the outcome-space image of the cost function’s

error. It carries useful information only when the cost model prices need correctly, and the residual bias tests provide a direct check. Even then, the residual combines measurement error, omitted variables bias and real differences in efficiency. The design choices here (state intercepts, three-year averages, alternative cost models and within-state comparisons) are ways of narrowing the first two so that more of what remains is the third.

Applied nationally, the approach yields three substantive findings. Most of the between-state variation in efficiency is absorbed by state frontier intercepts and is plausibly measurement error, possibly compounded by state-level costs the national model does not price. The dispersion of efficiency within states has roughly doubled since 2011. And the least efficient districts are not those with more money than they need: they are smaller, higher-cost districts, funded similarly relative to cost, with more staff of every kind than their state peers. Their small size points partly to scale costs the model underprices; their staffing may reflect real inefficiency, but is partly the same scale effect. Their administrative spending excesses are too small to account for their outcome gaps.

The limits above argue for using cost-adjusted efficiency estimates to identify districts for closer study, not to set funding or impose sanctions. Natural extensions include panel frontiers separating persistent from transient inefficiency (Colombi, Kumbhakar, Martini, & Vittadini, 2014), models of inefficiency as a function of scale, sparsity and governance, staffing comparisons among districts of similar cost and scale within states, and further work on the pricing of concentrated poverty in the cost model itself.

Data availability statement

The National Education Cost Model estimates and School Finance Indicators Database are publicly available at <https://www.schoolfinancedata.org> (Baker, Di Carlo, & Weber, 2024). Outcome data are from the Stanford Education Data Archive (Reardon et al., 2024), spending data from the U.S. Census Bureau Annual Survey of School System Finances and staffing data from the NCES Common Core of Data, all publicly available. Replication code is available from the author on request.

References

- Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), 21–37.
- Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20(1), 46–63.
- Angrist, J. D., Hull, P. D., Pathak, P. A., & Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics*, 132(2), 871–919.
- Baker, B. D. (2025a). *Framework for evaluating & reforming education finance systems* (EdWorkingPaper No. 25-1127). Annenberg Institute at Brown University. <https://doi.org/10.26300/pa0r-n548>
- Baker, B. D. (2025b). How and why racial isolation affects education costs and the provision of equal educational opportunity. *Education Policy Analysis Archives*, 33(55). <https://doi.org/10.14507/epaa.33.8843>
- Baker, B. D. (2026). *Technical efficiency in producing educational outcomes: A stochastic frontier analysis using the National Education Cost Model* [Working paper]. University of Miami.
- Baker, B. D., Di Carlo, M., & Weber, M. (2024). *School Finance Indicators Database and National Education Cost Model* [Data set]. Albert Shanker Institute. <https://www.schoolfinancedata.org>
- Baker, B. D., Weber, M., & Srikanth, A. (2021). Informing federal school finance policy with empirical evidence. *Journal of Education Finance*, 47(1), 1–25.
- Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, 38(3), 387–399.
- Chakraborty, K., Biswas, B., & Lewis, W. C. (2001). Measurement of technical efficiency in public education: A stochastic and nonstochastic production function approach. *Southern Economic Journal*, 67(4), 889–905.
- Colombi, R., Kumbhakar, S. C., Martini, G., & Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis*, 42(2), 123–136.
- Costrell, R., Hanushek, E., & Loeb, S. (2008). What do cost functions tell us about the cost of an adequate education? *Peabody Journal of Education*, 83(2), 198–223.
- De Witte, K., & López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, 68(4), 339–363.
- Duncombe, W., Ruggiero, J., & Yinger, J. (1996). Alternative approaches to measuring the cost of education. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 327–356). Brookings Institution.
- Duncombe, W., & Yinger, J. (2005). How much more does a disadvantaged student cost? *Economics of Education Review*, 24(5), 513–532.

- Duncombe, W., & Yinger, J. (2011). Are education cost functions ready for prime time? An examination of their validity and reliability. *Peabody Journal of Education*, 86(1), 28–57.
- Epple, D., & Romano, R. E. (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 88(1), 33–62.
- Fahle, E. M., saliba, j., Kalogrides, D., Shear, B. R., Reardon, S. F., & Ho, A. D. (2024). *Stanford Education Data Archive: Technical documentation* (Version 5.0). <https://purl.stanford.edu/cs829jn7849>
- Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2), 269–303.
- Gronberg, T. J., Jansen, D. W., Karakaplan, M. U., & Taylor, L. L. (2015). School district consolidation: Market concentration and the scale-efficiency tradeoff. *Southern Economic Journal*, 82(2), 580–597.
- Gronberg, T. J., Jansen, D. W., & Taylor, L. L. (2011). The adequacy of educational cost functions: Lessons from Texas. *Peabody Journal of Education*, 86(1), 3–27.
- Gronberg, T. J., Jansen, D. W., & Taylor, L. L. (2017). Are charters the best alternative? A cost frontier analysis of alternative education campuses in Texas. *Southern Economic Journal*, 83(3), 721–743.
- Grosskopf, S., Hayes, K. J., Taylor, L. L., & Weber, W. L. (1997). Budget-constrained frontier measures of fiscal equality and efficiency in schooling. *Review of Economics and Statistics*, 79(1), 116–124.
- Grosskopf, S., Hayes, K. J., Taylor, L. L., & Weber, W. L. (2001). On the determinants of school district efficiency: Competition and monitoring. *Journal of Urban Economics*, 49(3), 453–478.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029–1054.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141–1177.
- Horrace, W. C., Rothbart, M. W., & Yang, Y. (2022). Technical efficiency of public middle schools in New York City. *Economics of Education Review*, 86, Article 102216.
- Hoxby, C. M. (1996). How teachers’ unions affect education production. *Quarterly Journal of Economics*, 111(3), 671–718.
- Imazeki, J., & Reschovsky, A. (2004). Is No Child Left Behind an un- (or under-) funded federal mandate? Evidence from Texas. *National Tax Journal*, 57(3), 571–588.
- Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *Quarterly Journal of Economics*, 131(1), 157–218.
- Jackson, C. K., & Mackevicius, C. L. (2024). What impacts can we expect from school spending policy? Evidence from evaluations in the United States. *American Economic Journal: Applied Economics*, 16(1), 412–446.

- Jondrow, J., Lovell, C. A. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2–3), 233–238.
- Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1), 97–126.
- Kumbhakar, S. C., & Tsionas, E. G. (2006). Estimation of stochastic frontier production functions with input-oriented technical efficiency. *Journal of Econometrics*, 133(1), 71–96.
- Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, 10(2), 1–26.
- Lovenheim, M. F. (2009). The effect of teachers’ unions on education production: Evidence from union election certifications in three Midwestern states. *Journal of Labor Economics*, 27(4), 525–587.
- Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2), 435–444.
- National Center for Education Statistics. (n.d.). *Common Core of Data: Local education agency universe and staff data, 2017–18 through 2023–24* [Data set]. U.S. Department of Education. <https://nces.ed.gov/ccd/>
- Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., & saliba, j. (2024). *Stanford Education Data Archive* (Version 5.0) [Data set]. <https://purl.stanford.edu/cs829jn7849>
- Reardon, S. F., Kalogrides, D., & Ho, A. D. (2021). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Journal of Educational and Behavioral Statistics*, 46(2), 138–167.
- Ruggiero, J. (1996). Efficiency of educational production: An analysis of New York school districts. *Review of Economics and Statistics*, 78(3), 499–509.
- Simar, L., & Wilson, P. W. (2010). Inferences from cross-sectional, stochastic frontier models. *Econometric Reviews*, 29(1), 62–98.
- Strunk, K. O. (2011). Are teachers’ unions really to blame? Collective bargaining agreements and their relationships with district resource allocation and student performance in California. *Education Finance and Policy*, 6(3), 354–398.
- Taylor, L. L., & Fowler, W. J., Jr. (2006). *A comparable wage approach to geographic cost adjustment* (NCES 2006-321). U.S. Department of Education, National Center for Education Statistics.
- Taylor, L. L., Willis, J., Berg-Jacobson, A., Jaquet, K., & Caparas, R. (2018). *Estimating the costs associated with reaching student achievement expectations for Kansas public education students: A cost function approach*. WestEd. Prepared for the Kansas Legislature.
- Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113(485), F3–F33.

- U.S. Census Bureau. (n.d.). *Annual Survey of School System Finances, fiscal years 2009–2024* [Data set]. <https://www.census.gov/programs-surveys/school-finances.html>
- Waldman, D. M. (1982). A stationary point for the stochastic frontier likelihood. *Journal of Econometrics*, 18(2), 275–279.