Discussion

Do State Governments Matter?

Julian R. Betts*

Tom Downes is ambitious in his goal to summarize the impact of three distinct types of state policy changes on the ability of school systems to "equalize opportunity" across students of varying socioeconomic backgrounds. The policy reforms he considers are finance reforms emanating from court challenges to individual states' school finance systems, reforms deriving from state tax and expenditure limitations, and the advent of charter schools as a publicly funded alternative to regular public schools.

To those not familiar with the debate on whether educational spending matters, it is worth mentioning that the works that Downes reviews on court decisions and voter tax limitations are of great interest to economists studying public education. One of the most important areas in education economics is the extent to which changes in school funding cause changes in outcomes such as test scores, graduation rates, college attendance, and earnings of students years after graduation. A large body of literature examines the relationship between school resources and student achievement, years of schooling completed, and earnings after leaving school.[1] But does a positive correlation between school expenditures and student outcomes necessarily imply causation? There are many reasons to think not. Most obviously, in the United States today, students of lower socioeconomic background still typically attend schools with lower levels of resources, particularly when resources are measured by teacher qualifications such as credentials, years of experi-

* Professor of Economics, University of California, San Diego.

[1] For a review of the test-score literature, see Hanushek (1996); for a review of the relation between school resources and years of schooling and earnings, see Betts (1996).

ence, and education. If researchers find that disadvantaged students have both poorer educational outcomes and fewer resources at school, it certainly could signal that resources do matter. But the correlation could equally well be spurious. For instance, it could be that the true reason that disadvantaged students tend to have poorer educational outcomes is that they receive fewer educational resources in the home, fewer supports among the family, and fewer highly educated role models in the local community. In this instance, the positive correlation between school resources and student outcomes is merely that: a correlation induced by imperfectly measured variations in student socioeconomic status.

Of course, economists do more than look at simple correlations. Regression analysis attempts to control for all factors that may affect the dependent variable, in this case student outcomes. But we lack data sources that include rich measures of the educational supports in the home and community that I listed above. It is unlikely that commonly available measures of socioeconomic status, such as parental education and eligibility for school lunch assistance, fully capture variations in these factors. Hence, even careful regression analysis might overstate the impact of school resources on student outcomes because both variables are positively correlated with imperfectly observed family and community resources.

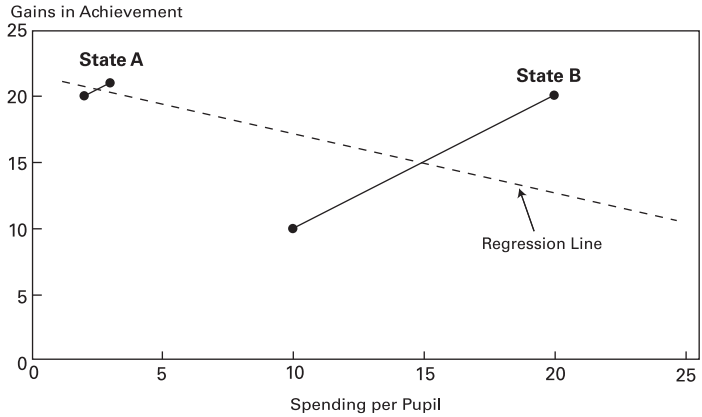
Conversely, one could argue that the many attempts by federal, state, and local governments to provide compensatory educational aid to schools in impoverished neighborhoods could induce a negative correlation between school resources and socioeconomic status. This, in turn, could induce a negative correlation between student outcomes and school resources that again is not causal, but merely reflects the correlations between both these variables and student disadvantage.

On the whole, I find the first of these arguments more persuasive, as most of our rather imperfect measures of student socioeconomic status tell us that there is still a predominantly positive relationship between socioeconomic status and the level of school resources that a student receives. In addition, few can doubt that socioeconomic status is a powerful determinant of cognitive development in children. For instance, a recent Educational Testing Service (ETS) study (Coley 2002) finds extremely large gaps in various measures of academic achievement between students with low and high socioeconomic status at the *start* of kindergarten. This surely speaks to the major contributions of home and neighborhood to early cognitive development.

Court-Mandated Educational Reforms and Tax and Expenditure Limitations

I would now like to discuss the court-mandated educational reforms and the tax and expenditure limitations that Downes reviews. Both of

Figure 1
Hypothetical Example of Incorrect Inferences about the Value of Spending per Pupil for Student Learning, Caused by Unobserved Variations in School Quality



these literatures attempt to reduce problems of endogenous school spending and omitted variable bias by seeking exogenous sources of variation in the resources that a school receives. In the case of court cases, one can argue that both the launching of a school finance lawsuit and especially the timing of its resolution are not caused by unobservable demographic or other personal attributes of state residents, or by any other characteristics of the state that could be causally related to student outcomes and school resources. If this assumption holds, then social scientists can perform before-and-after analyses of student outcomes that can potentially provide unbiased estimates of the impact of changes in school finances on student performance. In the more sophisticated approach that has quickly come to the fore, economists instead perform "difference-in-difference" analyses that compare changes in student outcomes over time in states that have undergone court-mandated finance reforms to changes in states that have not been subject to court mandates. This approach effectively takes account of national trends in the underlying variables and unobserved and constant characteristics of each state.

To those not familiar with difference-in-difference models, a simple example may be helpful. Figure 1 shows average annual gains in students' test scores plotted against spending per pupil in two hypothetical states, for two different years. The state in the upper left of the figure habitually spends less on schools but has higher rates of student learning,

perhaps because of some other unobserved factors affecting both variables. (In this hypothetical world, perhaps fiscally conservative parents not only vote to spend less on schools but also read more to their children at home!)

Let us suppose that court decisions in both states have caused spending per pupil to rise slightly, which in turn has quite literally *caused* student learning rates to increase in both state A and state B, as shown. But linear regression would not detect these causal effects. As shown by the dotted line representing the fitted regression line in Figure 1, we obtain the "wrong" result. There appears to be a negative relation between spending per pupil and student learning because the between-state differences in spending and learning completely dominate the visible but small effects of increases in spending per pupil in each state.

The difference-in-difference estimation strategy solves this problem by comparing changes in one state to changes in another. Social scientists typically estimate these models by expanding the list of explanatory variables from spending per pupil alone to also include a set of dummy (0,1) variables for states.[2] It can be shown that this is equivalent to subtracting the state mean from both gains in achievement and spending per pupil from each observation, and then running a linear regression using these "de-meaned" variables. Figure 2 illustrates what happens when we subtract the state means in this way. The changes over time in states A and B now line up perfectly along a positively sloped line. When we estimate a linear regression on these transformed data, we correctly estimate a positive causal relation between spending and learning. The trick in this analysis is to disregard all of the between-state variation, and instead focus only on the within-state variation.

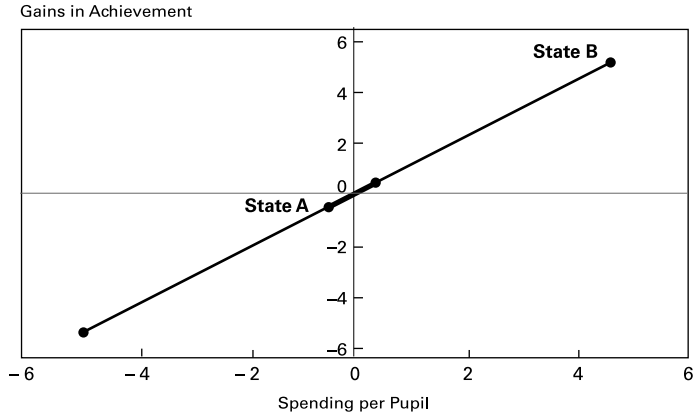
The tax-limitation and expenditure-limitation literature works on a similar premise: If voters pass these limitations for reasons that are not related to student outcomes in the state, then economists often consider the resulting reduction in school spending as occurring exogenously with respect to student outcomes. What we have, in both cases, is a natural experiment in which some outside or exogenous force has induced a change in school finance.

Downes reviews these literatures with care. He correctly concludes that the existing literature on court-mandated school finance changes has yet to deliver a consistent message about the impact on either the level or the distribution of student outcomes. The tax-limit and spending-limit literature provides slightly more definitive results suggesting that mean performance may fall if spending per pupil drops because of limitations.

Despite these methodological advances, using state-level variation in

[2] To keep the analysis simple, for this example, I will ignore the additional control or controls for time trends that researchers typically employ.

Figure 2
Hypothetical Example of Correct Inferences about the Value of Spending per Pupil for Student Learning, after Taking into Account Unobserved Differences between States



court decisions or tax limits carries certain risks. The crucial assumptions here are that court rulings on education finance and voter passage of tax or expenditure limitations occur in ways that are exogenous with respect to student outcomes. One can imagine scenarios in which either type of event occurred endogenously with respect to school quality. For instance, suppose that lower-income parents in one state become increasingly concerned about the quality of public schooling. This increased concern could manifest itself in several ways, for instance, in increased parental involvement in schools, which might improve student outcomes in these less-affluent areas. At the same time increased parental concern could lead to a lawsuit seeking to equalize school spending between have and have-not districts. If the court case is successful, these two events would lead—separately—to an increase in test scores in disadvantaged districts and an increase in school spending in the very same districts. Although a difference-in-difference analysis would lead us to infer that increased school spending had improved test scores, in reality, both changes would have been caused by something quite different—increased parental activism in the have-not districts.

A weakness in the above argument is that it ignores the fact that legal challenges to states' systems of education finance can typically take years and in some cases decades to draw to a final conclusion. This would make

the timing of the increase in test scores and the court-ordered change in spending less coincident.

Another example, this one related to the passage of tax or expenditure limitations, is that voters are more likely to support such limitations if they come to believe that state and local governments are not spending current tax revenues effectively. One event that could spur such a belief among voters is a downward trend (or stagnation) in student achievement, in spite of recent increases in spending per pupil. (Such increases in spending have been the norm over the last half-century.) This leads us into a situation of reverse causation, in which a decline or stagnation in student achievement causes the tax limitation measure to pass. It is not hard to see how even a very careful researcher might misconstrue this correlation as meaning that the new tax limitation had caused test scores to decline. Only by carefully removing ongoing trends in both variables can the researcher hope to obtain the correct inference.[3] The underlying issue in this second example is that the primary identification approach used in both literatures, the difference-in-difference method, is prone to error because before-and-after analyses can mistakenly attribute differing trends in different states to the change in policy. My goal here is not to dismiss the literatures that exploit the apparent exogeneity of court orders and tax limitations. On the contrary, they represent important developments in the broader literature on the determinants of school quality. Rather, my goal is to caution that the research and education policy communities would be wrong to treat either approach as a panacea.

Downes provides a careful and evenhanded summary of the findings that emerge from the court-mandate and the tax- and expenditure-limitation literatures. The results vary across data sets and the specific techniques used, which in part may reflect occasional violations of the assumptions underlying the natural experiments that these papers study. Overall, the body of work summarized by Downes suggests that changes in school spending are related to student outcomes in the expected direction, although the court-mandate literature is murkier in this regard than the tax- and expenditure-limitation work. My own reading of these papers is that the effects are modest, in the sense that complete equalization of school funding would go only part way toward equalizing student achievement.

The related literature on the impact of school resources on the earnings of students in the years after graduation points in the same direction. Betts and Roemer (2001) use the National Longitudinal Survey of Young Men to address the question of the extent to which educational

[3] Downes and Figlio (2000) represent a good attempt to tackle this specific possibility head on, and more work of this nature needs to be done.

funds would have to be reallocated across students to equalize opportunity across groups, defined, following Roemer (1998), as equalizing wages in an average sense. We find that equalizing spending per pupil, for instance between black and white students, would do virtually nothing to equalize the black–white wage gap years after the students had left school. Rather, spending per pupil would have to be many times larger for black students if policymakers wanted to take a significant chunk out of the black–white wage gap.

My final observation on Downes's summaries of the court-mandate and the tax- and expenditure-limitation literatures is that, as a central contributor to these literatures, he has written an extremely balanced review that points out the limitations not only of others' work but of his own. This is a model for others to follow.

Charter Schools

The third avenue of research reviewed by Downes is the advent of charter schools as an alternative to regular public schools. He asks whether students attending charter schools increase their rate of learning once enrolled, and the more difficult question of whether the advent of charter schools as a competitive force has induced regular public schools to improve.

Downes discusses two recent evaluations of charter schools in Texas and Arizona. These evaluations suggest a first-year slump for students enrolled in charter schools, followed by improvements for at least some charter school students in later years. This finding is of great importance given that school districts typically place charter schools under the accountability microscope practically from day one of their establishment. It will be important to see whether these dynamics can be replicated in other states. If so, administrators should be apprised of these patterns in order to avoid over-reacting to initial results at startup charter schools. At present, the results are not sufficiently solid for us to know for sure. (For a critique of the Arizona study, see Nelson and Hollenbeck 2001.)

On the question of whether the establishment of charter schools creates competitive pressures that spur nearby regular public schools to improve, Downes discusses the Michigan work of Bettinger (1999) at some length. Again, his review is on target in that data limitations restrict what we can know with certainty. The tentative conclusion from Bettinger's work is that we lack evidence of competitive pressures that improve student achievement at public schools located near charter schools.

However, another study, Hoxby (2002), finds statistically significant evidence that test scores, relative to spending per pupil, rise significantly in districts in which charter schools come to represent 6 percent or more of student enrollment. Hoxby uses a difference-in-difference approach, as

do many authors in the two aforementioned literatures. As I argued earlier, such approaches are susceptible to error if there are differences in the trends in student achievement among schools that are, by happenstance, correlated with the enrollment share of charter schools in the local district. To her credit, Hoxby successfully replicates her results by testing for a change in the *trend in gains* in school productivity after charter schools become a significant competitive force.

A key problem that remains, however, is that we do not know why charter schools become commonplace in some districts, yet remain so rare in other districts. It is quite easy to think of circumstances that would bias the estimated impact of charter schools on regular public schools up or down. For instance, suppose that one of the many omitted variables in existing analyses is the quality of district leadership and its openness to change. Suppose that a district hires a new, reform-minded superintendent, who simultaneously implements meaningful reforms in the public schools and, as part of the package, increases the number of charter schools. Even if charter schools had no real impact on the quality of regular public schools, a positive correlation between the number of charter schools in the district and public school productivity would result, and this, again, would not be causal. Even Hoxby's useful de-trended difference-in-difference approach would not capture the true causal relationships in such an instance.

Summing Up: What We Have Learned and What We Still Don't Know

While work on the question of charter schools' impact on regular public schools is still in its infancy, we have already learned important lessons. To date, little evidence supports those who warned that charter schools would be an educational disaster, as Downes points out. But we do not have as much positive to say as proponents of charter schools might like. Hoxby's work provides the strongest evidence to date that charter schools might have a positive competitive effect, even if it occurs after a threshold point has been reached. It also provides an intelligent check on the standard difference-in-difference approach.

To applied economists, the court-mandate and tax-limit and expenditure-limit literatures offer important examples of attempts to find exogenous sources of variation in an explanatory variable (in this case, spending per pupil) with the ultimate goal of unearthing the true causal impact of that variable on the outcome of interest (in this case, student achievement). Much of the literature has adopted the difference-in-difference approach, which in essence compares changes in outcomes in states (or districts) that have undergone a policy shock (such as a tax limitation) with changes in states that have not experienced the shock, all

the while removing fixed characteristics of each state and common trends that occur in all states equally.

The difference-in-difference approach has proven valuable, but it puts us at some risk of attributing changes in one state to the given policy shock when, in fact, another policy innovation or perhaps a demographic shock, imperfectly measured in the researchers' data, was in truth responsible for the change in student outcomes. A specific example of this is when an omitted variable "causes" both the change in student achievement and the change in policy, where the policy change could be either a court decision, a tax or expenditure limit, or the creation of a charter school. In my opinion, this issue is most severe in the charter school literature, where the existence of stark differences across districts in the rate of creation of charter schools suggests an underlying cause, perhaps related to changes in attitudes of the district administration or of local voters. A second risk is that the standard difference-in-difference approach misinterprets variations in trends across states or districts as being caused by the policy change in certain states. As I have noted, some researchers have started to find approaches that at least partially take these concerns into account.

Apart from his review of charter school studies that typically use a district-level or school-by-school analysis, Downes concentrates on lessons from natural experiments at the state level. Readers of the court-mandate and tax and expenditure literatures should be particularly concerned that—at this high level of aggregation—state fixed effects do not do enough to control for unobserved variations among states that, contrary to the assumptions of difference-in-difference work, are not always fixed. Furthermore, the problem of endogeneity (which reforms occur in which jurisdiction) does not disappear at the state level.[4] For these reasons, it will be important to supplement the state-level literatures on supposedly exogenous policy changes with similar analyses at the district level in order to check for consistency.

With these qualifications in mind, we have learned a great deal from all three bodies of literature, in particular from the tax- and expenditure-limit work. But much remains to be done before we can say with reasonable precision and certainty what the exact impact of spending changes or of the creation of charter schools might be on the quality of regular public schools.

---

[4] For a cautionary tale about the dangers of relying on state-level variation to identify the effects of school resources on students' earnings later in life, see Heckman, Layne-Farrar, and Todd (1996).

References

Bettinger, Eric. 1999. "The Effect of Charter Schools on Charter Students and Public Schools." National Center for the Study of Privatization in Education Occasional Paper No. 4, Columbia University.

Betts, Julian R. 1996. "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by G. Burtless. Washington, DC: Brookings Institution.

Betts, Julian R. and John E. Roemer. 2001. "Equalizing Opportunity through Educational Finance Reform." University of California, San Diego, Department of Economics, unpublished paper.

Coley, Richard J. 2002. *An Uneven Start: Indicators of Inequality in School Readiness.* Princeton, NJ: Educational Testing Service.

Downes, Thomas A. and David N. Figlio. 2000. "School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level Up or Dumb Down?" Tufts University, working paper.

Hanushek, Eric A. 1996. "School Resources and Student Performance." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by G. Burtless. Washington, DC: Brookings Institution.

Heckman, James, Anne Layne-Farrar, and Petra Todd. 1996. "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by G. Burtless. Washington, DC: Brookings Institution.

Hoxby, Caroline. 2002. "School Choice and School Productivity (or, Could School Choice be a Tide that Lifts All Boats?)." NBER Working Paper No. 8873 (April).

Nelson, Christopher and Kevin Hollenbeck. 2001. "'Does Charter School Attendance Improve Test Scores?' Comments and Reactions on the Arizona Achievement Study." W.E. Upjohn Institute Staff Working Paper No. 01-70.

Roemer, John E. 1998. *Equality of Opportunity*. Cambridge MA: Harvard University Press.