The Adequacy of Educational Cost Functions: Lessons from Texas

By

Timothy J. Gronberg
Department of Economics
4228 TAMU
Texas A&M University
College Station, TX 77843-4228
tjg@econmail.tamu.edu
979-845-8849

Dennis W. Jansen
Department of Economics
4228 TAMU
Texas A&M University
College Station, TX 77843-4228
dennisjansen@tamu.edu
979-845-8849

and
Lori L. Taylor
The Bush School of Government and Public Service
4220 TAMU
Texas A&M University
College Station, TX 77843-4220
lltaylor@tamu.edu
979-458-3015

May 2008

The Adequacy of Educational Cost Functions: Lessons from Texas

Abstract

Adequacy studies based on cost functions have come under attack. A recent Texas court battle featured two cost functions, one professional judgment study and three widely divergent estimates of the cost of adequacy. At the low end of the scale, the state's expert estimated that it would cost an additional $68,000 to raise the 46 plaintiff districts to a reasonable standard of adequacy. The plaintiff's experts estimated that it would cost at least $457 million to meet the same standard of performance. Enough time has passed to look back and determine which one had the more accurate predicted cost of adequacy. We think the evidence points to a clear winner in the horse race between the two cost function analyses put forward in Texas, and we point to some of the differences in the three studies as a potential guide to researchers, policymakers and others interested in studying the cost of adequacy.

Policymakers, educators, and increasingly, litigators, have expressed a demand for estimates of the resources needed for the provision of an adequate education. This demand has led to studies using a variety of techniques to generate estimates of the resources needed for an adequate education. Two popular methods of generating these estimates are the professional judgment method and econometric cost functions. Recently these methods have been put to the test in Texas, in a lawsuit that confounded the issue of the school funding mechanism with the issue resources needed to achieve an adequate education.

In the spring of 2004, 46 school districts brought suit against the state of Texas, arguing that the Texas school finance formula—locally known as the Robin Hood plan—was unconstitutional. The plaintiff districts argued that the formula provided inadequate funding, that cost adjustments for student need were too low, and that as a result school districts were forced to raise their local property tax rates to the maximum allowed under the formula, thereby removing local discretion over property tax rates. The latter point was particularly salient because the Texas Constitution prohibits a statewide property tax, and decisions in previous school finance cases had specifically guaranteed local discretion over property tax rates.

The Texas District Court ruled in favor of the plaintiffs on all three counts, and ordered the State to revamp the school finance formula and greatly increase spending on education. The Texas Supreme Court overruled the District Court's conclusions with respect to adequacy, but upheld the decision with respect to the unconstitutionality of limits on local property tax rates. As a result, the litigation did not lead to a substantial increase in per-pupil spending. On average, real per-pupil spending fell by 0.8 percent between 2002 and 2004, and by another 0.5 percent between 2004 and 2007.[1]

Three studies of educational adequacy were used as evidence in the trial. The state of Texas presented a cost function analysis that Timothy Gronberg, Dennis Jansen, Lori Taylor and Kevin Booker had conducted at the request of the Texas legislature before the litigation was filed and that Lori Taylor extended for litigation purposes (hereafter referred to as GJTB).[2] The plaintiffs presented two studies—a cost function analysis conducted at their request by Jennifer Imazeki and Andrew Reschovsky (hereafter referred to as I&R) and a Professional Judgment analysis conducted at their request by Management Analysis and Planning Inc. (hereafter referred to as MAP).[3]

The three studies made very different predictions about the cost of achieving adequacy in the 46 Plaintiff districts. The GJTB study indicated that all but two of the plaintiff districts were already spending enough to achieve performance thresholds consistent with the state's definition of adequacy. Assuming that funding would not decrease from 2002 levels in any district—a hold-harmless assumption—their model predicted that it would cost in total an additional $861,000 to raise all of the plaintiff districts up to the threshold in 2004.[4] Using the same performance thresholds, the I&R study predicted that it would cost at least an additional $457 million. The MAP study

[1] As in the Texas studies under evaluation, current operating expenditures per pupil have been adjusted for inflation using the Employment Cost Index.
[2] See Gronberg, et al. (2004), Gronberg et al. (2005) and Taylor (2004).
[3] See Imazeki and Reschovsky (2004) and Smith et al. (2004), respectively.
[4] The GJTB study used data through the 2001-2002 school year. In her extension of the GJTB model, Taylor (2004) updated the GJTB estimates including data for 2002-03. Because spending increased in most plaintiff districts between 2002 and 2003, the estimate of the total needed to bring all plaintiff districts up to the performance standard, holding all other districts harmless, fell to $68,000. We focus on the original GJTB analysis to remove any differences in model predictions attributable to differences in the time frame for the cost projections.

predicted that providing an adequate education in the 46 plaintiff districts would cost an additional $683-$830 million.[5]

The differences between GJTB and I&R had implications beyond the 46 plaintiff districts, as all districts in Texas would presumably have received additional funding if the court had decided, in the end, for the plaintiffs. Further, the differences between the GJTB and I&R studies have made their way into recent academic articles. Imazeki and Reschovsky (2005) write that the GJTB study "reached the conclusion that 'in aggregate, the level of education funding in Texas is more than sufficient to meet performance goals consistent with the state's accountability system.'" They also write that the I&R study "concluded that, in aggregate, Texas school districts would need at least $2 billion in additional revenue to satisfy the requirements of the accountability system." While Imazeki and Reschovsky (2005) overstate the differences in the two studies, the general impression that there are large differences is correct, and leads Duncombe (2006) to include these two studies as an example of the (lack of) inter-rater reliability in cost of adequacy studies.

At the end of the day, however, the proof is in the pudding. Although it is impossible to quantify the performance standard in the MAP analysis, the GJTB and I&R studies predicted the additional spending needed so that 55 percent of the students would pass the Texas Assessment of Knowledge and Skills (TAKS) in mathematics and reading. ***All*** of the 46 plaintiff districts met this standard in 2004. As Figure 1 illustrates, most did so with fewer real resources than in 2002.

[5] These estimates adjust for geographic variations in cost using the same index as in the GJTB and I&R models. Using other geographic deflators, the MAP estimates are as low as $504 million and as high as $990 million.

As an example, consider Dallas Independent School District (DISD). The I&R model estimated that DISD required an additional $1,200 per pupil to achieve a 55 percent passing rate in 2004. This district alone accounted for 47 percent of the total increase in funding indicated by the I&R model. The GJTB estimated that DISD would need an additional $4 per pupil to meet the 55 percent passing rate, a small increase that, because of the size of DISD, accounted for 77 percent of the total funding increase indicated by the GJTB model. In 2004, 55 percent of DISD students passed the TAKS in math, 63 percent passed in reading/language arts, and all of the other outcomes included in either model also increased from their 2002 levels. This occurred with an increase in real per-pupil spending between 2002 and 2004 of $300 per pupil, not $1,200 per pupil.

This paper examines the three studies used in the Texas litigation, paying particular attention to the two cost function studies. We discuss strengths and weaknesses of each model, highlighting the differences in approach across the studies. We begin with a discussion of general concerns that have been raised regarding use of cost functions for the purpose of estimating the cost of providing an adequate education.

1. The Cost Function Approach

There are several reasons that cost functions are attractive to economists studying public school spending behavior. First, the cost function allows estimation when output prices are unavailable, perhaps because data is missing or because output prices are not determined in a competitive market. Second, the cost function more easily handles a multi-product firm than the alternative of directly estimating the technology via a production function. Third, the cost function allows a relatively straightforward calculation of alternative cost indices for policy analysis. Fourth, the cost function approach does not require an assumption of profit maximizing objectives for the decision makers under study. The ability to employ a cost function based approach without assuming profit maximization is especially useful for non-profit institutions such as public schools.

Faux Cost Functions?

In a series of recent papers (Costrell, Hanushek, and Loeb (2007), Hanushek (2006, 2007)), Rick Hanushek and his co-authors argue that the econometric cost analyses used in school finance litigation are not actually cost functions. Their argument is at a fundamental, conceptual level. In economics, a cost function has a precise meaning. The cost function represents the relationship between the minimum spending required to produce a given outcome within a given input environment. The input environment would include input prices, the quantities of any quasi-fixed input factors, and the best available technology for transforming inputs into outcomes. Thus a cost function assumes technical (maximum outcomes for the inputs used) and allocative (best mix of inputs for the input prices faced) efficiency is achieved by the relevant decision-makers. Given the potentially weak incentives for efficient behavior among school managers, and given their reading of existing empirical studies of school performance, Hanushek and his co-authors assert that schools are not cost-minimizing operations. Under the inefficiency hypothesis, estimates of the observed relationship between outlays and outcomes (conditioning on input conditions) are merely spending function estimates, not cost function estimates. Hanushek and company find attempts to introduce control variables for efficiency as additional regressors in the spending estimation to be abject failures. As they state “the ‘efficiency controls’ do little to explain the variations in spending, and are rarely convincing measures of the full range of efficiency” (Costrell et al. 2007, p.10). In the absence of an appropriate methodology for uncovering the true cost function from the faux cost (spending) function, their advice is to shut-down the cost function enterprise.

We disagree with this conclusion because we believe such a method exists—stochastic frontier estimation. The raison d’etre for the stochastic frontier methodology is the realization that actual decision-makers may not be efficient. This reality can hold in the for-profit private, the not-for-profit private, and the public sectors. Indeed, this method has been meaningfully applied to analyzing behavior in diverse settings, from banks to hospitals. The standard stochastic frontier model specifies

$$C = C\,(w_1, \ldots, w_k; z_1, \ldots, z_h; y \mid \beta\,)\cdot\exp(v + u),$$

where C is actual or observed cost/spending, $C(w_1, \ldots, w_k; z_1, \ldots, z_h; y \mid \beta)$ is the cost frontier, w_i are input prices, z_j are quasi-fixed inputs, y is outcome(s), β is the cost parameter vector to be estimated, v is a random noise component (representing an exogenous random shock, such as a rainy testing day), and u is a one-sided error term that captures inefficiency. This cost frontier is the true neo-classical function (or deterministic cost frontier), the object of discovery here. Inefficiency increases cost above minimum cost, so $u \geq 0$, and cost efficiency is defined as $\exp(-u) \leq 1$. In the GJTB models, u is interpreted as measuring the cost of technical inefficiency, i.e. failure to reach the best practice student achievement frontier.

The stochastic cost frontier model is, therefore, structured to answer the Hanushek group challenge for an appropriate conceptual framework for school cost analysis. The stochastic cost frontier model will, at least in principle, separate the variation in observed spending due to differences in "true" deterministic cost factors from variation due to inefficiency while also allowing for variation due to stochastic, random factors. We agree with Hanushek that the movement from principle to practice involved in school cost estimation faces a large number of econometric hurdles. Indeed, the stochastic cost frontier model brings its own set of additional hurdles to the track (eg. how to model the one-sided error). Our point is simply that the stochastic frontier estimation approach—which was used in the GJTB analysis of Texas—is conceptually sound.

2. The Texas Cost Function Analyses

Both the GJTB and I&R studies modeled K-12 school district expenditures as a function of the outcomes produced, the prices of inputs, and the environmental factors that influence the educational production process, such as student characteristics and school district size. Both used regression techniques and generated adequacy estimates by using their models to predict expenditures at designated outcome levels, holding constant the characteristics of districts. However, within this general framework, there were substantial differences between the two models.

Estimation Period

GJTB estimated a cost function based on four years of data—a time series of cross sections—covering the period from 1998-99 through 2001-02. I&R estimated a cross-sectional cost function using data from the 2001-02 school year.

Functional Form

An important difference in the two studies is in the functional forms. GJTB used a flexible functional form, a variant of the translog cost function[6] while I&R used a Cobb-Douglas specification. There are several reasons to prefer a translog. First, the Cobb-Douglas specification is a restricted, special case of the translog. (Technically, the translog model nests the Cobb Douglas specification.) If the true model is in fact a Cobb-Douglas model, then estimation of the translog yields, asymptotically, the same coefficients as from the Cobb-Douglas specification. If the true model is in fact a Cobb-Douglas, then estimating a Cobb-Douglas results in some improvement in the statistical efficiency of the coefficient estimates. On the other hand, if the Cobb-Douglas is in fact too restrictive a functional form, then estimating a Cobb-Douglas will lead to biased and inconsistent parameter estimates. At the very least it is incumbent on the analyst to test the Cobb-Douglas restrictions.

A translog, however, is not just a less restricted functional form compared to a Cobb-Douglas model. Instead, the translog cost function is a local second-order approximation to an arbitrary cost function. Thus, up to a second order, the translog can serve to approximate any number of possible cost function specifications.

Finally, there are theoretical reasons for avoiding a Cobb-Douglas cost function. The Cobb-Douglas specification imposes concavity on the relationship among the outputs in a multiple-output cost function. This violates the theoretical assumption that the outputs form a convex set, a production possibilities frontier. Thus for multiple outputs

[6] GJTB used a variant of the translog adapted to incorporate percentage changes in place of natural logs. GJTB also include a cubic term for enrollment.

the Cobb-Douglas specification imposes a theoretically flawed parametric relationship among the outputs.

Given this list of reasons, what are the disadvantages of using a translog? There is a certain complexity in communicating results of the translog, especially in a public policy setting. In particular, this complexity shows up when evaluating marginal effects of changes in outputs or inputs, as the marginal effects vary conditional on the set of outputs and inputs. Other issues that are sometimes mentioned are the loss in degrees of freedom that can result from estimating a large number of parameters. If the number of observations is small and the number of inputs and outputs large, the translog specification may exhaust the available degrees of freedom.

Student Output Measures

The GJTB and I&R studies differ with respect to the variables under analysis, as illustrated in Table 1. While both sets of researchers used a measure of value added, GJTB used a student-based measure of value added whereas I&R used a cohort-based measure.

GJTB calculated the percentage of students in each district who passed the Texas Assessment of Academic Skills (TAAS) test one year and compared it to the percentage of those *same students* who passed two years previously. Their value-added measure was a three-year moving average of the average increase in the district passing rate for grades five through eight and 10.[7] The moving average was introduced to smooth variations in test score gains due to stochastic volatility in student performance.[8] (GJTB also included

[7] The TAAS was administered annually in grades three through eight and 10.
[8] Kane and Staiger (2001) investigated volatility among student test score gains.

the two-year lag of the passing rate in grades three through six as an environmental variable.)

I&R used a TAKS-equivalent passing rate in 2002 for grades five through eight and 10 as their outcome measure, and the TAAS passing rate one year earlier for grades three through eight and 10 as an environmental variable. The I&R cohort match is obviously imprecise, as a direct cohort match would exclude grades three, eight and 10 from the lagged score.

Cost Measures

Both GJTB and I&R described their dependent variable as current operating expenditures excluding food and student transportation expenditures. However I&R included in their definition of current operating expenditures a number of expenditure categories that GJTB excluded. Those categories are community service (function 61), debt service (71), facility acquisition and construction (function 81), and intergovernmental payments (functions 92, 93, 95, 97 and 99). As a result of the difference in definitions, the dependent variable for the I&R averaged $200 per pupil higher than the dependent variable for the GJTB analysis.

Endogeneity

The potential endogeneity of explanatory variables is an important problem, and failure to adequately control for endogeneity may lead to inconsistent parameter estimates. The standard method of dealing with endogenous explanatory variables is instrumental variables. Basically, the analyst must find additional variables that are both correlated with the potential endogenous variable and not correlated with the error terms in the regression. A good instrumental variable will also have a high correlation with the potential endogenous variable.

The widespread use of IV techniques, and the disappointing results sometimes encountered by practitioners, has led to a substantial literature on the pros and cons of using IV estimation, and a large literature on so-called weak instruments. Staiger and Stock (1997) provide a theoretical analysis and conclude that instruments that have low partial correlation with an endogenous explanatory variable can lead to substantial bias, even in large samples. They also suggest computing the F-statistic from the first stage regression and provide some guidelines as to critical values for this statistic.

In a technical appendix GJTB report tests for endogeneity of their output and wage variables (Gronberg at al. 2005). To capture the keeping-up-with-the-Jones's aspect of educational demand, they used the enrollment-weighted average value of each output in the surrounding districts as an instrument for the district's actual output. For each wage series (teacher wages and auxiliary wages) they used the predicted wage as an instrument for the actual wage.[9] GJTB then tested for endogeneity in the outputs and their interactions, for the wages and their interactions, and for endogeneity with respect to both outputs and wages (and their respective interactions). Hausman tests fail to reject the hypothesis that their instrumental variables method was unnecessary. The probability values of the test statistics were all above 95 percent (for outputs, 0.9999; for wages, 0.9545; and for outputs and wages, 0.9999).

I&R attempted to deal with the endogeneity of the school outcome variables by instrumenting for the four school outcome measures in their model (the TAKS-equivalent

[9]Instrumental variables is not commonly done with a stochastic frontier specification, so GJTB used two stage least squares to estimate the translog function.

passing rate, the SDAA passing rate, the retention rate and the share of students scoring above criterion on the SAT or ACT) using five indicators of educational demand -- median household income, property values per pupil (which I&R label "tax price"), the educational attainment of households, the share of households with children, and the share of households that own their homes. Because property values per pupil are likely to be correlated with school district characteristics that are not captured in the I&R model (such as school district facilities or urban location), it seems a problematic choice as an instrumental variable on conceptual grounds. Indeed, our analysis confirms that I&R's chosen instruments lack the necessary properties of good instruments. In particular, they are not well correlated with the potentially endogenous variables,[10] nor are they uncorrelated with the errors from the primary equation.[11] Thus I&R's concern for the endogeneity of the outcomes measures is entirely appropriate, but their choice of instruments and the extremely low correlation with the potential endogenous variables introduces the problem of weak instruments into their analysis, and casts doubt on the reliability of their cost estimates.[12]

---

[10] The most common indicators of the correlation between the endogenous variables and excluded instruments are an F-test of the joint significance of the excluded instruments in the first-stage regression, and the partial R2. Here, the relevant F-statistics are all significant at the 1-percent level. However, the partial R^2 is less than 0.04 for both the SDAA passing rate and the retention rate. In situations such as this, with multiple endogenous variables, inter-correlations among the instruments may render one or more of the instruments irrelevant, and leave the model unidentified despite a statistically significant F-test and a high partial R^2 (Baum, Schaffer and Stillman 2003). Shea (1997) developed a partial R^2 statistic that is a more appropriate measure of instrument relevance for multivariate models. The Shea's partial R^2's are 0.02, 0.02, 0.03 and 0.07 for the TAKS-equivalent passing rate, the SDAA passing rate, the retention rate and the SAT/ACT performance measure, respectively.

[11] Because the number of instruments excluded from the primary equation exceeds the number of endogenous variables, the I&R model is overidentified and a Sargan statistic can be used to test for whether or not the excluded instruments are uncorrelated with the errors from the primary equation. The Sargan statistic has a null hypothesis that the instruments are valid. The Sargan test statistic for the I&R application is 8.396, and the probability of a greater J statistic is 0.0038, rejecting the null hypothesis.

[12] We note that Imazeki and Reschovsky substantially altered the IV specification when their Texas cost analysis was published in a peer-reviewed journal (Imazeki and Reschovsky 2006). In particular, Imazeki and Reschovsky dropped one endogenous outcome variable (the annual retention rate) and modified their list of instruments to exclude median household income and include other demographic characteristics.

Regression Weighting and Heteroskedasticity

I&R weighted their regression by school district enrollment; GJTB did not. Pupil-weighting is often used in the literature on educational cost functions (e.g. Imazeki and Reschovsky 2004a) but not in the cost-function literature for other service industries (e.g. Bilodeau et al. 2000). The theoretical justification for per pupil weighting is unclear. If school districts are the decision-making units for the cost analysis, it is not immediately obvious why decision-making units should receive different weights due to enrollment. The statistical justification is potentially more appealing, as pupil-weighting might be justified as a way to correct for heteroskedasticity. Unfortunately, Taylor (2004) shows that the I&R model has heteroskedastic residuals even after pupil weighting.

Heteroskedasticity in the I&R model raises other concerns regarding hypothesis testing. It is well known in the econometrics literature that hypothesis tests on coefficients are biased in the presence of heteroskedasticity. The textbook solution is to use robust standard errors, i.e. standard errors that are constructed to result in valid t-statistics in the presence of heteroskedasticity.[13] Taylor shows that, using robust standard errors, the coefficient on the TAKS-equivalent passing rate in the I&R model is insignificant at the 10-percent level (Table 2). Thus the coefficient estimate that I&R relied on to generate their $2 billion cost estimate was statistically insignificant at standard significance levels.

[13] Wooldridge (2002) writes "it has become popular to estimate β by OLS even when heteroskedasticity is suspected but to adjust the standard errors and test statistics so that they are valid in the presence of arbitrary heteroskedasticity" (p. 56).

Inefficiency

Another important difference between the models is their treatment of inefficiency. A sizeable literature suggests that school districts do not all operate in an efficient, cost-minimizing fashion and that the degree of inefficiency varies considerably across districts. For example, in a previous analysis, Imazeki and Reschovsky (2004a) had found that "the average district in Texas is 59 percent as efficient as the most efficient districts in the state" (p. 41). In their analysis for the plaintiffs, I&R argued that school inefficiency is correlated with the degree of competition in the local education market, and the I&R model includes a Herfindahl index of market concentration to capture that inefficiency.[14] However, even if competition is correlated with inefficiency, there is no reason to believe that including the Herfindahl index is sufficient to capture the inefficiency in the system. When we re-estimate the I&R model without their weak instruments but otherwise unchanged, and in a stochastic frontier setting, we find that even with the Herfindahl index included, the average district in the sample is 12 percent inefficient (Table 3). If we estimate a frontier that is not pupil weighed, the average inefficiency increases to 17 percent (again with the Hefindahl index included in the model).

In contrast, GJTB used stochastic frontier analysis to allow for potential inefficiency. As mentioned above, the stochastic frontier method allows estimation of a cost function under the assumption that observed costs vary from the frontier for two reasons, a standard two-sided error term and a one-sided error term meant to capture

[14] I&R calculate their Herfindahl index based on traditional public school enrollments in each county. This measure does not reflect competition from charter schools or private schools.

inefficiency. This stochastic cost frontier is below the cost frontier that would be estimated using a cost function that only includes the standard two-sided error terms.

Differences in Predicting Adequacy

One of the attractions of the cost function approach is that it can be used to generate estimates of the cost associated with meeting any reasonable performance standard.[15] Both I&R and GJTB estimated the cost of achieving an average TAKS-equivalent passing rate of 55 percent, holding all other outcomes in their models constant at the state average.

The concept of a 55-percent TAKS-equivalent passing rate requires some explanation. Until 2003, Texas policy makers had a system for evaluating school districts in Texas that focused primarily upon student performance on the TAAS. In 2003, Texas replaced the TAAS with the Texas Assessment of Knowledge and Skills (TAKS). The TAKS is widely recognized as a more challenging test than the TAAS. Furthermore, the TAKS has been getting harder each year as the Texas Education Agency (TEA) phased in the recommended passing standard on the new, harder test.

TAKS data were not available for estimation, so the research teams were forced to improvise. On the basis of field trials, TEA had developed a schedule for converting TAAS scores into TAKS-equivalent scores. For example, the schedule indicated that where a score of 70 would have been high enough to pass the TAAS, a third grader would need to have scored an 82 to pass the math test at the recommended TAKS standard. GJTB re-graded each individual student's TAAS test for 2002 using the

[15] By design, statistical models describe relationships within the experience of the data. It is problematic to extrapolate beyond that experience to predict the costs associated with a level of performance that is not regularly achieved, or is not achieved by districts with a particular set of geographic and demographic characteristics.

recommended grading standards given in the conversion table, and using these re-graded exams, calculated the TAKS-equivalent passing rate for each district. District TAKS-equivalent passing rates ranged from 19 percent to 88 percent, with an average of 53 percent. GJTB used these TAKS-equivalent passing rates to calculate the change in passing rate required to reach a 55-percent passing threshold. I&R took a different approach, and used these TAKS-equivalent passing rates as an outcome measure in their estimation, and based their cost projections on the resulting coefficient estimate.

GJTB chose to evaluate the cost of having 55 percent of students passing TAAS at the TAKS level for a number of reasons. At the time of the GJTB analysis, the TEA had not yet released its accountability standards for 2004. However, it was reasonable to expect that the passing rate required for a school district to be considered academically acceptable would be no higher in the TAKS world than it had been in the TAAS world. In the final year of TAAS testing, 55 percent of each type of student had to pass TAAS for a district to be considered acceptable. In addition, Texas's Consolidated State Application Accountability Workbook detailed the state's plans for compliance with the No Child Left Behind Act of 2001. The workbook set out a progression of increasing passing rates on TAKS. The designated goal for 2006 was that the percent of students in all grades and demographic categories passing the TAKS be 53.5 percent for reading/language arts, and 41.7 percent for math.

For comparability with GJTB, I&R also evaluated the cost of achieving a passing rate of 55 percent. However, I&R extended their analysis to cover higher performance standards as well.

The large difference in the forecasted costs of adequacy from the two models turns on the estimated marginal impact of a value added performance measure on costs. While the two studies used different measures of student test performance, and slightly different measures of expenditures, the main reason that I&R estimated a *$2 billion* price tag for educational adequacy while GJTB estimated a *zero* price tag is differences in the estimated impact of student test performance on costs. GJTB estimated that the marginal impact on costs of a one percentage point increase in the passing rate on the Texas statewide TAKS test was 0.3354. In contrast, I&R estimate the marginal impact on costs was 2.2824. Differences in performance measures, in treatment of efficiency, in fungibility of funds, all played an important, but secondary, role in the respective COA estimates. We detail and differentiate the two forecasting approaches below.

The GJTB Approach.

To calculate predicted costs, GJTB first calculated the predicted cost for a benchmark district that had the mean value for each right-hand-side variable. Each of the non-interacted variables was set to the mean value, then the second order terms were calculated. The benchmark district's predicted cost was calculated from the cost function by multiplying each right-hand-side variable by the estimated coefficient and adding up each of the terms. Because measured inefficiency can reflect unmeasured outcomes as well as true inefficiency, GJTB presumed that the benchmark district had average inefficiency, and added the mean one-sided error for 2002 to the benchmark (log) predicted cost.

The above calculation provided predicted cost for a benchmark district. To calculate predicted costs for each individual district, GJTB calculated the marginal effect on cost of each right-hand-side variable, evaluated at the mean of all the other variables.[16] Table 3 presents those marginal effects. GJTB then linearized the cost function to generate cost predictions for each district. The predicted cost for each district was the predicted cost for the benchmark district, plus each of the linearized coefficients multiplied by the difference in value of the right hand side variable between the specific individual district and the benchmark district.[17]

The linearization was a first order Taylor expansion in all variables except for district enrollment and auxiliary worker salary. For enrollment, GJTB also included the second and third order terms, and for auxiliary worker salary they included the interaction term of auxiliary worker salary and district enrollment. These two adjustments were made to reduce the loss in accuracy associated with simple linearization. The relationship between cost and enrollment is clearly nonlinear, and including the enrollment-auxiliary worker interaction greatly improves the match between the predicted cost estimates from the (quasi-) linearized model and those of the full translog model for the largest districts.

GJTB argued that relying on the marginal effects to generate district-level cost estimates rather than each specific coefficient estimate enhances the usefulness of their cost function analysis in the design of a school finance formula. Providing guidance for the design of a school finance formula was the primary objective of the GJTB analysis.

For the adequacy study GGJTB generated a predicted cost of having 55 percent of students passing TAAS at the TAKS level. To do this, GJTB calculated the gap between

[16] For more on the calculation of marginal effects, see Gronberg et al. (2005).

[17] To ensure that their estimates are based on factors outside of school district control, GJTB substitute a teacher wage index for average actual wages when generating their cost estimates. Gronberg et al. (2005) demonstrate that the analysis generates similar cost projections if the wage index is used instead of the average actual wage in the estimation of the translog cost function.

each district's TAKS-equivalent average passing rate and 55 percent. They then calculated the cost of producing at least the state average on the outcome measures *plus* whatever additional value added would be required to close this performance gap. If a district had a TAKS-equivalent average passing rate of 52 percent (or higher), they calculated the cost of producing a three percentage point increase in TAAS average passing rates (together with the state averages for advanced courses and SAT/ACT performance). If the district had a TAKS-equivalent average passing rate below 52 percent, then GJTB calculated the cost of achieving enough percentage point increases in the passing rate to bring the district up to a 55 percent average passing rate. For example, if the district's TAKS-equivalent average passing rate was 50 percent, then GJTB calculated the cost of achieving a five percentage point increase in the passing rate.[18]

To calculate the additional resources needed to meet the 55-percent passing standard, GJTB compared each district's actual expenditures in 2002 with the level of expenditures their model predicted would be necessary to reach the 55-percent performance standard, given district-specific wages and environmental factors, and holding the other two outcome measures (the percent scoring above criterion on the SAT or ACT and the percent taking advanced courses) constant at the state mean. If the predicted expenditures were less than observed expenditures, GJTB concluded that the school district required no additional resources. If the predicted expenditures exceeded the observed expenditures, then GJTB concluded that the school district required

[18] In so doing, GJTB implicitly assumed that the cost of achieving a one percentage point increase in TAKS passing rates was no higher than the cost of achieving a one percentage point increase in TAAS average passing rates. This assumption has been borne out by the marginal effects estimates from models estimated using TAKS data (Gronberg et al. 2005).

additional resources equal to the difference between actual and predicted expenditures. In other words, GJTB treated school district expenditures as completely fungible.

For purposes of litigation, Taylor extended the GJTB model to predict costs in 2004, based on 2003 data (Taylor 2004). She found that only one of the 46 plaintiff districts had actual expenditures in 2003 that were lower than the level of expenditures predicted by the GJTB model. That district -- Kaufman ISD -- spent $68,000 less than the model predicted would be required to reach the performance standard (in $2004). Therefore, holding all districts harmless, she estimated that the plaintiff districts as a whole would require an additional $68,000 to meet the 55-percent passing rate performance standard.

<u>The I&R approach.</u>

To calculate the cost of achieving the 55-percent passing rate in each school district, I&R used their cost function coefficients to predict the level of expenditures required to achieve the observed TAKS-equivalent passing rate, holding all other outcomes constant at the state average, and holding the teacher wage index and all environmental variables except the Herfindahl index at their observed, district-specific values.[19] I&R set the Herfindahl index at the 90th percentile for competition, thereby assuming that all districts operate in highly competitive environments and by I&R's reasoning implicitly assuming low inefficiencies. However, this assumption also had the effect of assuming a lower degree of competition than the Herfindahl index would indicate (and, by I&R's reasoning, less efficiency) for 81 districts including Houston and San Antonio ISDs.

[19]For details, see Imazeki and Reschovsky (2004).

I&R used the cost predictions at the observed values for the passing rate (and the mean values for all other outcomes) as the baseline for their analysis. They then predicted cost for a number of alternative specifications of the TAKS passing rate, holding all other elements of the cost function constant at the baseline values. I&R's estimate of the additional resources needed by a school district was the difference between their baseline cost estimate for the district and their alternative cost estimate. That is, they compared two estimates of cost. Actual spending was not a direct part of the calculation.

The least-cost alternative that I&R evaluated was one in which school districts had three years to close the gap between the observed TAKS equivalent passing rate and 55 percent. If the gap was less than three percentage points or if the school was already above a 55 percent passing rate, I&R set the TAKS equivalent passing rate to 55 percent. If the gap was more than three percentage points, they set the TAKS equivalent passing rate equal to the observed TAKS equivalent passing rate plus one third of the gap. Under this alternative, I&R estimated that it would cost *$456.5 million* to meet the 55-percent passing rate standard. When they assumed that all gaps would be closed immediately, and that all districts would be expected to increase passing rates by at least 3 percentage points per year (the alternative they deemed most similar to the GJTB analysis), I&R estimated that the plaintiff districts would require an additional *$1,735 million*.

Notably, the I&R cost estimates were derived by comparing one prediction from their model with another prediction from their model. If the projected cost of meeting a designated TAKS performance standard was $100 higher than the projected cost of meeting their baseline performance standard, then I&R concluded that the district needed an additional $100—even if the district was already spending $300 more than the projected cost of meeting their baseline. In other words, I&R treated school district expenditures as completely non-fungible.[20]

3. Robustness

Duncombe (2006) calls attention to the issue of reliability in estimating the cost of educational adequacy. One method of checking reliability is to investigate the sensitivity of results to alternative specifications and alternative modeling choices. GJTB (2005) provide an extensive list of alternative specifications, which we include here as Tables 3a-3c. These alternative models explore alternative definitions of expenditures, alternative measures of family income, including predicted wages instead of actual wages, imposing a Cobb Douglas specification, estimating the model without the stochastic frontier, estimating the model with alternative student output measures, estimating the model without small districts, estimating the model with only 2004 data. While all these alternative specifications yielded somewhat different coefficient estimates, the mean predicted expenditures for achieving average performance on the models' outcomes measures were remarkably consistent across specifications, ranging from $6,283 per pupil in their baseline specification to a high of $6,478 and a low of $6,015 (Table 4). Moreover, the estimated cost of achieving various performance measures ranged from a mean of $6,389 in their baseline model to a high of $6,483 in a

[20] Had I&R treated school district expenditures as completely fungible within school districts, an interesting pattern would have appeared. We compared the predicted cost from their least-cost alternative with actual district expenditures rather than with the predictions from their baseline model. In all but seven plaintiff districts, actual expenditures in 2002 exceeded the model predictions. However, the I&R model predicted that costs for Dallas ISD were so far above actual expenditures that the district alone would need have needed an additional $479 million to cut the gap by one third. Houston ISD would have needed an additional $321 million. All told, had I&R treated school district spending as completely fungible, they would have predicted that the plaintiff districts required nearly twice what they predicted by assuming that expenditures were not fungible.

model using 2004 data (table 5). Thus the GJTB model held up well, both in terms of reasonableness of estimated marginal effects and consistency of predicted cost of adequacy across a wide array of alternative specifications.

In contrast, there is evidence that the I&R estimates are more fragile. We estimated their model without the suspect instruments and found the coefficient measuring the marginal effect of student performance on cost fell from 2.28 to 0.41. If we don't pupil weight, the marginal effect falls further, to 0.23. Finally, if we estimate a stochastic frontier (with or without pupil weights) the coefficient falls as well. These order-of-magnitude changes are troubling, especially when there is such strong evidence of weak instruments, when pupil weights fail to correct for heteroskedasticity and are otherwise unjustified, and when the stochastic frontier gives evidence of large average inefficiencies.

These differences in coefficient estimates lead to large differences in predicted costs. We took each model from Table 2 and followed the I&R procedure to generate estimates of the additional spending required to yield a passing rate of 55 percent. These calculations are summarized in Table 6. As the table illustrates, the estimates I&R included in their report are much higher than those generated by the alternative specifications.

As an example, consider again the case of Dallas ISD. I&R's baseline estimate is that it would have cost an additional $1,200 to bring the passing rates in DISD up to 55 percent. However, none of the other specification yield an estimate for DISD above $200 per student. Because DISD educates 160,000 students, the more than $1,000 difference across the models implies a $180 million dollar difference in the total cost estimate.

4. The MAP Professional Judgment Model

Management Analysis and Planning, Inc. (MAP) conducted a professional judgment analysis of the cost of providing an adequate education in the 46 plaintiff districts. In a professional-judgment study, researchers estimate the cost of an adequate education by asking panels of educators to design a school that could meet certain objectives. The researchers then tabulate the cost of replicating those prototypical schools in each school district. In this case, MAP asked the panelists to evaluate the instructional resources needed to offer "all Texas children access to a quality education that enables them to achieve their potential and fully participate now and in the future in the social, economic, and educational opportunities of our state and nation" (Smith and Seder 2004, 5).

MAP recruited 28 Texas educators to serve on their professional judgment panels. Those individuals were far from disinterested parties. Half of the panelists were drawn from districts that were plaintiffs in the lawsuit, even though those plaintiff districts employed less than a quarter of total Texas teachers.

MAP then divided the 28 educators into five panels, and asked each panel to develop seven prototypes for elementary schools, seven prototypes for middle schools, and seven prototypes for high schools. The seven prototypes each correspond to different combinations of student characteristics. The least-needy school modeled had 4 percent Limited English Proficient (LEP) students and 13.5 percent of students eligible for free or reduced lunch (FRL). The most-needy school modeled had 32 percent LEP and 80 percent FRL students. As the authors note, these ranges (from 4 percent to 32 percent for LEP and from 13.5 percent to 80 percent for FRL) roughly correspond to the 10^{th} and 90^{th}

percentiles of school district demographics for the 46 plaintiff districts. In other words, 80 percent of the children in plaintiff school districts attended a district with average characteristics within the designated range.

In order to estimate the cost of education in each school district, MAP calculated the cost of providing the resources called for in each of the prototype schools developed by the panels. MAP then "synthesized" the panel predictions to generate cost estimates for each school in a plaintiff district, and aggregated those school-level estimates to generate district level estimates.

MAP's method for deriving a synthesis model is a little peculiar. Within each grade level (elementary, middle and high school), MAP researchers ran a regression of school cost on the share of students receiving free and reduced price lunch (FRL). They then ran a second regression of cost on the share with limited English proficiency (LEP). Finally the researchers strung together the separately estimated slope coefficients from the two regressions into a single equation, and adjusted the intercept term so that their resulting equation generated the grade-level average spending when FRL equals 50 percent and LEP equals 15 percent. Such an approach is unbiased only if LEP and FRL are uncorrelated. If FRL and LEP are positively correlated in the data (and they are) then each slope coefficient will overstate the independent effect of the corresponding student characteristic. This pattern is illustrated most clearly for grades K-5. As Figure 2 illustrates, the MAP synthesis model is close to the lowest of the panel estimates for low-needs schools, but near the top of the panel estimates for high-needs schools.

In the end, this unusual technique generated a synthesis model that has little power to explain the variation in costs across the panel exercises. The MAP synthesis model has an explained variation (R-square) of barely 25 percent at the elementary level, and only 5 percent at the middle or high school levels. In contrast, traditional multivariate regression analysis can explain 61 percent of the variation at the elementary school level, and roughly 13 percent of the variation at the middle and high school levels.

MAP's unusual synthesis model is particularly problematic because most schools in the 46 plaintiff districts have student demographics that are outside of the range of demographics the panels evaluated. Only 372 of the 1464 schools in the plaintiff districts fall within the study parameters (Figure 3). Over 70 percent of the students in the plaintiff districts attend schools for which MAP had to extrapolate cost estimates using their synthesis model. As a result, 73 percent of the total estimated cost was based on extrapolation outside of the range of the data.[21] Cost estimates based on extrapolation from limited data using a questionable model are highly suspect.

To arrive at its final cost estimate, MAP adjusted its district-level estimates of school instructional cost for regional wage variations using the same cost index used by GJTB and I&R, and then added the district-specific amount of observed non-school and non-instructional expenditures.[22] As did GJTB, MAP then compared its final cost estimate to observed current expenditures to determine the additional funds needed to achieve adequacy. Assuming that spending could be reallocated among plaintiff districts,

[21] MAP tries to address this concern by offering a set of "bounded" cost estimates. Given their synthesis model, such an approach results in the cost estimate being rounded up for 34 districts and rounded down for eight districts. (The remaining four districts are unaffected) The baseline estimates are suspect because they derive from extrapolation outside of the data and the bounded estimates are biased upwards for most districts.

[22] The estimates of instructional cost are adjusted for geographic cost variation using a cost of education index developed by Taylor (2004). According to MAP data, the non-instructional share of cost among plaintiff districts ranges from 32 percent of current expenditures to over 86 percent of current expenditures.

MAP estimated that achieving adequacy in the plaintiff districts would cost an additional $683 to $830 million per year.[23]

Given the vagueness of the outcomes measure underlying the MAP analysis, we cannot determine whether or not the standards have been met, and thus cannot compare the cost estimates to the realizations. Indeed, this is one of the seldom mentioned weaknesses of the professional judgment approach—it is impossible to check the validity of the cost projections against any objective standard.

5. Conclusions

We think the conclusion to be drawn from the dueling Texas analyses is clear. The appropriate conclusion is *not* to abandon cost function estimation as either hopelessly flawed in general or as flawed and inapplicable to public schools. Such a move would cede the field to other approaches—such as the professional judgment approach—that can be subjected to equally vociferous criticism. Rather, the appropriate conclusion is that it is critically important to demand best-practice techniques from any analyst of educational adequacy. For cost function analysis, best practice requires researchers to adopt appropriate modeling and estimation strategies and to check carefully for robustness and reliability of results. For professional judgment analyses, best practice requires researchers to ensure that the panelists have no conflicts of interest, that the panels are asked to develop an appropriate range of prototypes, and that the panel recommendations are used appropriately to develop cost projections. It is not possible to produce a 'gold standard' study using 'base' practices.

[23] The lower estimate comes from the bounded forecast. The higher estimate comes from the unbounded forecast.

6. References

Bilodeau, Daniel, Pierre-Yves Cremieux, and Pierre Ouellette. 2000, Hospital Cost Function in a Non-market Health Care System. *Review of Economics and Statistics*, 82(3) pp. 489-98.

Baum, Christopher F., Mark E. Schaffer and Steven Stillman (2003) "Instrumental Variables and GMM: Estimation and Testing," Boston College Department of Economics Working Paper no. 545.

Costrell, Robert, Hanushek, Eric, and Loeb, Susanna (2007) "What do Cost Functions Tell Us About the Cost of an Adequate Education?" Conference paper manuscript (*From Equity to Adequacy to Choice: Perspectives on School Finance and School Finance Litigation* conference at Show-Me Institute and Truman School of Public Affairs of the University of Missouri).

Duncombe, William (2006), "Responding to the Charge of Alchemy: Strategies for Evaluating the Reliability and Validity of Costing-Out Research." *Journal of Education Finance,* forthcoming.

Gronberg, T, Jansen, D, Taylor, L. L, and Booker, K. "School outcomes and school costs: the cost function approach." Texas Joint Select Committee on Public School Finance, Austin, TX; 2004. http://bush.tamu.edu/research/faculty%5Fprojects/txschoolfinance/.

Gronberg, T, Jansen, D, Taylor, L. L, and Booker, K. "School Outcomes and School Costs: A Technical Supplement." Texas Joint Select Committee on Public School Finance, Austin, TX; 2005. http://bush.tamu.edu/research/faculty%5Fprojects/txschoolfinance/.

Hanushek, Eric A. (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.

Hanushek, Eric A. (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, J. and A. Reschovsky. 2006. “Does No Child Left Behind Place a Fiscal Burden on States? Evidence from Texas.” *Education Finance and Policy*, 1(2) 217-46.

Imazeki, J. and A. Reschovsky. 2005. “Assessing the Use of Econometric Analysis in Estimating the Costs of Meeting State Education Accountability Standards: Lessons From Texas.” *Peabody Journal of Education*, 80(3) 96-125.

Imazeki, J. and A. Reschovsky. 2004a. “Financing education so that no child is left behind: Determining the costs of improving student performance.” Developments in School Finance 2003: National Center for Education Statistics: 33-52.

Imazeki, J and Reschovsky A. “Estimating the Costs of Meeting the Texas Educational Accountability Standards.” manuscript. 2004.

Kane, Thomas J. and Staiger, Douglas, "Improving School Accountability Measures" (March 2001). NBER Working Paper No. W8156.

Shea, John (1997) “Instrument Relevance in Multivariate Linear Models: A Simple Measure,” *Review of Economics and Statistics*, 79(2): 348-52.

Smith, J.R. and Seder, “Estimating the Cost of Meeting State Educational Standards.” manuscript 2004.

Staiger, D. and J.H. Stock, “Instrumental Variables Regression with Weak Instruments,” *Econometrica* 65, 1997, 557-586.

Taylor, L.L. “Estimating the Cost of Education in Texas,” Texas Office of the Attorney General, 2004.

Wooldridge, Jeffrey M. Econometric Analysis of Cross Section and Panel Data. MIT Press, 2002.

Figure 1: Per-Pupil Cost Projections and Real Expenditure Changes, W.O.C. Plaintiffs

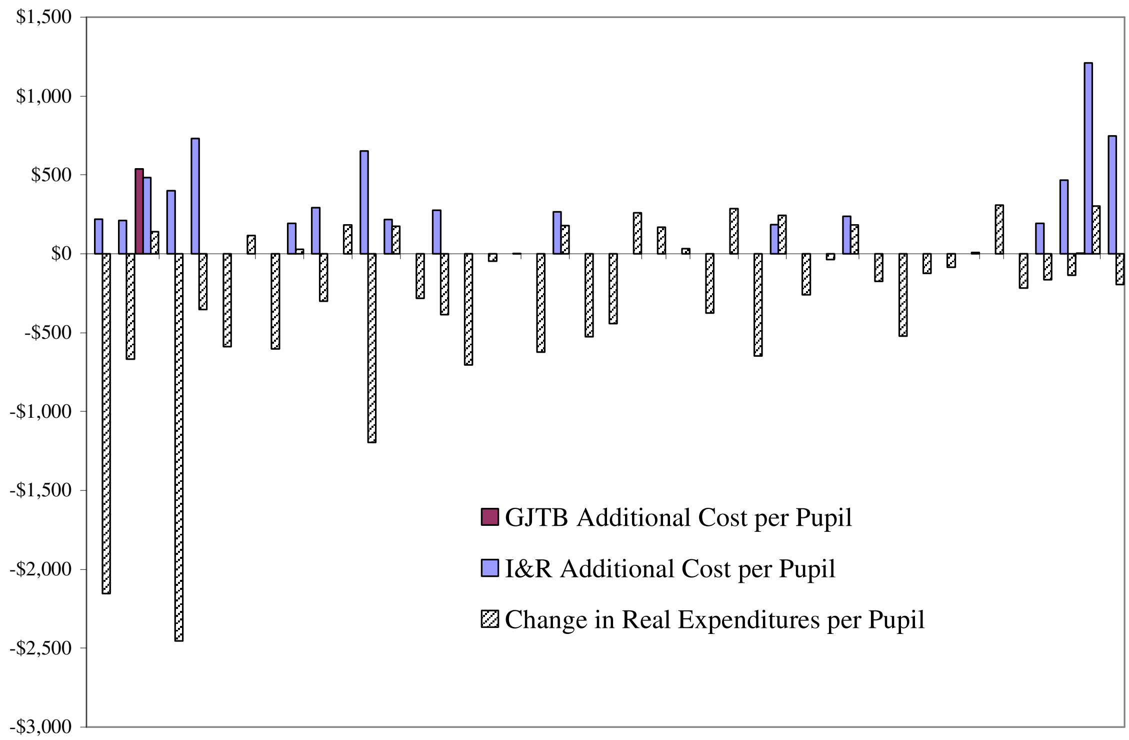

Figure 2: The MAP K-5 Synthesis Model Exaggerates the Cost Impact of Student Need

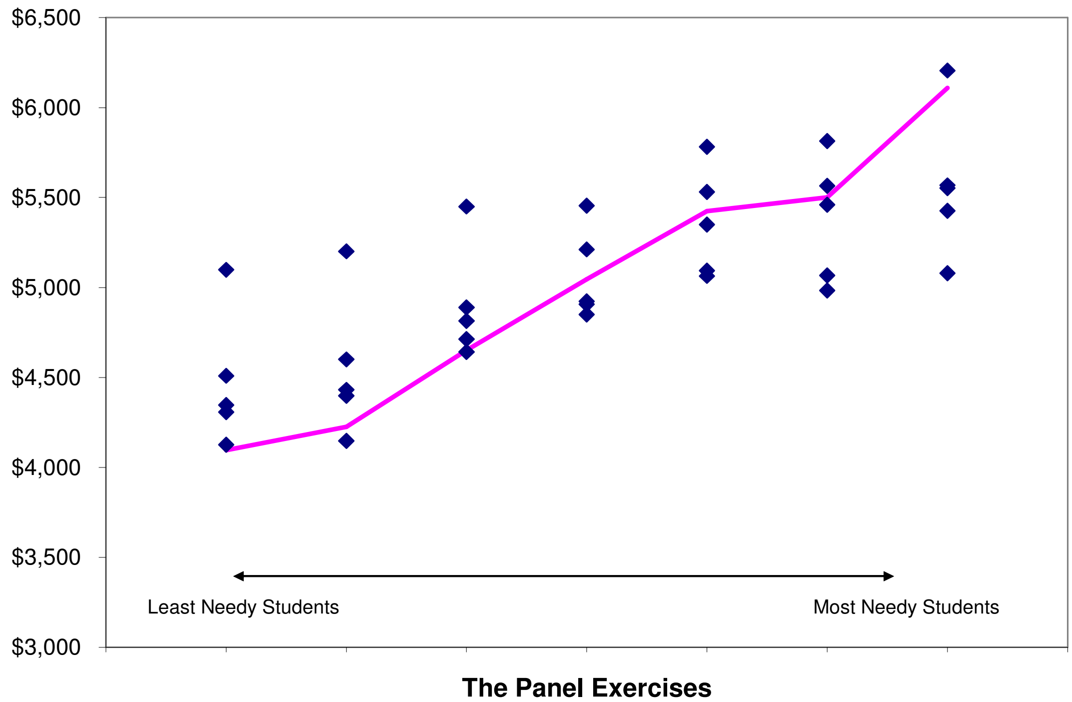

Figure 3: Most Plaintiff Schools are Outside the MAP Study Parameters

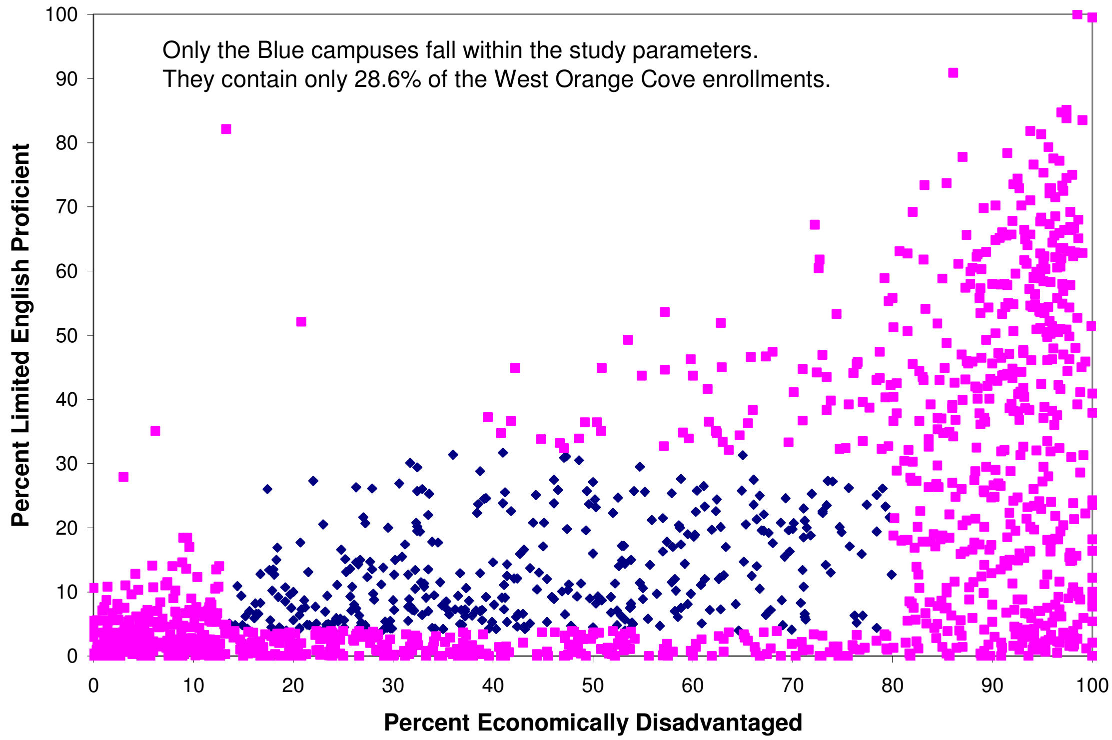

Table 1: The Variables Used in the Cost Function Analyses		
	GJTB	**I&R**
Dependent Variable	Per pupil current expenditures (log)	Per pupil current expenditures[i] (log)
Outcomes	Change in TAAS Passing Rate Percent Above Criterion SAT/ACT Percent Completing an Advance Course	TAKS Equivalent Passing Rate Percent Above Criterion SAT/ACT Retention Rate State Developed Alternative Assessment passing rate
Input Prices	Beginning Teacher Wage Average Auxiliary Wage	Teacher Wage Index
Student Characteristics	Percent low income[ii] Percent limited English[iii] Percent special education, learning or speech disabilities Percent special education, other Prior average TAAS score (two-year lag, grades 3-6)	Percent low income Percent limited English Percent special education, learning or speech disabilities Percent special education, other Prior average TAAS score (one-year lag, grades 3-8, 10) Percent black Percent Hispanic
Additional factors	Enrollment Distance from metro area	Enrollment Herfindahl index

[i] The cost measure used by I&R includes not only those current operating expenditures used by GJTB, but also current operating expenditures for community service function 61), debt service (function71), facility acquisition and construction (function 81), and intergovernmental payments (functions 92, 93, 95, 97 and 99). According to TEA, function 61 “is used for expenditures that are for activities or purposes other than regular public education and adult basic education services.”

[ii] GJTB use the share of students receiving free lunch in the elementary grades as their measure of student poverty. I&R use the share of students receiving free and reduced lunches in the district as a whole.

[iii] GJTB use the share of students who are limited English proficient in the elementary grades. I&R use the share of LEP students in the district as a whole.

Table 2: Coefficient Estimates from Alternative Specifications of the I&R Cost Function

	IV	SFA Weighted	SFA Unweighted	OLS weighted	OLS unweighted
TAKS-equivalent	2.282	0.432	0.148	0.411	0.229
passing rate	(1.593)	(0.001)**	(0.066)*	(0.119)**	(0.092)*
SDAA	0.006	0.001	0.000	0.001	0.001
	(0.005)	(0.000)**	(0.000)	(0.000)*	(0.000)
Annual Retention Rate	0.104	-0.006	0.004	-0.011	-0.002
	(0.138)	(0.000)**	(0.008)	(0.012)	(0.010)
Percent Above Criterion	0.059	0.425	0.145	0.411	0.134
on SAT or ACT	(0.547)	(0.001)**	(0.061)*	(0.097)**	(0.080)
Composite lagged	-0.033	-0.007	-0.001	-0.006	-0.001
TAAS passing rate	(0.017)	(0.000)**	(0.001)	(0.002)**	(0.002)
Teacher wage index	0.859	0.712	0.591	0.655	0.489
(log)	(0.345)*	(0.002)**	(0.129)**	(0.170)**	(0.171)**
Log Enrollment	-0.286	-0.284	-0.314	-0.322	-0.337
	(0.063)**	(0.000)**	(0.030)**	(0.030)**	(0.032)**
Log Enrollment,	0.013	0.013	0.015	0.015	0.016
squared	(0.003)**	(0.000)**	(0.002)**	(0.002)**	(0.002)**
Percent Black	0.350	0.153	0.231	0.165	0.281
	(0.125)**	(0.000)**	(0.038)**	(0.048)**	(0.048)**
Percent Hispanic	0.312	0.122	0.207	0.146	0.260
	(0.083)**	(0.000)**	(0.030)**	(0.038)**	(0.042)**
Percent low SES	0.424	0.270	0.140	0.200	0.059
	(0.183)*	(0.001)**	(0.046)**	(0.076)**	(0.064)
Percent LEP	-0.645	-0.299	-0.056	-0.222	0.093
	(0.358)	(0.002)**	(0.141)	(0.211)	(0.199)
Percent LEP, squared	0.499	0.225	-0.125	0.154	-0.292
	(0.518)	(0.003)**	(0.268)	(0.313)	(0.319)
Percent learning or	1.201	0.603	0.582	0.723	0.904
speech disabilities	(0.860)	(0.002)**	(0.151)**	(0.266)**	(0.274)**
Percent other disabilities	-0.429	-0.111	0.411	-0.151	0.020
	(1.708)	(0.009)**	(0.540)	(0.843)	(0.783)
Percent other	2.710	4.619	2.250	4.902	2.794
disabilities, squared	(10.829)	(0.079)**	(3.294)	(4.583)	(3.246)
Percent high school	-0.464	-0.158	0.556	-0.293	0.657
	(0.309)	(0.002)**	(0.131)**	(0.221)	(0.213)**
Herfindahl index (log)	-0.008	-0.004	-0.011	-0.005	-0.013
	(0.014)	(0.000)**	(0.004)**	(0.005)	(0.005)*
Constant	-5.717	5.262	4.781	6.485	6.317
	(12.401)	(0.015)**	(1.209)**	(1.548)**	(1.630)**
Observations	827	827	827	827	827
R-squared	0.18			0.39	0.45

Robust standard errors in parentheses. * significant at 5%; ** significant at 1%

Table 3a: Marginal Effects from Alternative GJTB Specifications

Marginal Effect of:	A:Baseline Model	B: Model A Excluding Object Codes 6401-6499	C: Model A but with Free and Reduced Lunch	D:Model A but with Predicted Wages	E:"Cobb Douglas" Version of Model A
Change in Average TAAS Passing Rate	0.3354	0.3301	0.3033	0.4167	0.0076
	(.1377)	(.13737)	(.1379)	(.1450)	(.1117)
% Completing an Advanced Course	0.1646	0.1437	0.1799	0.1209	0.0955
	(.0379)	(.0378)	(.0379)	(.0397)	(.0325)
% Above Criterion on SAT/ACT	0.0716	0.0859	0.0512	0.1296	0.2515
	(.0421)	(.0419)	(.0421)	(.0449)	(.0362)
Prior TAAS Passing Rate	0.0353	0.0210	0.0079	0.0872	-0.3010
	(.0632)	(.0630)	(.0629)	(.0671)	(.0530)
Average Monthly Salary for Beginning Teachers (log)	0.4201	0.4145	0.4246	0.5995	0.4048
	(.0345)	(.0341)	(.0345)	(.0458)	(.0316)
Average Monthly Salary for Auxiliary Workers (log)	0.2551	0.2540	0.2512	0.3051	0.2852
	(.0193)	(.0192)	(.0193)	(.0319)	(.0196)
District Enrollment (log)	-0.0809	-0.0777	-0.0794	-0.0780	-0.0809
	(.0037)	(.0037)	(.0037)	(.0047)	(.0037)
% Free Lunch (non-high school)	0.2816	0.2918	0.2475	0.2547	0.2048
	(.0201)	(.0201)	(.0183)	(.0215)	(.0164)
% Limited English Proficient (non-high school)	0.1915	0.1857	0.2202	0.1987	0.1263
	(.0438)	(.0436)	(.0433)	(.0472)	(.0223)
% Less Severe Special Education	0.5490	0.5502	0.5564	0.2907	0.6527
	(.0823)	(.0821)	(.0824)	(.0886)	(.0759)
% High school	0.5973	0.5957	0.6212	0.5471	0.3981
	(.0810)	(.0807)	(.0825)	(.0868)	(.0813)
Miles to Nearest Metro Center (log)	0.0208	0.0174	0.0228	0.0374	0.0267
	(.0033)	(.0033)	(.0033)	(.0046)	(.0027)
% Severe Disability Special Education	0.8253	0.7897	0.8174	0.5923	-0.1805
	(.2004)	(.1993)	(.2009)	(.2127)	(.1618)
Log Likelihood	2754.65	2768.07	2753.67	2568.25	2451.24
Observations	2755	2755	2755	2749	2755

Table 3b: Marginal Effects from Alternative GJTB Specifications

Marginal Effect of:	A:Baseline Model	F: Model A but without frontier errors	G: Model A but percent passing both math & reading	H: Model A with low income TAAS scores	I: Model A but only districts with more than 1600 students
Change in Average TAAS Passing Rate	0.3354	0.4180	0.1954	0.0763 (0.0940)	0.2435
	(.1377)	(.1498)	(.09509)		(0.2289)
% Completing an Advanced Course	0.1646	0.1610	0.1592	0.1958	0.1486
	(.0379)	(.0407)	(0.0379)	(0.0442)	(0.0681)
% Above Criterion on SAT/ACT	0.0716	0.0828	0.0744	-0.0356	-0.0312
	(.0421)	(.0453)	(0.0432)	(0.0483)	(0.0718)
Prior TAAS Passing Rate	0.0353	0.0264	0.0263	-0.0275	0.1496
	(.0632)	(.0680)	(0.0481)	(0.0462)	(0.1042)
Average Monthly Salary Beginning Teachers (log)	0.4201	0.4173	0.4112	0.4074	0.4002
	(.0345)	(.0381)	(0.0345)	(0.0385)	(0.0524)
Average Monthly Salary Auxiliary Workers (log)	0.2551	0.2800	0.2635	0.2933	0.3291
	(.0193)	(.0208)	(0.0192)	(0.0225)	(0.0326)
District Enrollment (log)	-0.0809	-0.0803	-0.0811	-0.0814	-0.0676
	(.0037)	(.0042)	(0.0038)	(0.0043)	(0.0107)
% Free Lunch (non-highschool)	0.2816	0.2690	0.2736	0.2657	0.3364
	(.0201)	(.0210)	(0.0200)	(0.0227)	(0.0317)
% Limited English Proficient (non-highschool)	0.1915	0.1942	0.1963	0.1615	0.1644
	(.0438)	(.0471)	(0.0438)	(0.0474)	(0.0629)
% Less Severe Special Education	0.5490	0.5469	0.4968	0.5001	0.6533
	(.0823)	(.0914)	(0.0819)	(0.0959)	(0.1527)
% Highschool	0.5973	0.6181	0.5929	0.7429	0.7688
	(.0810)	(.0890)	(0.0808)	(0.0955)	(0.1553)
Miles to Nearest Metro Center (log)	0.0208	0.0238	0.0204	0.0194	0.0100
	(.0033)	(.0035)	(0.0033)	(0.0037)	(0.0054)
% Severe Disability Special Education	0.8253	0.7594	0.8667	0.1618	0.5882
	(.2004)	(.2148)	(0.2005)	(0.2440)	(0.3841)
Log Likelihood	2754.6	2644.6	2754.3	2348.1	1745.0
Observations	2755	2755	2755	2179	1418

Table 3c: Marginal Effects from Alternative GJTB Specifications, continued

Marginal Effect of:	J: Model A passing both math & reading, free & reduced	K: Model J using 2004 data	L. Model K excluding Advanced Courses & SAT/ACT	M: Model L including completion rate
Change in Average Percent Passing Both Math & Reading (TAAS or TAKS)	0.1774 (0.0953)	0.0913 (0.1186)	0.1376 (0.1181)	0.0903 (0.1220
Completion Rate				-0.1615 (0.2172
% Completing an Advanced Course	0.1744 (0.0380)	0.2508 (0.0657)		
% Above Criterion on SAT/ACT	0.0530 (0.0432)	-0.0704 (0.0761)		
Prior Passing Rate (TAAS or TAKS)	0.0091 (0.0479)	0.0116 (0.0558)	0.0215 (0.0660)	0.0197 (0.0752
Average Monthly Salary Beginning Teachers (log)	0.4149 (0.0345)	0.3672 (0.0641)	0.3869 (0.0613)	0.4070 (0.0761
Average Monthly Salary Auxiliary Workers (log)	0.2595 (0.0193)	0.1844 (0.0387)	0.2247 (0.0438)	0.2189 (0.0420
District Enrollment (log)	-0.0795 (0.0032)	-0.0736 (0.0074)	-0.0809 (0.0077)	-0.0862 (0.0082
% Free and Reduced Lunch (non-highschool)	0.2401 (0.0182)	0.1831 (0.0361)	0.1790 (0.0410)	0.1682 (0.0440
% Limited English Proficient (non-highschool)	0.2253 (0.0433)	0.2572 (0.0842)	0.2928 (0.0814)	0.2853 (0.0846
% Less Severe Special Education	0.5056 (0.0822)	0.8037 (0.1648)	0.7251 (0.1781)	0.7256 (0.1971
% Highschool	0.6159 (0.0808)	0.2326 (0.1664)	0.3676 (0.1528)	0.2871 (0.1425
Miles to Nearest Metro Center (log)	0.0222 (0.0033)	0.0280 (0.0066)	0.0286 (0.0073)	0.0296 (0.0069
% Severe Disability Special Education	0.8598 (0.2011)	1.3051 (0.4192)	1.5847 (0.4681)	1.5997 (0.5099
Log Likelihood	2752.70	713.94	691.51	694.00
Observations	2755	714	726	726

Table 4. Distribution of GJTB Predicted Expenditures (in 2004$)*

Alternative Models	Correlation with predicted expenditures from Model A	Mean Predicted Expenditures	Standard Deviation of Predicted Expenditures	Minimum Predicted Expenditures	Maximum Predicted Expenditures
Model A	1.0000	6,283.06	561.94	5,252.07	14,832.45
Model B	0.9952	6,168.64	530.97	5,161.60	14,291.06
Model C	0.9971	6,270.18	562.42	5,210.97	14,154.71
Model D	0.8446	6,014.69	703.80	5,026.33	14,889.53
Model E	0.9718	6,336.16	538.30	5,391.94	13,534.03
Model F	0.9978	6,335.48	553.73	5,318.52	15,105.36
Model G	0.9998	6,294.06	551.27	5,281.84	14,749.30
Model H	0.9903	6,282.11	555.36	5,237.39	14,190.90
Model I	0.9364	6,015.34	552.60	5,047.81	11,730.79
Model J	0.9967	6,290.31	548.89	5,247.73	14,070.87
Model K	0.9428	6,477.51	505.07	5,419.63	16,077.38
Model L	0.9562	6,381.04	537.24	5,306.30	17,684.42
Model M	0.9512	6,444.52	538.67	5,346.61	17,531.71

*The predicted cost of average performance on the outcome measures in the model, weighted by 2004 enrollment.

Table 5. The Predicted Cost of Achieving 55-Percent Passing Rate (in 2004$) GJTB

Alternative Models	Correlation with Model A at 55-Percent Passing Rate	Mean	Standard Deviation	Minimum	Maximum
Predicted Cost per Student					
Model A	1.0000	6,389.23	617.00	5,252.07	14,832.45
Model J	0.9930	6,410.25	601.60	5,247.73	14,283.39
Model K	0.9337	6,483.18	507.72	5,419.63	16,077.38
Model L	0.9340	6,395.35	539.80	5,309.24	17,664.62
Additional Cost per Student					
Model A	1.0000	178.48	340.69	0.00	4,355.60
Model J	0.9868	188.67	343.78	0.00	4,566.02
Model K	0.9007	193.72	344.73	0.00	4,378.64
Model L	0.9156	175.40	335.72	0.00	5,563.13

*The predicted cost of performance on the outcome measures in the model, weighted by 2004 enrollment. There are 966 observations.

Table 6. The Predicted Cost of Achieving 55-Percent Passing Rate (in 2004$) I&R

Alternative Models	Correlation with I&R's IV Model	Mean	Standard Deviation	Minimum	Maximum
Additional Cost per Student					
IV Model	1.0000	395.09	444.00	0.00	3250.76
SFA weighted	0.9925	62.77	67.60	0.00	389.51
SFA unweighted	0.9887	20.43	21.77	0.00	123.55
OLS weighted	0.9909	64.18	68.89	0.00	402.74
OLS unweighted	0.9865	35.43	37.51	0.00	215.55

Additional cost predictions based on I&R method, comparing model forecast at observed outcomes with model forecast at 55 percent passing (or one third the gain in passing rate). Weighted by enrollment in 2004. There are 966 observations.