Test-Based Accountability and the Effectiveness of School Finance Reforms

By Christian Buerger, Seung Hyeong Lee, and John D. Singleton

A growing body of work provides compelling evidence that school resources can improve student outcomes. For example, changes in funding induced by school finance reforms across US states reduced the incidence of adult poverty, especially among low-income students, and contributed to closing test score gaps between low- and high-income school districts (Jackson, Johnson, and Persico 2016; Lafortune, Rothstein, and Schanzenbach 2018). Though these and related findings establish that “money matters” in principle, under what conditions additional school spending improves student outcomes remains an open question (Jackson 2018).

One hypothesis is that increases in school resources are especially likely to translate to improvements in learning in contexts with stronger incentives to promote student outcomes. For example, test-based school accountability systems create rewards or sanctions for schools based on aggregate student performance, with the goal of incentivizing school improvement. Such consequences might be explicit (and include threats of closing persistently low-performing schools, for example) or may be implicit, as with the provision of information on measured student performance (Figlio and Loeb 2011). Even prior to No Child Left Behind (NCLB), 30 states adopted “consequential” accountability policies that both publicly reported school results and attached sanctions or rewards to school performance (Hanushek and Raymond 2005).

The staggered adoption of school accountability within the “adequacy era” of school finance reforms forms the basis of our empirical approach to testing for complementarity between resources and incentives. Specifically, we estimate and compare the effects of school finance reforms on low-stakes National Assessment of Educational Progress (NAEP) scores between states that had an accountability system in place at the time of the finance reform and states that did not. This approach follows prior work in treating the timing of adequacy-based finance reforms as exogenous, but examines heterogeneity in their impact on student learning according to the presence of test-based accountability.

Our analysis produces two main findings. First, we find that test scores in low-income districts improve by nearly 0.012 standard deviation each year following a school finance reform when that reform was accompanied by consequential school accountability. This magnitude is over 60 percent larger than the comparable estimate from Lafortune, Rothstein, and Schanzenbach (2018), who instead pool all states. We find no evidence for pre-trends, and this finding is robust to numerous robustness checks.

In contrast, we do not find evidence that school finance reforms raise test scores in the absence of school accountability. An apparent effect on low-income districts in states without accountability in place is confounded by a trend in scores leading up to finance reforms. This pre-trend suggests that the parallel trends assumption is violated for these states. Test score growth appears to decline following finance reforms in states without test-based accountability.

These findings suggest that test-based accountability may raise the efficiency with which additional school resources are used. To
provide further evidence supporting this interpretation, we report tests of two alternative explanations. First, we do not find evidence that the results can be explained by the pattern of resource effects: impacts on school spending in low-income districts are largely similar between the two groups of states. Second, we find no evidence that the differences in effectiveness can be accounted for by other differences between states, such as preferences for education quality. Using variation from states with multiple finance reforms, we find that—within-state—those reforms accompanied by accountability are more effective.

I. Data

Our dataset combines student-level records of performance on NAEP exams with information about when states reformed their school finance system and implemented test-based accountability.

For finance reforms, we use tabulations from Lafortune, Rothstein, and Schanzenbach (2018), who identify those events “most consequential” for school funding in each state. Information on school accountability prior to NCLB is taken from Dee and Jacob (2011). Dee and Jacob (2011) label accountability systems as consequential if they are accompanied by (i) publicly available information on school performance and (ii) sanctions for low-achieving and rewards for high-achieving schools. We adopt this definition and assign the arrival of consequential accountability with the implementation of NCLB for those states without accountability prior to 2002.

Student performance is measured using restricted-access NAEP microdata. Including over 100,000 students nationwide every other year since 1990, the NAEP provides a representative sample of mathematics and reading test scores in grades four and eight. We follow previous research and standardize individual test scores by subject and grade to the distribution in the first year tested. We drop students attending charter and private schools. Like Lafortune, Rothstein, and Schanzenbach (2018), we aggregate the student-level records to summarize average achievement in low-income (first quintile) and high-income (fifth quintile) school districts by state, grade, subject, and year.

The final sample covers the period 1990 to 2011 for the 48 continental states. Twenty-five states had a school finance reform during the period. Of these, 13 had test-based accountability in place at the time of the school finance reform, while accountability arrived after the finance reform in 12 states. We summarize characteristics and differences between these groups in the online Appendix.

II. Empirical Approach

Let \( t^*_s \) indicate the calendar year that a state adopted test-based accountability and \( t^F_s \) the year of school finance reform. Note, \( t^F_s \) is undefined for the 23 states without a reform during the sample period. Our main estimating equation is given by

\[
y_{sjgt} = \theta(t - t^F_s)1(t > t^F_s)1(t^*_s \geq t^F_s)
\]

\[
+ \delta(t - t^F_s)1(t > t^F_s)1(t^* < t^F_s)
\]

\[
+ \pi_s + \lambda_{jgt} + \epsilon_{sjgt},
\]

where \( y_{sjgt} \) is average performance in low-income school districts in state \( s \), subject \( j \), and grade \( g \) at time \( t \). The treatment effects are modeled as linear post-reform trends.\(^1\) The variable \( \delta \) measures the per year effect of school finance reforms in states without accountability at the time \( t^F_s < t^*_s \), while \( \theta \) measures the finance reform effect in states where accountability is present \( t^F_s \geq t^*_s \). The terms \( \pi_s \) and \( \lambda_{jgt} \) represent state and subject-grade-year fixed effects, respectively.\(^2\)

Identification of \( \delta \) and \( \theta \) rests on the parallel trends assumption that \( y_{sjgt} \) would have trended similarly to control states in the absence of school finance reforms. While not directly testable, we augment equation (1) to test for trends in outcomes prior to finance reforms. Specifically, we estimate linear trends in test scores during the five years immediately leading up to finance reforms. We also present an event study figure that allows for visual inspection for

\(^1\) This allows for the effects to accumulate as more students are exposed to additional resources longer. We include results in the online Appendix that also estimate post-reform indicators (which are not statistically significant).

\(^2\) We weight observations in the estimation by the sum of the NAEP student weights within the cell.
Table 1—Estimates of Effects of School Finance Reforms on Test Scores in Low-Income Districts

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accountability in place</td>
<td>Pre-trend</td>
<td>Pre-trend</td>
</tr>
<tr>
<td>× Years elapsed</td>
<td>0.012</td>
<td>0.011</td>
</tr>
<tr>
<td>× Pre-trend</td>
<td>0.000</td>
<td>−0.003</td>
</tr>
<tr>
<td>No accountability</td>
<td>Pre-trend</td>
<td>0.004</td>
</tr>
<tr>
<td>× Pre-trend</td>
<td>0.029</td>
<td>0.041</td>
</tr>
</tbody>
</table>

State FE Yes Yes Yes
Subject-grade-year FE Yes Yes Yes
Accountability controls No No Yes

Notes: Table presents results of estimating the effects of school finance reforms on student achievement in low-income districts for accountability and non-accountability states. The number of observations is 1,436. Standard errors clustered at the state level reported in parentheses. Observations are weighted by the sum of the NAEP student weights within the quintile-state-grade-subject-year cell. Columns 2 and 3 estimate pre-trends over the five years immediately prior to finance reforms (but do not report estimates on indicators that the calendar year is six or more years before a finance reform). Accountability controls include a post-accountability adoption indicator and linear pre- and post-accountability trends (each interacted with whether accountability was NCLB).

III. Results

Table 1 reports the main results on student achievement in low-income school districts. The estimates in column 1 indicate that test scores in low-income districts improve by 0.012 standard deviation (σ) each year following a school finance reform in accountability states. Low-income districts’ test scores in non-accountability states increase by around 0.006σ each year. Finance reform effects on low-income districts in both groups of states in the baseline specification are statistically significant.

The second column of Table 1 adds pre-trend terms to the model, and Figure 1 plots event study estimates. For those states in which the finance reform is accompanied by accountability, the estimated impact on test scores is unchanged, and consistent with parallel trends, the trend prior to reform is zero. By contrast, the pre-trend in non-accountability states is positive and statistically significant, indicating that test scores are increasing prior to finance reforms in these states. Moreover, the pre-trend is larger in magnitude than the post-reform effect; this leveling off of test score growth in non-accountability states can also be seen in Figure 1. The third column of Table 1 reports results controlling also for the impacts of school accountability on test scores, which are similar to column 2.3

The results in Table 1 facilitate hypothesis tests for whether finance reform effects are the same between the two groups of states. The p-value for whether the post-reform trends are the same for accountability and non-accountability states is 0.27 in column 1. However, this test is only valid under the assumption that there are no pre-trends. Testing instead for whether the changes in test score trends following finance

3 We include an indicator for post-accountability adoption in the model as well as separate linear trends leading up to and following accountability (we also include interactions of these variables with an indicator for whether the accountability was NCLB).
reforms are equal yields $p$-values of 0.05 and 0.02 in columns 2 and 3, respectively. Note that this test instead relies on the strong assumption that test scores in non-accountability states would have continued to grow by $0.03-0.04\sigma$ per year in the absence of a finance reform.

We examine and report results for several robustness checks on the results in Table 1 in the online Appendix. In particular, we examine sensitivity to heterogeneity according to the calendar year of the reform, to considering only court-ordered reforms, to alternative dating of reforms, and to removing states from the analysis in which the finance reform and accountability occur in close succession. The main findings are robust to these changes. We do not find effects of finance reforms on test scores in high-income districts (with or without accountability in place).

IV. Mechanism

The hypothesized mechanism for the findings is that test-based accountability enhances the efficiency with which additional school resources are used. In this section, we discuss additional results that test alternative explanations.

On the one hand, it may be that the results are explained by the pattern of resource effects. For instance, the effect of finance reforms on low-income district spending could be larger or directed toward more effective resources when accountability is in place. Alternatively, the heterogeneity in finance reform effects may not arise from differences in accountability per se: accountability states might exhibit stronger preferences for education quality, and those preferences—rather than the incentives embedded in school accountability policies—enhance finance reform impacts.

We marshal data on district spending by year to estimate the effects of finance reforms on school resources. The data and results are reported in full detail in the online Appendix. We find that resource effects on low-income districts are largely similar between accountability and non-accountability states. Low-income districts in non-accountability states, where we do not find robust evidence for increases in student learning, increase spending by around 9 percent on average following a school finance reform. This compares with a 7 percent increase on average in low-income districts where the reform is accompanied by accountability. States without accountability also experience slightly larger increases in instructional and student support spending per pupil. The one input that appears to increase relatively more in the presence of accountability is teacher pay (4 percent versus 3 percent).

We also test whether differences between states other than the presence of accountability can account for the differences in two ways. First, we augment equation (1) to allow the effect of finance reforms to also depend on observed differences between states. We examine heterogeneity according to state median income and educational attainment (both measured in 1990), the 1988 Democratic vote share, and teacher union strength (Brunner, Hyman, and Ju 2020). The main results are unchanged when these interactions are included.

The second way we test this explanation is using states that experienced finance reforms both prior to and following the adoption of accountability during the sample. We create and stack duplicate test score series for each individual reform, including state-specific finance reform effects in the model. This test relies on the within-state variation while controlling for all (observed and unobserved) time-invariant state characteristics that may enhance the effectiveness of finance reforms. The results indicate that reforms accompanied by accountability are relatively more effective.

V. Conclusion

Education policies that target school spending are often juxtaposed with reforms that reward or sanction teachers and schools based on measured student performance. In this paper, we present new evidence that resources and incentives are complementary inputs to education production. This finding has important policy implications, as efforts that relax accountability

---

4If we relax this assumption by estimating pre-trends over the full pre-reform period (results in Table 1 censor the pre-trend variables below $-5$), we can still reject the null that the changes are the same.

5We also do not find evidence of effects on the local spending share or student demographics in either accountability or non-accountability states.
standards may also reduce the effectiveness of increasing school resources.

Our findings also suggest several avenues for future work, such as examining impacts on longer-run indicators of student success and for certain subgroups of students. In addition, higher-resolution measures of accountability pressure—for example, variation across schools—combined with spending changes could isolate the role of incentives with greater precision.

REFERENCES


