School Finance Redesign Project

center on reinventing public education

# DIFFICULTIES OF ESTIMATING THE COST OF ACHIEVING EDUCATION STANDARDS

**Susanna Loeb**
*Stanford University*

**Working Paper 23**
**August 6, 2007**

Daniel J. Evans School of Public Affairs
University of Washington

## The School Finance Redesign Project

The School Finance Redesign Project (SFRP) encompasses research, policy analysis, and public engagement activities that examine how K-12 finance can be redesigned to better support student performance. The project addresses the basic question, "How can resources help schools achieve the higher levels of student performance that state and national education standards now demand?"

Check in with us periodically to see what we're learning and how that information may re-shape education finance to make money matter for America's schools. You can find us at www.schoolfinanceredesign.org.

Jacob Adams, Principal Investigator

## The SFRP Working Paper Series

The Working Paper Series presents analyses that are complete but have not undergone peer review. The papers are subject to change and should be cited as working papers. Their purpose is to promote discussion and to solicit reactions.

## Support from the Bill & Melinda Gates Foundation

This work was supported by the School Finance Redesign Project at the University of Washington's Center on Reinventing Public Education through funding by the Bill & Melinda Gates Foundation, Grant No. 29252. The views expressed herein are those of the author and are not intended to represent the project, center, university, or foundation.

## Acknowledgements

I am especially grateful to Jacob Adams, Eric Hanushek, and Katharine Strunk for their comments and suggestions.

## Introduction

State governments have increasingly taken responsibility for funding K-12 schooling. When districts raise money locally, they can describe to voters what those tax dollars are going to fund. This clarity becomes more difficult at the state level unless the state either collects detailed information about resource allocation and disseminates it clearly or regulates how each district allocates its funds. Perhaps because of this, many states, including California, that do not have transparent revenue and expenditure systems do prescribe the use of many of the dollars given to districts. Yet, in the era of standards setting and assessment-based accountability, there is an interest in holding districts responsible not for their precise allocation of resources but for the outcomes of their students. This movement is anchored in the notion that with clear standards and aligned incentives schools will work towards meeting the goals set out for them. Because the optimal allocation of resources likely varies across districts and because local actors have the best information about their own students' needs, they may be in the best position to choose how resources are used—as long as they are working toward achieving shared goals.

In order to actually reach the goals set out for them, districts must have the resources as well as the intention to do so. Resources here can be viewed broadly to include not only dollars but also the capacities of administrators, teachers, and other staff to make the best use of the available funds. Schools and districts without these capacities may need to invest in professional development or more effective recruitment to build capacity to use resources well.

Court cases throughout the country have challenged, many successfully, the adequacy of resources available to districts for meeting state goals. Determining whether the dollars provided by the state are adequate requires a good estimate of the amount needed. For a number of reasons, such estimation is not an easy task. This paper addresses the difficulties with estimating resource needs, reviews methods used in this endeavor including their contributions and limitations, and concludes by emphasizing the importance of treating the results of these methodologies not as accurate estimates of costs but rather as one of many sources of imperfect information that together improve our dynamic understanding of resource needs and our ability to develop and implement effective policies and practices.

## The Difficulty of the Endeavor

Determining the dollars necessary to provide an adequate education is not an easy task. In particular:

- Our current state goals often are far higher than many of our current outcomes.
- Students vary substantially in their needs.
- The prices of education inputs differ across districts.
- Districts vary in current capacity.
- The data are poor.
- Any estimate is specific to the current state of knowledge and educational practice.

This section addresses each of these difficulties in turn.

First, as noted above, our current state goals often are far higher than our current outcomes. While it is appealing to think that our education systems used to be excellent, and now we must simply regain the quality that we had before, the evidence is clear that this is not the case. Test scores are at least as high as they were a generation ago, substantially higher for some groups. The reality is that we have set higher goals now than we have in years past, especially for children in poverty and in other traditionally low-achieving groups such as black students, Hispanic students, and special education students. At the same time, the importance of skills for later economic success has increased dramatically, meaning that the differences across groups in educational achievement and attainment have greater repercussions today than they did in the past. Because today's education systems are trying to do what yesterday's did not, they require different resources and educational approaches. However, without having met these new goals either through our current or previous systems of education, it is very difficult to know exactly what resources and approaches are needed. In turn, without knowing what resources are required, it is difficult to produce an accurate estimate of the cost of these necessary resources.

A second difficultly in estimating the cost of reaching outcome goals is that students vary substantially in their needs and costs. Clearly there is a meaningful difference, on average, in the resources required for students in poverty to meet the same goals as students from families with higher incomes. Students in poverty come to school with less prior preparation and have fewer resources at home to support their learning while in school. The interventions required to address these extra needs can be quite costly. For example, recent successes in some schools suggest that extended coherent instructional time can greatly benefit students from low-income families; this is an intervention that likely requires additional investment (David et al. 2006). Any reliable estimate of the cost of providing different kinds of students with an adequate education must identify the important drivers of variation in need and develop appropriate weights to account for the resultant differences in costs.

Not only can differences among students drive variation in the costs of education, but differences in the prices of educational resources can also result in cost differences. For example, consider teachers and administrators—arguably the most important education resources. Variation in local labor markets leads to variation in the salaries required to attract similarly skilled workers. Regions with higher wages in alternative occupations or, conversely, with few college graduates to draw on face higher prices for teachers and administrators. As another example, small or remote districts can face higher prices because they do not have the scale to buy in bulk or because the prices of goods include additional transportation costs. Again, any reliable estimate of costs must identify and account for important sources of variation in the price of inputs. Price indices are available for some goods in some areas—for example, Rose and Sengupta (2006) provide a cost index intended to equalize the purchasing power of California school districts in their acquisition of labor resources—but they are not readily available for all regions or all education inputs.

Districts and schools vary not only in their needs and prices, as described above, but also in their current capacities to transform resources into achievement. Districts and schools have different starting points in areas as basic as school facilities and infrastructure. Some schools and districts have new buildings with up-to-date science laboratories, computer systems, educational technologies, and the like. Other districts suffer from old and decrepit buildings and a lack of modern infrastructure. Perhaps even more importantly schools and districts have different personnel with diverse and often unequal abilities and knowledge. Cost estimates are likely to be

inaccurate if they are based on the assumption that all personnel are created and trained equally, such that a principal in one school is interchangeable with the principal in another. Human capacities differ across districts, allowing some districts greater ability to implement reforms and transform resources into achievement. The cost of transitioning to a school system that meets our educational goals requires understanding where each school and district is now and adjusting cost estimates to account for the current state.

How to implement policies that account for these differences in current capacities without creating incentives for districts and schools to appear less effective so as to get additional funds is not a straightforward task. Yet, this policy problem does not eliminate the fact that different schools and districts face different costs due to current capacities, and accurate cost estimates must account for these differences.

A fifth difficulty in assessing the costs of needed resources is that the data for doing so are often of poor quality. It is rare that we have good information on student learning linked to the resources that students received in a way that allows for accurate evaluation of the effects of those resources. As an extreme example, in California the data to assess the effects of expenditures on student outcomes are limited to average test scores at the school level and dollars per pupil at the district level. Average test scores (even student test scores at one point in time) are not a good measure of learning, and district dollars are not a good measure of expenditures on students aimed at improving learning. As data systems develop and revenues are distributed in ways that make evaluation more accurate (e.g., through random assignment), our understanding of the link between dollars and learning may improve; however, currently this link is very difficult to establish.

Finally, any estimate is specific to the current state of knowledge of education practice. Future innovations in curricula or instruction, for example, may reduce the cost of achieving a given education goal. While this might seem evident, it is central to understanding the usefulness of these estimates. As discussed above, we have yet to achieve many of the goals we have set, especially for students in poverty. As such, we look to innovation to help reach these goals—new curriculum, new instructional practices, new supplementary services. We focus not only on estimating the costs of providing education, but also the elements of education that make the spending of resources more efficient. Yet each innovation that works changes the current costs of reaching our goals and will require us to reexamine the costs of providing an adequate education to students.

Education practice here need not refer only to the technology of instruction. The governance of the education system can powerfully influence how effectively dollars are used to meet outcome goals. As an example, states vary in the extent to which they delegate education decisions to districts, and districts vary in the extent to which they delegate decisions to school personnel. Similarly, some education systems have more clearly defined goals and aligned systems of incentives than others, likely influencing the efficiency with which dollars are used to reach these specified goals, even given similar capacities at the school and district level. As governance systems change, reallocating authority and responsibility and changing incentives, so likely will the costs associated with reaching outcome goals. Education is a set of many dynamic systems, and cost estimates are best viewed as part of this dynamic system. Cost estimates provide information about the extent to which additional resources can help us move closer to our goals, and as we get closer to these goals our knowledge and our resultant cost estimates should change.

In summary, if we do not acknowledge the difficulties involved in estimation and the subsequent impact of these difficulties on the estimates, we are bound to misinterpret and misapply the results. That is not to say that there is nothing to be learned from such exercises. The following section reviews the four primary methods of estimating resources needed to reach given outcome goals, attempting to identify both their weaknesses and their contributions.

## Estimating Resource Needs

There are four methods typically used for estimating the cost of attaining a given student outcome:

- successful schools approach
- regression-based approaches
- evidence-based approach
- professional judgment approach

This section addresses each of these in turn.

### Successful Schools Approach

A common response to the question of how to best estimate the costs of improving student outcomes is, "Why not just look at schools that are doing well and see how much they spend and how they are spending this money?" This intuition is actualized in the successful schools or beating-the-odds approach. For this approach, analysts identify schools that achieve the chosen goals and then estimate the cost of providing the same services in all schools. This method has been implemented in many states, most often by Augenblick and Myers, Inc. and Augenblick, Palaich, and Associates (for examples, see the Augenblick et al. successful schools studies in Colorado, Illinois, Kansas, Louisiana, Maryland, Mississippi, Missouri, New Hampshire, and Ohio, among others). Each study uses different criteria to identify successful schools, but all recent studies attempt to account for differences across schools in the student population served, finding schools that are particularly successful for a given population of students. As Perez et al. (2007) point out, the appeal of this approach is that the concept is logical. It is also straightforward and relatively inexpensive to implement. Moreover, the successful schools concept is easily understood by policymakers, the Courts, and the general public, lending to its popularity.

**Shortcomings of this method include difficulties identifying successful schools and assumptions regarding scale-up.** While the appeal of the learning from successful schools is evident, this approach has two shortcomings—and they are important ones. First, it is not always easy to identify effective schools. Schools face different student populations and analysts may not be able to adjust adequately for these differences. Districts differ in their communities and their enrolled students in ways that are not easily measured nor adjusted for when identifying schools. One school may look as though it is doing better than a school with the same student population as measured by the percent eligible for free/reduced-price lunch and/or by the racial/ethnic composition of students but may in fact serve quite different populations. Community contexts as well as the different mechanisms by which students select schools can lead to different outcomes in ways that are outside of the control of the school, or the analyst. Perez et al. (2007), for example, adjust for the percentage of students that receive free/reduced-

price lunch, the percentage of English learners in the school, the percentage of English learners that speak Spanish, the percentage of students with disabilities, the percentage of parents that did not complete high school, the percentage of parents that have some college education, the percentage of parents that have a college degree, and the percentage of parents that have a graduate degree. These controls capture some but clearly not all of the student population differences that schools face.

The ease with which successful schools are identified is also hampered by the variation schools experience from year to year in the learning gains of their students. This is due in part to the differing strengths of particular cohorts of students and in part to random fluctuations of test score outcomes. For instance, a cohort of students in a given year may be particularly strong, which may lead to enhanced student performance in a given year. In order to identify consistently strong schools multiple years of learning gain data are necessary.[1]

Once accounting for differences in population and variation across years, it is not always clear whether there are schools that are consistently better, particularly at serving traditionally low-performing student populations. For example, in a study of Maryland successful schools, Augenblick and Myers (2001) attempt to account for the increased costs of teaching disadvantaged, special education, and limited English proficient (LEP) students, but because the "successful" schools in their sample tended not to have high proportions of these students, the method provided little information with which to adjust cost estimates to account for differences in student needs.

Second and perhaps even more importantly, successful school studies make the assumption that the schools can be scaled up at no additional cost. They assume that other schools could costlessly emulate these successful schools—and thus could reach the same outcomes with the same expenditures. This is not a reasonable assumption. Differences in outcomes across schools are driven at least in part by differences in the knowledge and ability of teachers and leaders. While highly effective individuals may move across schools, the current systems may not be able to produce enough of these exceptional individuals to provide all schools with such strong human resources.

**This method may also indicate little cost differences between successful and unsuccessful schools.** It is not unusual for successful schools studies to find little difference in expenditures between schools that beat the odds and those that do not. For example, Perez et al. (2007) find that of the group of high-poverty schools successful schools spend only $266 more per pupil, on average, than low-performing schools and conclude that this difference in spending was extremely unlikely to explain the differences in outcomes. Similarly, Augenblick et al. (2002) find that successful school districts in Kansas in the 2000-2001 school year were spending approximately six percent more than other school districts.

There are a few possible reasons why schools that beat the odds do not spend more than schools that do not. It could be that there is enough money in the system but that most schools

[1] Perez et al. (2007) define successful schools in California as those that in all four years perform at least 0.75 standard deviations higher in English Language Arts and mathematics overall and for subgroups based on free/reduced-price lunch participation, English learner status, Hispanic proportion, and African-American proportion. This criterion produced 61 elementary schools, 7 middle schools, and 35 high schools, excluding magnet schools and charter schools. Defining low performing schools as those that have a negative error term for all groups in all years, Perez et al. (2007) identifies 76 elementary schools, 32 middle schools, and 5 high schools.

are mismanaged and funds are simply wasted. If these schools would simply follow the lessons of the successful schools, they too would succeed. Similarly, it is possible that schools have sufficient funds to pursue their goals, but that most schools' goals are not aligned with the state-set goals as measured by the assessments, whereas "successful" schools' goals are those captured by the assessments. In this case, if most schools realigned their goals with those of the assessments, they would also be "successful."

It is unclear in this latter case whether it is beneficial for all "unsuccessful" schools that are actually quite successful at achieving their own specific goals to realign their goals with those of the state assessment. For example, the seemingly unsuccessful schools may be succeeding in other shared goals, but these goals may not be in the domain of outcomes measured by the assessments used in the study. These schools may excel in art, science, or school completion, while the study may only use math and reading scores to categorize schools. This possibility is one form of misclassification of schools as successful or not successful. However, while potentially an issue for a small number of schools, such misclassification is unlikely to completely explain the observed outcome differences across schools.

Another possibility, also a form of misclassification is that studies may not adequately control for differences across schools in their student population. The data available with which to adjust for cost and price differences are not ideal. For example, researchers rarely have access to data on the academic achievement of students as they enter the school. Without such measures, it is very difficult to establish the causal effect of schooling on student performance because researchers cannot fully control for students' prior achievement. Another particular concern here is the measure of poverty used by most researchers. Most studies use eligibility for subsidized lunch as the measure of poverty, but this measure is only a weak proxy. As Bryk (2007) demonstrates in his Chicago study, schools that share the same proportion of students eligible for these subsidies may see substantial differences in the income, employment, and education of their families. Simply using the proportion of students eligible for subsidized lunch doesn't control for the *severity* of poverty within a given school.

Perhaps most importantly, a final reason that successful schools studies may find little difference in expenditures between successful and unsuccessful schools is that the successful schools have gotten a good draw from the distribution of non-monetary resources. While not all successful schools studies explore why the identified schools are more successful, some of these studies do look more closely at these schools. The results, unfortunately, do not point to silver bullets for improving all schools. For example, Perez, et al. (2007) find that the allocation of resources explains little of the differences in achievement across schools though many beating-the-odds schools share three common factors: high quality teachers and staff, implementation of a standards-based curriculum. and coherent instruction. The authors also identify teacher support and training, school control over hiring, effectiveness in removing teachers, assessment data that informs instruction, high expectations for students, parental involvement, and teacher collaboration time as practices that appear to support these shared features. These findings imply that non-monetary resources such as strong leaders and teachers are central to a school's success. A poor teacher costs approximately the same as a great teacher, and a poor administrator costs approximately the same as a great administrator. However, the differences between poor and great teachers and administrators seem to be a large part of what is driving the differences between unsuccessful and successful schools. Unfortunately, we do not have the technology either to consistently identify or to replicate the most effective individuals.

Thus, while the achievement differences across all schools may not be associated with spending differences, the cost of replicating success may be great. Consider the extreme example in which the set of skilled leaders is fixed. Improving one school by bringing in a strong leader will then equally hurt another school. We may learn something about the cost of improvement through professional development or implementing curricula or new data systems from beating-the-odds schools. However, in practice we likely will need to expand the pool of highly effective teachers and leaders with innovation and investment outside of schools. The successful schools approach sheds little light on the best ways to do this or on the cost of these processes that occur outside of schools.

The limitation of using successful schools studies to inform estimates of resource needs does not imply that there is nothing to be gained from such studies. It is through studies of successful schools and successful districts that we can find practices and resource allocations that may be particularly effective. Additional instructional time and teacher collaboration time are examples of these effective practices. Subsequently, more causal analyses such as those built on random assignment designs can better assess the causal effects of these practices.

## Regression-Based Approaches

While the successful schools approach may not provide the information needed to estimate resource needs, there are other data on current schools that may be of more help with this. Regression-based approaches, often called "cost function analyses," use data on district spending and the educational achievement of students to determine the cost of reaching standards. In particular, these approaches use regression analysis to estimate the relationship between spending and student outcomes and then use these estimates to predict the amount of money each district needs to achieve a given outcome. In theory, this approach could be applied to the school level instead of to the district level; but, in practice, the school-level spending data that are needed to do such analysis are rarely available.

There are a number of difficulties in applying the cost function approach effectively, many of which align with the discussion of the Successful Schools approach above. The issues facing regression-based models are of two overarching types: technical problems that skilled analysts *with sufficient data* can correct in their models and conceptual problems that bring the overall approach into question.

**With sufficient data, skilled analysis can overcome technical problems**. Given sufficient data, a skilled analyst can adjust a regression-based model to produce accurate estimates of the relationship between resources and outcomes. One of the most important issues that must be taken into account is that districts differ in the prices they face for inputs and in the resource needs of their students for achieving a given outcome. For example, districts in areas with higher wages for college graduates need to spend more on teacher salaries to attract an equivalent pool of teachers—the price of teachers is higher. If districts facing higher prices spent more, on average, but not enough more to get equivalently skilled teachers as those in districts paying lower prices, it would appear that there is negative relationship between spending and teacher quality unless the analysis adjusted for the differential prices. Similarly, if one school serves more children in poverty, who, on average come to school with lower achievement levels, they may spend more and yet achieve lower results than a comparison district serving students from higher income families. Unless the regression analyses adjust for these cost differences, it will

indicate a negative relationship between spending and outcomes, which will not be a good estimate of the causal effect of spending. Most regression-based studies attempt to control for many prices and costs including labor market costs, student characteristics, and size (both geographic and number of students).

**Conceptual issues represent another challenge.** Even after adjusting for cost and price differences, there is still variation across districts in their outcomes for students, even in districts with the same expenditures. There are a number of reasons for these differences that draw the regression-based approaches into question. In particular, we have little way of knowing how much of these differences are driven by (1) unobserved cost or price differences, (2) efficiency or mismanagement, (3) district goals that are not within the domain of state goals, and (4) unobserved differences in district achievement of the full domain of state goals.

First, there may be unobserved cost differences across districts for which it is simply too difficult to adjust the estimates. For example, while the cost function may have controlled for poverty, as noted above, it is very difficult to obtain measures of poverty that sufficiently adjust for differential manifestations of poverty in a community. Similarly, districts that, in the past, have hired less-skilled teachers may now face higher costs, given their workforce. One district also may have particularly skilled community members helping in the schools, while another doesn't, allowing the former district to spend less on instructional aides. Thus, some of the differences observed in outcomes across districts may be due to inadequate adjustments for costs or prices.

In addition to inadequate adjustments for costs and prices, differences in outcomes may be driven simply by differences in districts' levels of "efficiency." Many regression-based analyses attempt to adjust for district differences in management effectiveness and in districts pursuing non-state goals (together these are often termed "efficiency"), but it is not an easy thing to do. It is difficult for analysts using any adequacy approach to completely assess whether or not a district and schools have weak or strong managers, but it is especially hard to ascertain differences in managerial skill levels by looking at the macro-level data. For example, one district may not utilize cost-saving approaches in their purchases or they may not do as good job of recruiting, selecting, training, and managing teachers. In the extreme, corruption may affect outcomes. Regression-based approaches have difficulty accounting for these issues of mismanagement due, at least in part, to lack of available data on resource use.

Another source of inefficiency stems from using resources towards goals that are not those deemed most important by the state. Consider a state that sets achievement goals in reading and math but a district that provides excellent instruction in music, encourages civic responsibility and/or responsible social interactions, or works towards a goal of all students being able to do back flips. This district's resources may be used effectively to reach their own goals, but ineffectively to reach the goals that have been set by the state. From the state perspective, this district is inefficient.

However, not all goals espoused by states are measured by the state assessments that are used as student outcomes measures in regression analyses. Because of this limitation, some districts that are actually achieving unmeasured state goals may appear to have poor outcome results in a regression-based approach that uses information on only measured state goals. Starting with the above example, assume the state changed its goals to include music, but did not measure this on

its assessments of student learning. The district, which is very good at music, but not as good in the measured outcomes, would appear less effective than it, in fact, is at achieving state goals.

Researchers have used a variety of statistical techniques to attempt to separate mismanagement and non-state goals from the other sources of difference in measured outcomes such as cost differences, price differences, and unmeasured outcomes for students. Adjusting for these two sources of inefficiency is important when estimating the cost of achieving a given outcome because without this adjustment additional spending by districts would appear to signal higher costs when it was actually just misdirected dollars. The first method for making this adjustment is to simply include controls in the regression model for efficiency. This way we can ask about needed dollars "holding efficiency constant." This is not a particularly compelling approach because we do not tend to have good measures of efficiency.[2] A second approach, which can be used in combination with direct controls for efficiency, is to use a statistical technique called Instrumental Variables (IV). This approach aims to adjust for possible unmeasured cost and price differences that are not within the district's control and that bias the estimated effects of outcomes on resource needs.[3] Unfortunately, the IV technique also is not likely to be completely effective because instruments are hard to find—namely, it is difficult to find variables that strongly predict a district's achievement that do not also predict their spending.

In addition to these limitations, regression-based approaches face another important problem: the results can be extremely sensitive to the specification of the regression. The form of regression analysis called "cost functions" estimates the effect of achieving higher outcomes on spending; that is, spending is modeled as a function of student outcomes. However, it makes just as much sense theoretically, arguably more sense, to model student outcomes as a function of spending. This second approach is called a "production function." In certain cases, ordinary least squares regression will give similar results when estimating the effect of spending on achievement as when estimating the effects of achievement on spending. However, in most cases, it will not.

In practice, the difference between the cost and the production function approach can be dramatic. For example, Imazeki (2007) finds that the cost function approach in which spending is assumed to be a function of outcomes leads to an estimate that California would need to spend an additional $1.7 to $5.7 billion to reach current state goals for student performance. The alternative approach, in which outcomes are modeled as a function of spending, leads to a predicted need of $1.5 trillion. She uses an example of a district that is spending $8,000 and has an API of 750. Using the cost function estimates, she finds that a fifty-point increase in API (to

[2] For example, in a recent study, Jennifer Imazeki (2006) uses a measure of competitive local markets as a proxy for efficiency. She assumes that districts will be more efficient in their use of resources if there are other districts in the area that residents can choose to live. Duncombe and Yinger (2007) use variables in their regression analyses that adjust for incentives that lead voters to monitor school officials and incentives directly on school officials themselves.

[3] For example, if the pool of potential leaders for a given district is stronger, we might actually see both higher achievement and lower spending both driven by this third, unobserved variable. In theory the IV method gets around this problem by predicting student outcomes based only on measured factors that don't influence spending except to the extent that they affect outcomes. For example, instead of using the district's actual outcomes, it uses the outcomes of surrounding districts as well as the characteristics of surrounding districts. These measures can predict district student outcomes and create a proxy for student outcomes that does not depend on the unmeasured cost or price differences specific to that district.

800) would require an increase of $181 per pupil. Using the production function estimates, the same fifty-point increase in API would require an increase of $11,600 per pupil. As Imazeki points out in her paper, these are clearly huge differences that draw into question the usefulness of the regression-based approach to assessing spending needs.

Close examination of regression-based estimates of the cost of achieving given goals illuminates serious shortcomings of this approach. But, while at this time analysts may not be able to use this method to produce reliable estimates of the effect of resources on student learning, this does not imply that under different circumstances this method would be without merit or contribution. If analysts had access to data that linked resources to students and followed students over time so as to measure student learning and if resources were allocated in such a way as to make evaluation more reliable (such as randomly staggered implementation), then researchers could more accurately estimate the effects of resources on student outcomes and use those estimates to predict the cost of needed resources. Unfortunately, in current education systems, such information is rarely available.

Both the regression-based approach and the successful schools approach rely on the performance of schools in the current system to estimate the dollars needed to reach a given a standard. In the regression approach there are very little adjustments for the effectiveness of resource allocation, assuming that most schools could achieve the gains of the average school given the student composition. This is an over-simplification of this approach but not far from an accurate characterization. In the successful schools approach for estimating the cost of reaching a goal, the assumption is that all schools could reach those outcomes with the resources that currently successful schools have. This is unlikely to be the case due to the difficulties of scaling up successful programs. Because of the limitations of using these data-based approaches for estimating resource needs, states have turned to alternative methods. The two other approaches do not rely on current schools or the current distribution of resources and outcomes. Instead, they design prototype schools based on research evidence and professional experience and then calculate the costs of such schools.

## Evidenced-Based Approach

In theory, this method creates prototype schools based on research evidence of the effectiveness of resources at improving student outcomes. In practice, the research base is not strong enough to support a full school design. Instead, this approach usually starts with the current school system and then estimates the cost of supplementing the current system based on the resources that appear to be effective. For example, given the evidence from Tennessee on the effects of smaller class sizes in early elementary school, the analysts might estimate the cost of reducing the class size in those grades. In Arkansas, Picus and Associates (2003) used research on the effectiveness of pre-school, full-day kindergarten, school size, curriculum offerings in secondary schools, class size, principals, school-based coaches, preparation time, professional development, struggling student strategies, gifted and talented programs, student support services, technology, and teacher salaries to inform their estimates of necessary resources to achieve an adequate education spending level. Sometimes, as in the Arkansas study, this approach also uses a group of teachers and administrators to review and discuss the proposed school reform, in keeping with the professional judgment approach described below. The cost of the reform is then estimated.

This method, unlike the regression-based methods, assumes spending on pre-determined resources, such as class size reduction, or a particular curricular or instructional intervention. Analysts assume set costs for the "research-based" education strategies and then decide what adequate levels of these resources should be provided to students. They then aggregate the costs of these reforms to produce estimates of increased spending necessary at the state, district, and school levels.

The primary drawback of this method is that the research base is not strong enough to support it; we simply do not have enough conclusive evidence about educational interventions to justify necessary expenditures. For example, the research base on the effectiveness of *every* one of the resources listed above that served as a basis for the prototype school in Arkansas is questionable. Most of the research is questionable because it doesn't do a good job of establishing the causal effect of the resources; that is, it doesn't account for the initial differences between students with and without access to the resources.

There are a number of factors that contribute to our lack of information regarding successful education reforms. The first, as noted earlier in this paper, is that some goals are so far above the current state of performance that we do not have examples of achieving the goals we set out. It is difficult to predict the utility of additional resources when the outcomes in question are never observed. The second factor is the lack of data available with which resources can be evaluated. Some states—North Carolina, Texas, and, particularly, Florida—have developed strong databases, but many other states have not. Third, even with base data as is available in Florida, it is difficult to estimate resource effects if schools select their resources. Schools with higher achieving students, for example, may select different resources, and researchers would not be able to distinguish selection from effectiveness.

State-led experiments—for example, on subject matter pedagogy coaching, planning time for teachers, and specific curriculum implementation—would help provide some of the evidence needed for this approach to be feasible. However, we must be careful to consider any unintended consequences of reforms when we utilize evidence ascertained via small-scale experiments. California's response to the Tennessee Class Size experiment provides a good example of why. The Tennessee class size experiment showed, overall, that smaller classes lead to increased student performance. In response, California instituted the Class Size Reduction Initiative, providing substantial monetary incentives to districts that lowered their classes in kindergarten through third grades to 20 students per class. In the relatively small-scaled Tennessee experiment, reducing class sizes did not necessitate hiring substantially more teachers. However, in California, this was the result. Because the supply of quality teachers in California was not simultaneously increased as class sizes were lowered, it is possible that California hired teachers of lower quality to teach smaller classes, possibly negating the benefits of reduced class sizes. A slower implementation of this policy may have been able to realize the potential benefits of class-size reduction without these transitional costs.

In summary, the evidence-based approach to estimating resource needs requires a thorough understanding of the effects of resources on student outcomes. We do not currently have that understanding. New technologies allow for the collection of much more detailed information than were possible just a few years ago. This new information collection combined with purposeful policy implementation, such as pilot programs with random assignment, may dramatically improve our ability to implement this approach in the future. Yet, even with improved information, the diversity of student needs, the resulting importance of context in the

effectiveness of any resource, and the dynamic nature of the education system are so great that research, and information more broadly, are more likely to contribute to effective schooling as formative assessments, giving us a better sense of what to do, than as definitive (summative) measures of the effectiveness of any given resource.

## Professional Judgment Approach

The final and most common method for estimating costs is the professional judgment approach. There are two forms of this: professional judgment panels and a new professional judgment budget-simulation or survey approach. In professional judgment panels (see Chambers et al. 2007), respected educators gather and reach consensus on the resources they think are required for different kinds of schools and districts to achieve a given adequacy standard (or standards). They design prototypical classrooms, schools, and districts, and then researchers estimate the costs of implementing the schools across the state. These educators draw solely on their own best judgments and experiences. In contrast, the survey approach (see Sonstelie 2007) samples a random group of educators—teachers, principals, and superintendents—and asks respondents to allocate resources to achieve the best student outcomes. Respondents then estimate the outcomes they would expect for students in a school with such resources.

Like the evidence-based approach, but unlike the regression-based approach, the professional judgment method assumes that resources will be used for specific purposes. The equivalent dollars simply inserted in the current system would not be expected to achieve the same results. The professional judgment panel assumes that the resources would be spent as specified. The professional judgment survey assumes that resource allocation would be determined entirely by unconstrained local actors.

Both of the professional judgment approaches rely on professional educators. There are clearly drawbacks to reliance. First, educators benefit from increased resources for schools and thus have an incentive to overestimate the resource needs associated with a given outcome. In addition, because these approaches draw solely on the knowledge of current educators, they may not be as innovative as they would be if their experiences were broader and thus they may overlook important and effective options for schools. Finally, neither method allows for an empirical test of whether the resources described actually would achieve the outcomes in questions.

There are additional pitfalls inherent to the traditional panel approach that are better addressed by the newer professional judgment survey approach. In the traditional approach, expert panels are convened and asked to determine necessary resources given an *unconstrained budget*. Rather than make strategic decisions about what is *needed* to provide adequate education expenditures, participants can easily slide into what is *wanted.* This is an important distinction. An unlimited budget may lead panel participants to add in resources that have little true benefit. While within a panel, members may learn from each other and thus develop a more effective school, they may also bargain in unproductive ways, especially because their budgets are not constrained: “I like librarians and you like guidance counselors, I’ll agree to include more guidance counselors if you agree to include more librarians.” The newer budget-simulation method restricts the budgets and asks respondents to allocate resources within those given budgets; this change in task is likely to reduce the tendency to include resources that are not effective.

Traditional professional judgment panels are also limited in the number of different outcome goals for which they can estimate costs. It would be too time consuming to ask panels to develop separate school prototypes for each of multiple potential outcome goals. Most professional judgment panel studies identify one outcome goal and then design multiple schools each aiming to achieve that single outcome goal for a different population of students. Because the studies are choosing a single goal, they usually choose one corresponding to a state's stated goal, which is often far higher than current outcomes. Because of the large gap between the goal and current outcomes, the estimates of needs for these goals can be quite inaccurate. The budget-simulation approach to professional judgment estimation avoids this problem because it has many more respondents and it can ask each respondent to allocate resources under three possible budgets (low, medium, and high). Thus, the newer approach estimates resource needs for a variety goals, many of which are closer to current outcomes and thus can better draw on the expertise of participants.

This is not to say that the traditional panel approach does not have its benefits. For instance, the panel approach relies on highly respected experts, while the surveys are completed by a random sample of educators. Not all survey respondents are likely to be skilled in resource allocation. In addition, panel members may learn from each other. Each may bring a different area of expertise to the table, and together they may design more effective schools than any participant would on his or her own. However, the modifications inherent in the survey approach—limiting budgets and addressing a variety of goals, some of which are closer to current outcomes and thus better draw on the expertise of practitioners—may be more important than the potential advantages of the panel approach.

Two recent studies in California are the first to compare the older expert panel approach with the newer budget-simulation approach. The results are similar in many respects, though the simulation approach (Sonstelie 2007) gives cost estimates for a greater range of outcomes and finds a stronger relationship between student poverty and resource needs. Using the responses of 568 teachers, principals, and superintendents, Sonstelie (2007) develops equations for predicting resource needs at the school level in California given a school's student population. He estimates, for example, that the average elementary school with 573 students and a budget of $4,000 per student at the school-level, about average for the state, could, on average, achieve a state test outcome of 843, well above the state's standard of 800, if none of the students were poor and an outcome 698 if all the students were poor. An increase in the school's budget of $1,000 per pupil increases this prediction, but only by 13 points. These estimates are quite imprecise, even as estimates of the potential responses of the full population of teachers, principals, and superintendents. For the average elementary school, the school in which 52 percent of students participate in the subsidized lunch program, the estimated budget is $7,430 per pupil and the 90 percent confidence interval runs from $6,403 per pupil to $8,368 per pupil.

Given the imprecision of even the more defensible approach to utilizing professional judgment in combination with the potential biases discussed above, it is not unreasonable to ask why we would trust estimates provided by either professional judgment approach. If the goal is accurately estimate a specific dollar value that all schools should receive to guarantee that they can reach a given outcome goal, then this method is going to fail, as would the other methods described above. If instead, the goal is to use the best information available to refine current policies in a dynamic system, then professional judgment, combined with evidence from high

quality evaluations of the effects of resources, are the best and, in many cases the only, sources of information available.

## Conclusion

States' recent efforts to develop standards of achievement for students and to align their curriculum and assessments to those standards naturally leads to a desire to align the school finance system to these same standards, guaranteeing that schools and students have access to resources that allow them to achieve the chosen goals. However, the appeal of linking finance to standards and outcomes does not alleviate the difficulty in actually implementing such a system.

None of the methods described above accurately estimate resource needs. The successful schools approach does not consistently identify "successful" schools or provide information needed on how to scale up success (e.g., by expanding the supply of effective leaders and teachers). The regression-based approach is limited by poor data in most states and by resource allocation that makes evaluation of the effects of resources very difficult. The evidence-based approach does not have adequate evidence to draw on, and the professional judgment methods are subject to incentive bias.

Nonetheless, oddly, the inaccuracy of the estimates may not be the fundamental difficulty in accurately estimating resource needs. When assessing the adequacy (or sufficiency) of a school finance system for meeting goals set forward by the state, it is necessary to ask more than simply, "How much money do we need to meet state goals?" Even after clearly defining the goals—a process that in itself is difficult—such an analysis requires us to answer a number of other questions. "Do we have the technology for reaching the stated goals? Do the current institutional structures allow the use of the necessary technology? And, can we alter institutions to allow us to make better use of resources and/or develop new and more effective technologies (e.g., strategies, curriculum, instructional techniques) for reaching our goals?" Estimates of the cost of reaching goals inherently make assumptions about these questions. Yet, they are very difficult questions to answer, leading to substantial uncertainty.

Because of the inherent link between institutions, technologies, and costs, any estimate of cost that we produce can only be expected to be error-prone and short-lived. This is not to say that estimates are not needed. Policies are never made with perfect knowledge of their effects. In the best cases, they are made based on mass of information that provides insight into the likelihood of potential benefits and costs. School funding is no different. We have no method to accurately estimate resource needs, yet we need to use the evidence that we can compile to help with these decisions. Successful schools are one source of information; quantitative data for regression analyses is another. Professional judgment and research evidence are yet other sources. The specific nature of the results of these studies should not be mistaken for the accuracy of the predictions, but the inaccuracy of the predictions should also not be mistaken for complete lack of useful information.

## References

Augenblick J., and J. Myers, Inc. 2001. *Calculation of the Cost of an Adequate Education in Maryland in 1999-2000 Using Two Different Analytic Approaches.* Prepared for the Maryland Commission on Education Finance, Equity and Excellence (Thornton Commission).

Augenblick, J., J. Myers, J. Silverstein, and A. Barkis. 2002. *Calculating the Cost of a Suitable Education in Kansas in 2000-2001 Using Two Different Analytical Approaches.* Prepared for the Kansas Legislative Coordinating Council. Augenblick and Myers, Inc.

Bryk, A. 2007. *The Influences of Community Context on School Improvement*. Working Paper. Stanford University.

Chambers, J., J. Levin, and D. DeLancey. 2007. *Efficiency and Adequacy in California School Finance: A Professional Judgment Approach*. Getting Down to Facts Project: Stanford University.

David, J. L., K. Woodworth, E. Grant, R. Guha, A. Lopez-Torkos, and V. M. Young. 2006. *Bay Area KIPP Schools: A Study of Early Implementation First Year Report 2004–05.* SRI International: Menlo Park, CA.

Duncombe, W., and J. Yinger. 2007. *Understanding the Incentives in California's Education Finance System*. Getting Down to Facts Project: Stanford University.

Imazeki, J. 2007. *Assessing the Costs of K-12 Education in California Public Schools*. Getting Down to Facts Project: Stanford University.

Perez, M., P. Anand, C. Speroni, T. Parrish, P. Esra, M. Socias, and P. Gubbins. 2007. *Successful California Schools in the Context of Educational Adequacy.* Getting Down to Facts Project: Stanford University.

Picus, L.O., and Associates. 2003. *An Evidenced-Based Approach to School Finance Adequacy in Arkansas.* Lawrence O. Picus and Associates.

Rose, H., and R. Sengupta. 2006. *Teacher Compensation and Local Labor Market Conditions in California: Implications for School Funding.* Getting Down to Facts Project: Stanford University.

Sonstelie, J. 2007. *Aligning School Finance With Academic Standards: A Weighted-Student Formula Based on a Survey of Practitioners.* Getting Down to Facts Project: Stanford University.