ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-277

Test-Based Accountability and the Effectiveness of School Finance Reforms

Christian Buerger
Indiana University-Purdue University Indianapolis

Seung Hyeong Lee
Harvard University

John D. Singleton
University of Rochester

A recent literature provides new evidence that school resources are important for student outcomes. In this paper, we show that school finance reform-induced increases in student performance are driven by those states that had test-based accountability policies in place at the time. By incentivizing school improvement, accountability systems (such as the federal No Child Left Behind act) may raise the efficiency with which additional school funding gets spent. Our empirical approach leverages the timing of school finance reforms to compare funding impacts on student test scores between states that had accountability in place at the time of the reform with states that did not. The results indicate that finance reforms are three times more productive in low-income school districts when also accompanied by test-based accountability. These findings shed new light on the role of accountability incentives in education production and the mechanisms supporting the effectiveness of school resources.

VERSION: August 2020

Suggested citation: Buerger, Christian, Seung Hyeong Lee, and John D. Singleton. (2020). Test-Based Accountability and the Effectiveness of School Finance Reforms. (EdWorkingPaper: 20-277). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/zk4r-fr93

Test-Based Accountability and the Effectiveness of School Finance Reforms*

Christian Buerger†
Indiana University-Purdue University Indianapolis

Seung Hyeong Lee‡
Harvard University

John D. Singleton§
University of Rochester

August 2020

Abstract

A recent literature provides new evidence that school resources are important for student outcomes. In this paper, we show that school finance reform-induced increases in student performance are driven by those states that had test-based accountability policies in place at the time. By incentivizing school improvement, accountability systems (such as the federal No Child Left Behind act) may raise the efficiency with which additional school funding gets spent. Our empirical approach leverages the timing of school finance reforms to compare funding impacts on student test scores between states that had accountability in place at the time of the reform with states that did not. The results indicate that finance reforms are three times more productive in low-income school districts when also accompanied by test-based accountability. These findings shed new light on the role of accountability incentives in education production and the mechanisms supporting the effectiveness of school resources.

*We thank John Beshears, Robert Bifulco, Thomas Dee, and participants at AEFP for helpful comments. We are also grateful to Jesse Rothstein and Josh Hyman for assistance with the data construction.

†cbuerge@iupui.edu
‡seunghyeong_lee@fas.harvard.edu
§john.singleton@rochester.edu

1 Introduction

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 U.S. 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 et al., 2016; Lafortune et al., 2018). Though these and related findings based on quasi-experimental variation 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 (Chubb and Moe, 1990; Hanushek and Jorgenson, 1996; Ladd, 2007). We investigate this question in by focusing on the role of test-based accountability. Accountability systems create rewards or sanctions for schools based on aggregate student performance with the goal of incentivizing school improvement (Figlio and Loeb, 2011). 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 of measured student performance. Even prior to No Child Left Behind (NCLB), thirty U.S. states adopted accountability policies, termed "consequential" by Hanushek and Raymond (2005), that both publicly reported school results and attached sanctions or rewards to school performance. The incentives embedded in these accountability systems raise the question whether increases in school resources, as for instance by school finance reforms, are more likely to improve student outcomes in settings where such policies are in place.

To answer this question, we estimate the causal effects of school finance reforms on student performance while accounting for the role of consequential school accountability. Our empirical approach leverages variation in the timing of school finance reforms relative to states' adoption of test-based school accountability. Specifically, we estimate and compare effects on student achievement in the thirteen states that had accountability systems in place

at the time of their school finance reform with the effects in those twelve states that did not. We draw on National Assessment of Education Progress (NAEP) records from 1990 to 2011, during the "adequacy era" in school finance reforms, to examine the impacts on students in both high- and low-income school districts (as well as the performance gap).

The results reveal that the effects of school finance reforms on student learning are driven entirely by those states that had test-based accountability in place at the time. For low-income districts in these states, the estimates indicate that test scores improve around 0.012 standard deviation (σ) each year following a school finance reform. In contrast, the corresponding point estimate for low-income districts in states without an accountability policy is only about a third of this size. Moreover, after accounting for trends leading up to the finance reform, the estimate for non-accountability state cannot be regarded as statistically different from zero. We examine the sensitivity of these results to several sensitivity checks, including controls for the timing and impact of accountability adoption.

While these findings suggest that accountability policies raise the efficiency with which school resources are used, it nonetheless may be that the results are explained by the pattern of resource effects. To examine this possibility, we examine heterogeneity in school finance reform impacts on school spending and other education inputs in low-income districts. The results indicate that resource effects of finance reforms are largely similar across states with and without accountability policies at the time. Low-income districts in states without accountability, where we do not find robust evidence for increases in student achievement, increase spending by around 9% on average following a school finance reform as compared to 7% for accountability states. Finance reforms are therefore considerably more productive when accompanied by the presence of test-based accountability policies, indicative of the importance of incentives.

Recent work, leveraging quasi-experimental variation in spending, provides evidence that resources can matter (e.g. Jackson et al. 2016; Hyman 2017; Lafortune et al. 2018; Biasi 2019), but raise the key questions of when and which resource increases translate to gains in outcomes (Jackson, 2018). Brunner et al. (2019) examine the role of teachers' unions

in allocating finance-reform induced spending increases, while Baron (2019) compares the effectiveness of operational as opposed to capital spending. Our paper instead examines the importance of incentives, specifically those embedded in consequential school accountability systems, with this motivation. Our focus on incentives is shared with Lastra-Anadón and Peterson (2019), who find that districts where a high local share accompanies spending increases experienced greater increases in student test scores.

In studying the interaction between incentives and resources, our paper also connects to a literature on policy instruments and mixes, which emphasizes that policy instruments can either supplement or substitute for one another (Gunningham et al., 1998; Hou and Brewer, 2010; Yi and Feiock, 2012). While much of the empirical literature focuses on environmental policy[1], an exception related to our work is Johnson and Jackson (2019), who examine complementarity between Head Start exposure and finance reform-induced spending increases. They find that both policies individually increased long-term student outcomes, but that the effects of spending during K-12 education are largest for low-income students exposed to both programs. We similarly find that the combination of consequential accountability with increases in school funding improved outcomes for students in low-income districts. This finding has important implications for designing effective policies that expand school resources.

The remainder of the paper is organized as follows. We describe the background and dataset we assemble in the next section. Section 3 details our empirical approach, while Section 4 presents the main findings and robustness checks. We conclude in Section 5.

2 Background and Data

Prior to school finance reforms, which were first initiated in the 1970s, local governments provided the majority of funds for K-12 education in the United States. Since these funds

[1]For instance, Yi and Feiock (2012) analyze the relationship between minimum requirements for renewable energy and incentives set by tax and rebate programs in the U.S. states. They find that that incentives on the consumer side spur the production of renewable energies by providers.

relied heavily on property taxes, education budgets were largely a function of local tax bases in addition to voters' ability and willingness to tax themselves. Consequently, large disparities in school resources arose between school districts (for an overview see Yinger 2004; Corcoran and Evans 2015).

Our paper focuses on the second wave of finance reforms, which began in 1989. These "adequacy" court cases were driven by provisions in state constitutions that require legislatures to guarantee a minimum level of free education to all students. Induced by judicial rulings (or the threat of them), state governments typically implemented foundation plans, which transfer to targeted districts the difference between a legislature-determined minimum level of spending and a local contribution. The resulting school funding schemes substantially raised state transfers to low-income school districts (Ladd and Yinger, 1994; Enrich, 1995; Minorini and Sugarman, 1999a,b; Lukemeyer, 2003).

Test-based accountability policies gained momentum during the time of adequacy reforms. Although NCLB ensured nationwide adoption, thirty states adopted consequential school accountability systems prior to 2002. Aimed at correcting institutional incentives facing teachers and administrators through rewards and sanctions, the available evidence suggests positive effects of accountability reforms on student performance. Carnoy and Loeb (2002) and Hanushek and Raymond (2005) find positive impacts of pre-NCLB accountability adoption when accompanied by consequential penalties for missing performance standards, such as state interventions into local school systems.[2] Dee et al. (2010) and Dee and Jacob (2011) find that NCLB led to increases in mathematics, though not reading, test scores while leading to increases in school resources Dee et al. (2013)

We examine how test-based accountability interacts with increases in school resources to impact student learning. In particular, the incentives to use resources efficiently suggest that finance reforms may be more likely to translate to student performance in settings where accountability policies are in place. The next subsection describes the dataset we use to test

[2]Jacob (2005) also shows positive effects of accountability policies on students' test scores for Chicago, but the findings are more nuanced. The test-score increases are not mirrored in low-stakes examinations and he finds evidence of teachers responding strategically to accountability pressure.

for this complementarity.

2.1 Data Sources

Our study draws on several data sources to combine student-level test performance and district-level variables with information about when states reformed their school finance system and implemented test-based accountability. To determine the year of school finance reforms, we utilize tabulations from Lafortune et al. (2018). These tabulations include court ordered and legislative events and, when states have multiple reforms in the adequacy era, determine the most consequential reform by identifying events that had the largest impact on the state's finance system.

Information on test-based accountability prior to No Child Left Behind is taken from Dee and Jacob (2011), who provide the most recent and comprehensive effort to classify these policies. Dee and Jacob (2011) label accountability systems as consequential if they are accompanied with: (1) publicly available information on school performance and (2) sanctions for low achieving and rewards for high achieving schools. Only reforms that fulfill both criteria are expected to create incentives for increasing student performance. We adopt this definition for our analysis and assign the arrival of consequential accountability with the implementation of NCLB for those states without accountability prior to 2002.[3] We summarize the timing of finance reforms and accountability adoption across states in the next subsection.

For outcomes, we employ information on student performance and school district resources. Student performance is measured utilizing restricted-access microdata from the National Assessment of Educational Progress (NAEP), administered by the U.S. Department of Education. The NAEP provides a representative sample of mathematics and reading test scores for grades four and eight, including over 100,000 students nationwide for every other year since 1990. We follow previous research (Lafortune et al., 2018; Brunner et al., 2019) and standardize individual test scores by subject and grade to the distribution in the first

[3] 2003 is coded as the first post-accountability year for these states.

year tested. We also drop observations recorded for students attending charter and private schools, focusing only on public schools.

Information on school district resources are taken from the Local Education Agency (School District) Finance Survey (F-33), maintained by the National Center for Education Statistics (NCES). The F-33 contains detailed information on annual revenues and expenditures for all school districts in the United States starting in 1990.[4] The two missing years in the F-33 (1993, and 1994) are replaced with data from the Annual Survey of School System Finances, conducted by the U.S. Census Bureau, which contains the same fiscal information as the F-33.[5] We augment these variables with information on student enrollment and staff counts from the NCES Common Core of Data (CCD) school district universe survey.

To measure differential impacts on achievement and resources, we classify districts as low- or high-income using information on average household income in 1990 (the first year in our data) from the School District Data Book. We create income quintiles and average the test score microdata and district-level variables to the state by year by quintile level. In doing so, each test score is weighted by the sum of NAEP student weights and each district variable by average log enrollment.[6] Our analysis focuses on the fifth (high-income) and first (low-income) quintiles (as well as the gaps in test score perfomrance between them).

Our final sample covers the period from 1990 to 2011 for forty-eight states.[7] The next subsection describes data patterns and presents summary statistics.

2.2 Data Summaries

Table A1 presents states' adoption of school finance reforms and test-based accountability policies. Twenty-five states had school finance reforms sometime between 1990 and 2011. We

[4]We exclude outlier districts following Lafortune et al. (2018): districts with a small number of students, with extreme increase/decrease in enrollment, and with extreme revenue and expenditure.

[5]All the values were converted to 2011 dollars by using the annual average of the seasonally adjusted Consumer Price Index. There is no finance data available for the fiscal year 1991.

[6]We utilize a crosswalk provided by Jesse Rothstein for the years prior to 2000. For all other years, NCES's unique district ID is available in the NAEP.

[7]We exclude Hawaii and the District of Columbia from the analysis as both jurisdictions consists of a single school district. Alaska is also dropped from the analysis because the cost of providing education differ greatly from other states and transfers to school are based on a highly volatile severance tax.

define these states as "treatment" states. Of these, thirteen had test-based accountability in place at the time of the school finance reform. For instance, California's school finance reform took place in 2004, but the state had adopted consequential accountability five years earlier. We define these states as "accountability" states. On other hand, twelve other states, who we define as "non-accountability" states, did not have accountability in place. As an example, Ohio reformed its school finance system in 1997, but accountability was not implemented until NCLB. We define the twenty-three states without a school finance reform during the period as "control" states.

Table 1 presents summaries for student achievement and school resource for these groups of states in 1990, the first year of our sample. The first and second columns present summaries for treatment states – those that ever had a finance reform – and control states, while the third column reports differences between the two groups of states. The first row shows that low-income districts in treatment states had NAEP scores around 0.24 standard deviations (σ) lower than low-income districts in control states on average. High-income districts in control states also had higher test scores than treatment states, though the difference is not statistically different from zero. Low-income districts in treatment states' average total expenditure per pupil in 1990 was slightly higher than that of control states ($8,363 vs. $8,060), a difference that is not statistically significant. The corresponding standard errors for spending highlight that, while differences on average are limited, there is considerable variation within each group. Treatment-control gaps among high-income states are even smaller. Table 1 shows similarly minor differences among high- and low-income districts between treatment and control states in pupil teacher ratios and minority student share on average. Teacher salary differences on average are larger, with the standard errors indicating a lot of variation within treatment and control groups of states.

The fourth through sixth columns of Table 1 compare the two groups of treated states: those that had test-based accountability in place at the time of their school finance reform and states that did not. Among low-income districts, NAEP scores were about 0.05σ higher on average in accountability states in 1990. On the other hand, test scores were much higher

in non-accountability states on average among high-income districts. Accountability states' total revenue and expenditure per pupil were higher than those of non-accountability states on average (by around $1,360 and $1,512, respectively) among low-income districts. The standard errors again indicate large difference among accountability and non-accountability states in spending. Differences in pupil-teacher ratios and minority student shares also cannot be distinguished from zero statistically. One difference of note among low-income districts is mean household income between accountability states (around $46,000) and non-accountability states (about $43,000), indicating that low-income districts in accountability states are somewhat more affluent in absolute terms. These summaries provide evidence that – at the onset of the adequacy era – accountability and non-accountability states were diverse groups in terms of school resources and student characteristics.

3 Empirical Approach

Our empirical approach examines the heterogeneity of school finance reforms on test scores and resources for states with and without accountability reforms. This approach leverages the variation in the timing of school finance reforms relative to the implementation of test-based accountability.

More formally, we define t_s^{SFR} as the first year state s was exposed to a school finance reform. If a state did not have school finance reforms, it is a control state and t_s^{SFR} is undefined. The average effects of school finance reforms are estimated by examining changes in outcome variables associated with the timing of the reform:

$$y_{st} = \beta(t - t_s^{SFR}) \times \mathbf{1}(t > t_s^{SFR}) + \pi_s + \lambda_t + \epsilon_{st} \tag{1}$$

where y_{st} is an outcome variable of interest (e.g. student performance) for state s in year t, π_s represents state fixed effects, λ_t year fixed effects, and ϵ_{st} is an error term. In this setup, β measures the post-reform per year effect of school finance reforms relative to control states (for who the term is zero) for states with and without accountability policies at time t_s^{SFR}.

Causal inference for β rests on the "natural experiment" that the timing of school finance reforms is as good as random (i.e. "parallel trends"). This specification corresponds closely to the one estimated by Lafortune et al. (2018).

Our empirical approach adapts equation (1) to test for complementarity with consequential school accountability policies. Define t_s^{ACC} as the year a state adopted test-based accountability. Among the group of "treatment" states, a state belongs to the accountability group if $t_s^{SFR} \geq t_s^{ACC}$ and belongs non-accountability states if $t_s^{SFR} < t_s^{ACC}$. Our main equation can be written as:

$$\begin{aligned} y_{st} = \delta(t - t_s^{SFR}) \times \mathbf{1}(t > t_s^{SFR}) \times \mathbf{1}(t_s^{SFR} < t_s^{Acc}) \\ + \theta(t - t_s^{SFR}) \times \mathbf{1}(t > t_s^{SFR}) \times \mathbf{1}(t_s^{SFR} \geq t_s^{Acc}) + \pi_s + \lambda_t + \epsilon_{st} \end{aligned} \tag{2}$$

where δ and θ are the parameters of interests. δ measures the per post-reform year effect of school finance reforms in states without accountability at the time of a finance reform (relative to control states), while θ measures the effect in accountability states. If test-based accountability enhances the effectiveness of finance reforms, we expect that $\theta > \delta$ when y_{st} equals student performance. We cluster the standard errors by state when estimating equation (2). We also weight NAEP scores by the inverse squared standard error in estimation to improve efficiency (consistent with Lafortune et al. 2018).[8]

Identification of δ and θ follows from the parallel trends assumption that y_{st} would have trended similarly to control states in the absence of school finance reforms. While this is not directly testable, we pursue two checks that examine trends in student performance *prior* to school finance reforms. The first check is that we expand the main equation to explicitly allow for linear pre-trends. Specifically, we include the variable:

$$Pretrend_{st} = (t - t_s^{SFR}) \times \mathbf{1}(t \leq t_s^{SFR})$$

in the regression interacted with separate indicators for belonging to accountability ($t_s^{SFR} \geq$

[8]We examine robustness of our test score findings to the use of weights.

t_s^{ACC}) or non-accountability ($t_s^{SFR} < t_s^{ACC}$) states. To focus only on the periods immediately leading up to school finance reforms, however, we set $Pretrend_{st}$ to zero for years more than five years before the reform.[9] For the second test, we estimate "event study" specifications that interact the effects of school finance reforms with the time before and after the reform:

$$\begin{aligned} y_{st} = &\sum_{k=-2}^{k=5} \delta_k^{'} \mathbf{1}(W_{st} = k) \times \mathbf{1}(t_s^{SFR} < t_s^{Acc}) + \\ &\sum_{k=-2}^{k=5} \theta_k^{'} \mathbf{1}(W_{st} = k) \times \mathbf{1}(t_s^{SFR} \geq t_s^{Acc}) + \pi_s + \lambda_t + \epsilon_{st} \end{aligned} \tag{3}$$

where W_{st} indexes the number of two year windows relative to one year before school finance reforms. We bin years in windows up to 5 years before and 10 years after school finance reforms and measure relative to one year before and the year of school finance reforms ($\delta_0^{'} = \theta_0^{'} = 0$). Therefore, $\delta_k^{'}$ estimates the effects of school finance reforms in a given window k relative to the window $k = 0$ for non-accountability states and $\theta_k^{'}$ for accountability states.

We provide several robustness checks for the results of these specifications in a later part of this study.

4 Results

Our empirical analysis starts with documenting the impact of school finance reforms on test-scores for states with and without accountability policies and for districts in different income quintiles. Based on our event study specification (Equation (3)), we begin with several figures of the same basic form. Coefficients for reforms in states with accountability policies are depicted by a blue solid line, while the effects for states without them are displayed by a red solid line. Confidence intervals are in the corresponding colors, but in whiskers.

Figure 1 shows the results for districts in the lowest income quintile. For the pre-period, the coefficients for the accountability states are negative and reveal a "v"-shaped pattern.

[9]The pretrends are thus "local" linear. This adjustment is important because accountability states have many more pre years on average than do non-accountability states. We also include intercepts for years beyond five years before school finance reforms in the pretrend regressions.

The confidence intervals include zero at all times. The effects in the post-reform, meanwhile, period are positive and increase over time. The coefficients in the last two post periods are around 0.15 standard deviations. For non-accountability states on the other hand, the pre-trend is positive and statistically significant. The estimates increase marginally before reaching a plateau and then increase again in years nine and ten after the school finance reform.

Figure 2 presents coefficients for a similar model, but this time only districts in the highest income quintile are used for the analysis. No clear change in test-scores can be established for either set of states.

Figure 3 focuses on the test-score gaps between low and high-income districts in accountability and non-accountability states. Accountability states show a small (for two and three years prior to treatment) and non-accountable states a large positive pre-trend. The post coefficients, for accountability states, are positive at all times and increase with the exception of year three and four. The non-accountability states have negative coefficients up to year five and small positive coefficients in the later time periods. The confidence intervals include zero at all times.

We report coefficient from our parametric models in Table 2. Columns with odd numbers present specifications employing a linear time trend to estimate reform effects, while the remaining columns examine the robustness of these results, when linear pre-trends are added to the model. The coefficients indicate that test scores in low-income districts (first quintile) improve by 0.012 standard deviations (σ) each year following a school finance reform in accountability states. Test score for the same districts in non-accountability states increase by around 0.006σ each year. Effects for both groups of states are statistically different from zero. When pre-trends are added, the effect for the accountability states remains the same in magnitude and precision, while the pre-trend is essentially zero. Conversely, the coefficient for the non-accountability states is greatly reduced and not statistically significant anymore. The pre-trend is positive and precisely estimated.

In contrast to these results, the estimates for high-income districts in Columns (3) and

(4) are much smaller in magnitude and do not show statistical significance. The coefficient, on the trend variable, for non-accountability states switches signs and is now negative.

As the result of these findings, the performance gap between low and high income districts declines over time. For non-accountability states, the effect is initially statistically significant, as a result of test scores decline in high-income districts, but the inclusion of the pre-trend takes some of that magnitude and precession away. The positive and statistically significant pre-trend, moreover, undermines the causal relationship between finance reforms and test-scores for these states even more. Places with accountability policies have now a positive and marginally significant coefficient, whereas the pre-trend is negative and also statistically significant at the 0.1 level.

In sum, the results show that the effects of school finance reforms on student learning in low-income districts are driven by accountability states. Based on the pre-trends check and event-study framework, we find evidence for parallel pre-trends in accountability states (consistent with our identifying assumption) and conclude that school finance reforms have meaningful causal effects on test scores in low-income districts when accompanied by test-based accountability. We find that the comparable point estimate for states without accountability at the time of their finance reform is about a third as large in magnitude and that this apparent effect is due to a significant trend prior to the reform.[10]

4.1 Robustness Checks

Besides controlling for pre-trends in our main specification, we run several additional robustness checks to analyze the sensitivity of our results. These tests focus on performance changes in low-income districts that had statistically significant results in the main specification.

A first check controls for the effects of test-based accountability policies themselves. Accountability adoption precedes finance reform for accountability states, but accountability arrives subsequent to their reform for non-accountability states. We add several variables

[10]We do not have statistical power to reject the null hypothesis that the effects in low-income districts between accountability and non-accountability states are the same. We can reject the one-sided hypothesis that the effect is larger in non-accountability states, however.

to our estimating equation to flexibly control for accountability effects: an indicator that switches "on" post-accountability, a linear pre-trend for the immediate five years before the implementation of accountability, and a linear post-trend. Additionally, we also allow these effects to vary by whether or not the accountability reform was NCLB. The top panel of Table 3 presents the results for these tests. The estimates are comparable to the baseline findings, indicating that our results are robust the effects of accountability policies.

We also examine whether our findings are explained by the possibility that later finance reforms – which happen to be more likely to occur after accountability adoption – are simply more productive than earlier reforms in Table 3. To test this, we allow the effect of school finance reforms to be additionally heterogeneous by the calendar year of reform. Specifically, we interact our treatment effect, $(t - t_s^{SFR})$, with $(t_s^{SFR} - 2000)$. The main parameters in these regression are thus interpreted as the effects of a school finance reforms in year 2000, the average year a school finance reform occurs. The results, reported in the middle panel of Table 3, indicate that although the earlier finance reforms are indeed less effective in low-income districts, the main effect of interest is similar to Table 2 even after controlling for this.

Our estimating equation, equation (2), models the effect of finance reforms as a linear post-reform trend variable, which parsimoniously captures the effects of learning dynamics on test scores. We examine sensitivity to this model of the "treatment effect" by additionally including variables in the regression that capture a level shift in test scores post-reform. The results, shown in the bottom panel of Table 3, do not indicate statistically significant level shifts in scores following reform, consistent with the exclusion of these variables from the main results. Moreover, we find that the post-reform trend for low-income districts in accountability states is nonetheless significant at the 90% level and comparable in size to the baseline results.

A final robustness check, reported in Table A2, re-estimates the effects on NAEP scores without weights. These results are consistent with our main findings, showing no post-reform effect on test scores in low-income districts in non-accountability states, but significant

positive effects in low-income districts in accountability states. The unweighted results also show a statistically significant reduction in the test score gap between high- and low-income districts in accountability states.

4.2 Mechanisms

The results indicate that school finance reforms cause test score increases in accountability states but not in non-accountability states. The proposed mechanism is that accountability policies raise the efficiency with which school resources are used. However, it may be that the results are explained by the pattern of resource effects if, for instance, the effect of finance reforms on low-income district spending is larger in accountability states. Alternatively, it may be that accountability states direct additional spending increases to more productive inputs or that the student composition changes.

We investigate these possibilities in this section by examining heterogeneity in school finance reform impacts on school spending and education inputs. Given we mainly find test score effects in low-income districts, we focus on spending in these districts. If the pattern of resource effects cannot explain the results, it highlights the importance of incentives embedded in accountability systems. To examine effects on spending and inputs, we estimate models similar to equation (2) at the state-level with pre-trend controls. In contrast with equation (2), however, we model the reform effects as just intercept level shifts because it fits these data better.

Table 4 presents the results of estimating the effects of school finance reforms on district components for accountability and non-accountability states. The results indicate that resource effects of finance reforms are largely similar across accountability and non-accountability states. Low-income districts in non-accountability states, where we do not find robust evidence for increases in student achievement, increase spending by around 9% on average following a school finance reform. This effect compares with a 7% increase in low-income district spending in accountability states on average.

Non-accountability states experienced slightly larger increases in other financial aspects

relative to accountability states, including in instructional expenditures, spending on teacher salaries and benefits, and spending on student support. Table 4 also shows that pupil-teacher ratios slightly decreased about the same amount in both groups on states due to finance reforms. At the same time, we do not see evidence that the local spending share or student demographics, such as the share of students who qualify for subsidized lunch, change in low-income districts in either accountability and non-accountability states. The only input that appears to increase relatively more in accountability states are teacher salaries. Teacher salaries in accountability states increase by 4%, while they increase by less than 3% (and the effect is not statistically significant) in non-accountability states.

Overall, the results suggest that changes in spending patterns cannot explain the difference in the test score impacts of school finance reforms. Finance reforms are therefore considerably more productive when accompanied by the presence of test-based accountability policies, indicative of the importance of incentives to promote student performance.

5 Conclusion

While a recent literature provides new evidence that "money matters" for student outcomes, we consider the role of incentives in raising the efficiency with which increases in school resources are used in this paper. We do this by comparing the effects of school finance reforms on student test scores between states that had a consequential accountability system in place at the time with those that did not.

The results indicate that school finance reform-induced increases in student performance are driven by those states that had test-based accountability in place at the time. Test scores in low-income (first quintile) districts improve by 0.012σ each year following a school finance reform in accountability states. The corresponding point estimates for non-accountability students is around a third as large in magnitude and statistically insignificant after accounting for trends prior to the finance reform. In addition, we find that impacts on school resources are unlikely to account for this pattern of effects.

Our findings have several important implications. First, our results show that the positive impacts of school finance reforms on test scores, measured in previous studies (e.g. Lafortune et al. 2018), are almost entirely driven by states that implemented accountability policies prior to changes in states aid. Second, we reveal an important complementarity between school finance reforms and test-based accountability systems. School finance reforms are much more effective when they are accompanied by accountability policies that create rewards or sanctions for schools based on student performance. Third, as our analysis of the finance mechanism uncovers, the incentives set by accountability policies raise the efficiency with which increases in school resources are used, a finding that has significant implications for policy.

At the same time, our study does not come without caveats. For instance, our research is poorly suited to analyze which combination of resources in interaction with consequential accountability policies leads to an optimal mix of school inputs. We can only say the average impact of finance reforms on school resources is much greater in states with performance incentives than in states without them. Another limitation is our focus solely on student test score performance, as measured by NAEP scores. A direction for future work is identifying the complementarity of incentives and resources for longer-run student outcomes of human capital accumulation.

References

Baron, E. J. (2019). School spending and student outcomes: Evidence from revenue limit elections in wisconsin. Working paper, Available at SSRN 3430766.

Biasi, B. (2019). School finance equalization increases intergenerational mobility: Evidence from a simulated-instruments approach. Working Paper 25600, National Bureau of Economic Research.

Brunner, E., J. Hyman, and A. Ju (2019). School finance reforms, teachers' unions, and the allocation of school resources. *Review of Economics and Statistics Forthcoming*.

Carnoy, M. and S. Loeb (2002). Does external accountability affect student outcomes? a cross-state analysis. *Educational evaluation and policy analysis 24*(4), 305–331.

Chubb, J. E. and T. M. Moe (1990). Politics, markets, and america's schools. *Washington, DC: Brookings Institution*.

Corcoran, S. P. and W. N. Evans (2015). Equity, adequacy, and the evolving state role in education finance. *Handbook of research in education finance and policy*, 353–371.

Dee, T. S. and B. Jacob (2011). The impact of no child left behind on student achievement. *Journal of Policy Analysis and management 30*(3), 418–446.

Dee, T. S., B. Jacob, and N. L. Schwartz (2013). The effects of nclb on school resources and practices. *Educational Evaluation and Policy Analysis 35*(2), 252–279.

Dee, T. S., B. A. Jacob, C. M. Hoxby, and H. F. Ladd (2010). The impact of no child left behind on students, teachers, and schools [with comments and discussion]. *Brookings papers on economic activity*, 149–207.

Enrich, P. (1995). Leaving equality behind: New directions in school finance reform. *Vand. L. Rev. 48*, 100.

Figlio, D. and S. Loeb (2011). School accountability. In *Handbook of the Economics of Education*, Volume 3, pp. 383–421. Elsevier.

Gunningham, N., P. Grabosky, and D. Sinclair (1998). Smart regulation. *REGULATORY THEORY 133*.

Hanushek, E. A. and D. W. Jorgenson (1996). *Improving America's Schools: The Role of Incentives*. Washington, DC: The National Academies Press.

Hanushek, E. A. and M. E. Raymond (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management 24*(2), 297–327.

Hou, Y. and G. A. Brewer (2010). Substitution and supplementation between co-functional policy instruments: evidence from state budget stabilization practices. *Public Administration Review 70*(6), 914–924.

Hyman, J. (2017). Does money matter in the long run? effects of school spending on educational attainment. *American Economic Journal: Economic Policy 9*(4), 256–80.

Jackson, C. K. (2018). Does school spending matter? The new literature on an old question. Working Paper 25368, National Bureau of Economic Research.

Jackson, C. K., R. C. Johnson, and C. Persico (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics 131*(1), 157–218.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the chicago public schools. *Journal of public Economics 89*(5-6), 761–796.

Johnson, R. C. and C. K. Jackson (2019). Reducing inequality through dynamic complementarity: Evidence from head start and public school spending. *American Economic Journal: Economic Policy 11*(4), 310–49.

Ladd, H. F. (2007). Holding schools accountable revisited. In *Spencer Foundation Lecture in Education Policy and Management, presented at the 2007 APPAM Fall Research Conference, Washington DC, November*, Volume 8.

Ladd, H. F. and J. Yinger (1994). The case for equalizing aid. *National Tax Journal 47*(1), 211–224.

Lafortune, J., J. Rothstein, and D. W. Schanzenbach (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics 10*(2), 1–26.

Lastra-Anadón, C. X. and P. E. Peterson (2019). Who benefits from local financing of public services? A causal analysis. Working paper, Program on Education Policy and Governance Working Papers Series.

Lukemeyer, A. (2003). *Courts as Policymakers: School Finance Reform Litigation*. LFB Scholarly Publishing LLC.

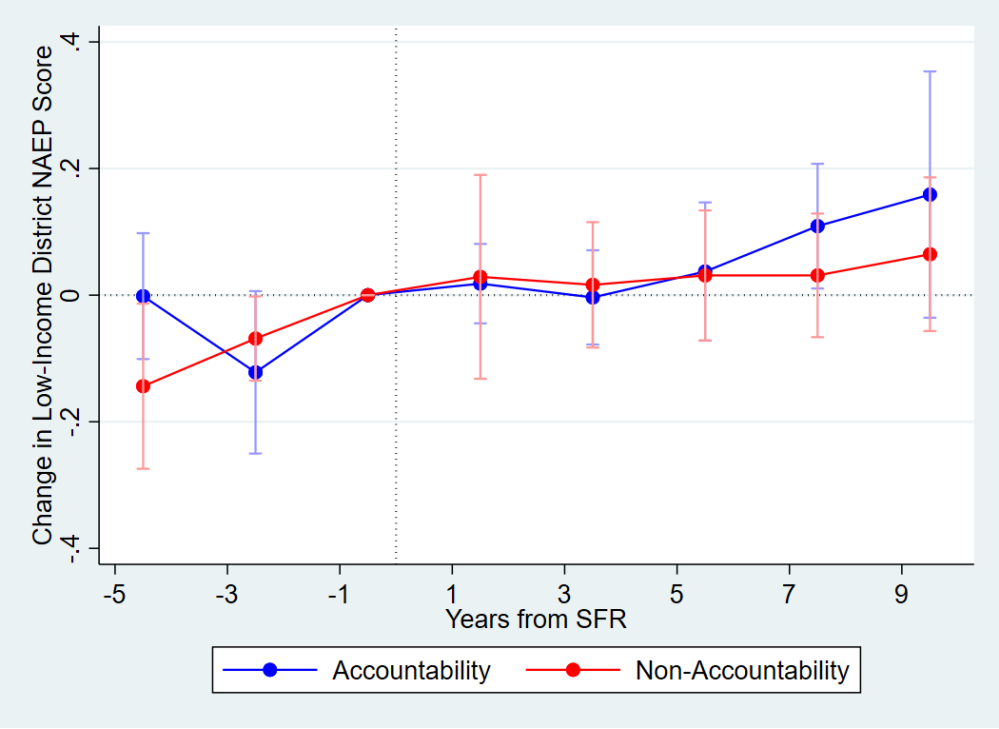
Minorini, P. A. and S. D. Sugarman (1999a). Educational adequacy and the courts: The promise and problems of moving to a new paradigm. *Equity and adequacy in education finance: Issues and perspectives*, 175–208.

Minorini, P. A. and S. D. Sugarman (1999b). School finance litigation in the name of educational equity: Its evolution, impact, and future. *Equity and adequacy in education finance: Issues and perspectives*, 34–71.

Yi, H. and R. C. Feiock (2012). Policy tool interactions and the adoption of state renewable portfolio standards. *Review of Policy Research 29*(2), 193–206.

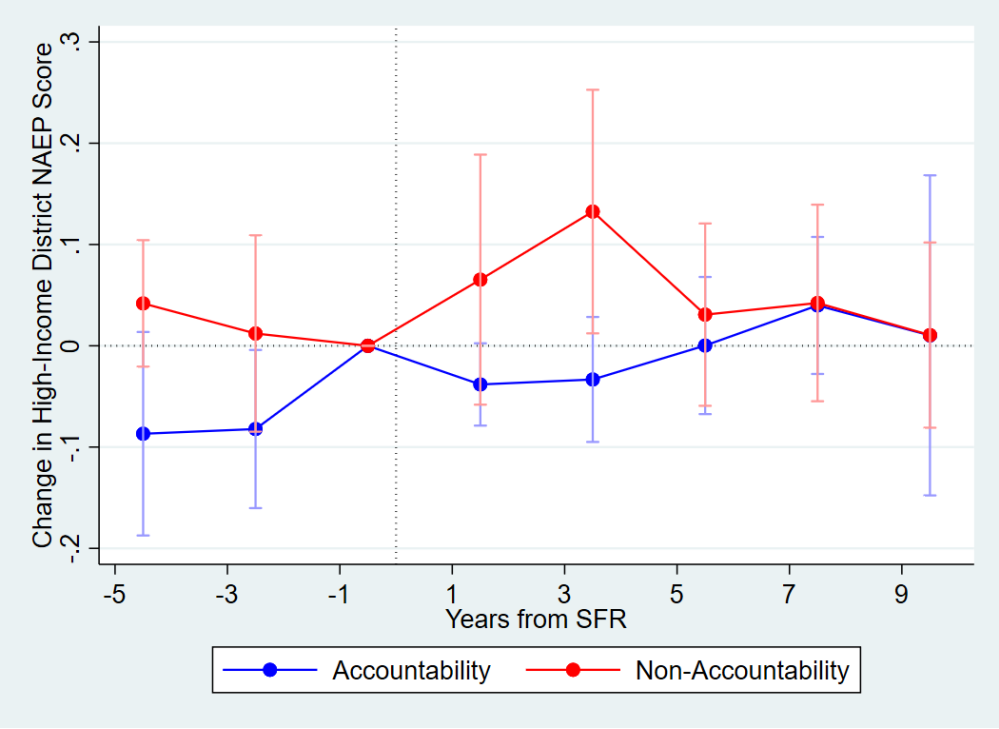
Yinger, J. (2004). *Helping children left behind: State aid and the pursuit of educational equity*. MIT Press.

Figure 1: Estimates of Effects on Low-income District NAEP Score



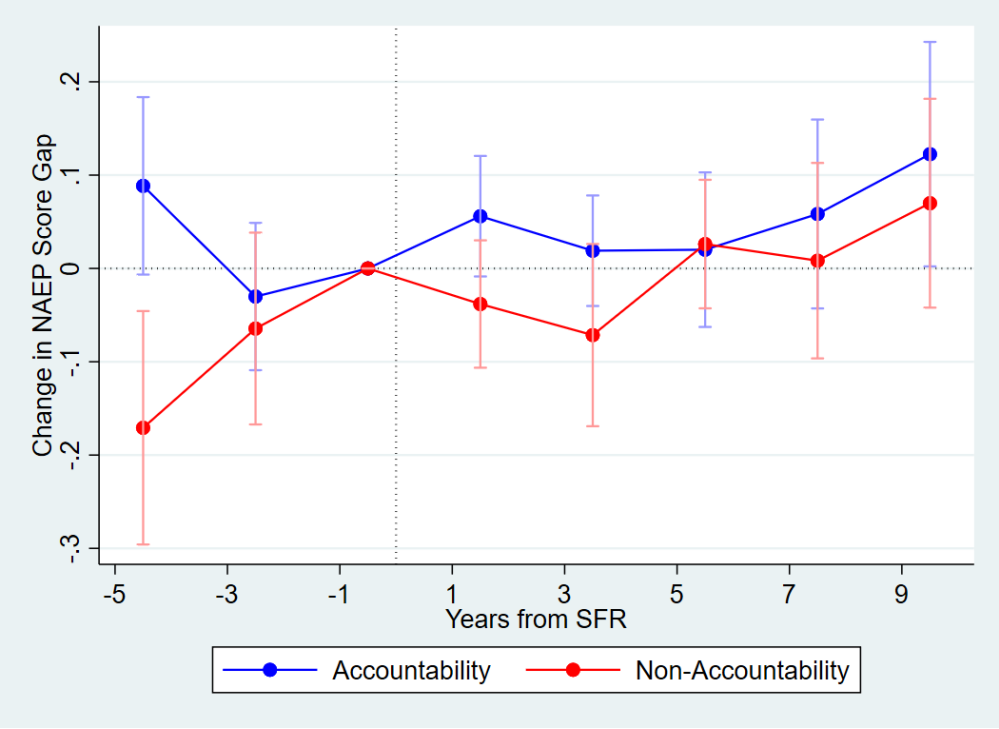
Notes: Figure presents trends in test scores for accountability and non-accountability states in low-income districts by using the event-study framework Equation (3). Blue line represents trends in accountability states, red line represents trends in non-accountability states, and whiskers represent 95% confidence intervals. The specification includes state and subject-grade-year fixed effects and weighted by the inverse squared standard error of the dependent variable. We do not include control states in the sample. 6 years before and 11 years after school finance reforms are calculated but not represented in the figure. Standard errors clustered at the state level.

Figure 2: Estimates of Effects on High-income District NAEP Score



Notes: Figure presents trends in test scores for accountability and non-accountability states in high-income districts by using the event-study framework Equation (3). Blue line represents trends in accountability states, red line represents trends in non-accountability states, and whiskers represent 95% confidence intervals. The specification includes state and subject-grade-year fixed effects and weighted by the inverse squared standard error of the dependent variable. We do not include control states in the sample. 6 years before and 11 years after school finance reforms are calculated but not represented in the figure. Standard errors clustered at the state level.

Figure 3: Estimates of Effects on NAEP Score Gap



Notes: Figure presents trends in gaps of test scores between low- and high-income districts for accountability and non-accountability states by using the event-study framework Equation (3). Blue line represents trends in accountability states, red line represents trends in non-accountability states, and whiskers represent 95% confidence intervals. The specification includes state and subject-grade-year fixed effects and weighted by the inverse squared standard error of the dependent variable. We do not include control states in the sample. 6 years before and 11 years after school finance reforms are calculated but not represented in the figure. Standard errors clustered at the state level.

Table 1: Summary Statistics in 1990

	Treat.	Control	Diff.	Acc.	Non-Acc.	Diff.
Standardized NAEP Score in Low-income	-0.32	-0.08	-0.24**	-0.30	-0.35	0.05
	[0.25]	[0.32]	(0.09)	[0.27]	[0.23]	(0.12)
Standardized NAEP Score in High-income	0.36	0.49	-0.13	0.29	0.46	-0.17*
	[0.22]	[0.37]	(0.10)	[0.22]	[0.21]	(0.10)
Total Revenue p.p in Low-income	8,341	7,935	406	8,994	7,633	1,360
	[2,071]	[1,618]	(540)	[1,680]	[2,288]	(798)
Total Revenue p.p in High-income	9,209	9,001	208	9,886	8,532	1,354
	[3,126]	[2,362]	(811)	[3,435]	[2,763]	(1,273)
Total Expenditure p.p in Low-income	8,363	8,060	303	9,089	7,577	1,512
	[2,055]	[1,782]	(557)	[1,857]	[2,037]	(779)
Total Expenditure p.p in High-income	9,287	9,269	18	9,871	8,704	1,167
	[3,062]	[2,387]	(803)	[3,516]	[2,550]	(1,254)
Pupil Teacher Ratio in Low-income	16.7	16.1	0.5	16.9	16.4	0.4
	[2.7]	[2.1]	(0.7)	[3.1]	[2.2]	(1.1)
Pupil Teacher Ratio in High-income	17.6	17.1	0.5	17.8	17.4	0.5
	[2.3]	[2.5]	(0.7)	[2.6]	[2.0]	(1.0)
Mean Teacher Salary in Low-income	52,760	50,936	1,824	55,346	49,703	5,642
	[10,792]	[12,017]	(3,363)	[12,645]	[7,553]	(4,358)
Mean Teacher Salary in High-income	62,486	58,985	3,501	66,183	58,453	7,731
	[15,203]	[13,330]	(4,270)	[17,548]	[11,639]	(6,273)
Minority Student Share in Low-income	0.18	0.21	-0.03	0.17	0.20	-0.04
	[0.17]	[0.23]	(0.06)	[0.17]	[0.18]	(0.08)
Minority Student Share in High-income	0.10	0.11	-0.02	0.10	0.09	0.01
	[0.09]	[0.11]	(0.03)	[0.11]	[0.05]	(0.04)
Mean Household Income in Low-income	44,956	45,624	-668	46,333	43,464	2,869
	[6,188]	[5,499]	(1,696)	[6,335]	[5,925]	(2,459)
Mean Household Income in High-income	90,462	88,759	1,703	90,097	90,827	-730
	[23,895]	[23,137]	(6,865)	[23,240]	[25,563]	(9,973)
Average Enrollment	5,332	7,464	-2,132	5,814	4,810	1,004
	[5,804]	[9,700]	(2,285)	[7,852]	[2,379]	(2,364)
Total Students	18,234,560	18,234,560		9,019,561	9,214,999	
Number of States	25	23		13	12	

Notes: The entries represent mean of the variables in fiscal year 1990 with standard deviations in bracket and standard errors in parenthesis. "Low-income" corresponds to first quintile districts in each state in terms of household average income in 1990; "high-income" to fifth quintile. NAEP scores in 1990 are for eighth grade math and are only available for 36 states. NAEP variables are weighted by the inverse squared standard error. All finance variables are in 2011 dollars. See Table A1 for which states belong to which category. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: Estimates of Effects of School Finance Reforms on Student Achievement

		Low-income		High-income		Gap	
		(1)	(2)	(3)	(4)	(5)	(6)
Non-Accountability State	Yrs. Elapsed since SFR	0.0060**	0.0044	-0.0029	-0.0021	0.0084**	0.0057
		(0.0030)	(0.0034)	(0.0028)	(0.0032)	(0.0036)	(0.0039)
	Pre-trend		0.0293***		0.0080		0.0236*
			(0.0096)		(0.0089)		(0.0122)
Accountability State	Yrs. Elapsed scine SFR	0.0117**	0.0116**	0.0079	0.0042	0.0026	0.0066*
		(0.0057)	(0.0051)	(0.0061)	(0.0052)	(0.0054)	(0.0038)
	Pre-trend		0.0000		0.0135		-0.0144*
			(0.0134)		(0.0136)		(0.0086)
State fixed effects		Y	Y	Y	Y	Y	Y
Subject-grade-year fixed effects		Y	Y	Y	Y	Y	Y
Observations		1,436	1,436	1,436	1,436	1,434	1,434

Notes: Table presents results of estimating the effects of school finance reforms on student achievement for accountability and non-accountability states. The dependent variable is the weighted mean NAEP score in low-income districts, high-income districts, and gaps between them for columns (1) and (2), columns (3) and (4), and columns (5) and (6), respectively. All specifications include state and subject-grade-year fixed effects, and are weighted by the inverse squared standard error of the dependent variable. Note that columns (2), (4), and (6) do not report estimates for the 5 years and before dummies that are also included. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the state level in parentheses.

Table 3: Sensitivity of Effects of School Finance Reforms on Student Performance

		Low-income	High-income	Gap
Robustness Check 1: Accountability Adoption after School Finance Reform				
Non-Accountability State	Yrs. Elapsed since SFR	0.0046	-0.0005	0.0045
		(0.0035)	(0.0027)	(0.0034)
	Pre-trend	0.0411***	0.0063	0.0369**
		(0.0139)	(0.0072)	(0.0139)
Accountability State	Yrs. Elapsed since SFR	0.0109**	0.0054	0.0050
		(0.0053)	(0.0049)	(0.0041)
	Pre-trend	-0.0034	0.0101	-0.0148
		(0.0127)	(0.0135)	(0.0089)
Robustness Check 2: Timing of School Finance Reform				
Non-Accountability State	Yrs. Elapsed since SFR	-0.0085	-0.0079	0.0040
		(0.0069)	(0.0073)	(0.0068)
	Pre-trend	0.0554***	0.0145*	0.0375***
		(0.0129)	(0.0083)	(0.0130)
Accountability State	Yrs. Elapsed since SFR	0.0123*	0.0066	0.0051
		(0.0072)	(0.0069)	(0.0041)
	Pre-trend	-0.0007	0.0122	-0.0147*
		(0.0123)	(0.0125)	(0.0087)
Year of Reform - 2000		-0.0018**	-0.0011	-0.0001
		(0.0009)	(0.0010)	(0.0008)
Robustness Check 3: Level and Slope Shift				
Non-Accountability State	Yrs. Elapsed since SFR	0.0046	-0.0016	0.0053
		(0.0035)	(0.0034)	(0.0035)
	Pre-trend	0.0410**	-0.0053	0.0444***
		(0.0164)	(0.0065)	(0.0149)
	Level Shift	0.0005	0.0561	-0.0358
		(0.0493)	(0.0457)	(0.0356)
Accountability State	Yrs. Elapsed since SFR	0.0099*	0.0083	0.0012
		(0.0058)	(0.0070)	(0.0055)
	Pre-trend	-0.0067	0.0197	-0.0271
		(0.0102)	(0.0154)	(0.0169)
	Level Shift	0.0178	-0.0510	0.0655
		(0.0309)	(0.0348)	(0.0477)
Accountability control		Y	Y	Y
State fixed effects		Y	Y	Y
Subject-grade-year fixed effects		Y	Y	Y

Notes: Table presents results of sensitivity of the estimated effects of school finance reforms on student achievement for additional controls. The dependent variable is the weighted mean score in low-income districts, high-income districts, and gaps between them. All specifications include accountability control, state and subject-grade-year fixed effects, and are weighted by the inverse squared standard error of the dependent variable. Note that these results do not report estimates for the 5 years and before dummies that are also included. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the state level in parentheses.

Table 4: Estimates of Effects of School Finance Reforms on District Components

	Non-Accountability	Accountability	Observations
Mean			
Log Total Revenue p.p	0.0552*	0.0415**	1,008
	(0.0278)	(0.0186)	
Log Total Expenditure p.p	0.0628*	0.0622**	1,008
	(0.0314)	(0.0237)	
Low-income Districts			
Log Total Expenditure p.p	0.0916***	0.0702***	1,008
	(0.0336)	(0.0240)	
Log Instructional Expenditure p.p	0.0876**	0.0528**	1,008
	(0.0353)	(0.0198)	
Log Teacher Salaries + Benefits p.p	0.0836*	0.0470**	960
	(0.0448)	(0.0204)	
Log Student Support p.p	0.0535*	0.0402*	1,008
	(0.0282)	(0.0233)	
Log Mean Teacher Salary	0.0275	0.0398**	972
	(0.0234)	(0.0153)	
Pupil Teacher Ratio	-0.305*	-0.338*	972
	(0.174)	(0.187)	
Local Revenue Share	-0.0059	-0.0158	1,008
	(0.0138)	(0.0147)	
Subsidized Lunch Share	-0.0555	0.0026	893
	(0.0398)	(0.0087)	
Minority Student Share	0.0089	0.0006	977
	(0.0113)	(0.0068)	

Notes: Table presents the results of estimating the effects of school finance reforms on district components for accountability and non-accountability states. Each row represents a separate regression, where the reported effects correspond to level-shifts post-finance reform. The specification includes state and year fixed effects. Note that these results do not report estimates for pre-trends and the 5 years and before dummies that are also included. All finance variables are in 2011 dollars. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the state level in parentheses.

Appendix

Table A1: States Information

State	SFR	Accountability	Category
Alabama		1997	Control
Arizona	1998	2002 (NCLB)	Non-Accountability
Arkansas	2002	1999	Accountability
California	2004	1999	Accountability
Colorado	2000	2002 (NCLB)	Non-Accountability
Connecticut		1999	Control
Delaware		1998	Control
Florida		1999	Control
Georgia		2000	Control
Idaho	1993	2002 (NCLB)	Non-Accountability
Illinois		1992	Control
Indiana	2011	1995	Accountability
Iowa		2002 (NCLB)	Control
Kansas	2005	1995	Accountability
Kentucky	1990	1995	Non-Accountability
Louisiana		1999	Control
Maine		2002 (NCLB)	Control
Maryland	2002	1999	Accountability
Massachusetts	1993	1998	Non-Accountability
Michigan		1998	Control
Minnesota		2002 (NCLB)	Control
Mississippi		2002 (NCLB)	Control
Missouri	1993	2002 (NCLB)	Non-Accountability
Montana	2005	2002 (NCLB)	Accountability
Nebraska		2002 (NCLB)	Control
Nevada		1996	Control
New Hampshire	2008	2002 (NCLB)	Accountability
New Jersey	1998	2002 (NCLB)	Non-Accountability
New Mexico	1999	1998	Accountability
New York	2006	1998	Accountability
North Carolina	1997	1996	Accountability
North Dakota	2007	2002 (NCLB)	Accountability
Ohio	1997	2002 (NCLB)	Non-Accountability
Oklahoma		1996	Control
Oregon		2000	Control
Pennsylvania		2002 (NCLB)	Control
Rhode Island		1997	Control
South Carolina		1999	Control
South Dakota		2002 (NCLB)	Control

Table A1: States Information

State	SFR	Accountability	Category
Tennessee	1995	2000	Non-Accountability
Texas	1992	1994	Non-Accountability
Utah		2002 (NCLB)	Control
Vermont	2003	1999	Accountability
Virginia		1998	Control
Washington	2010	2002 (NCLB)	Accountability
West Virginia	1995	1997	Non-Accountability
Wisconsin		1993	Control
Wyoming	2001	2002 (NCLB)	Non-Accountability

Notes: The years for school finance reform are based on Lafortune et al. (2018) and the years for the accountability policies are based on Dee and Jacob (2011).

Table A2: Estimates of Effects of School Finance Reforms on Student Achievement: Unweighted

		Low-income	High-income	Gap
Robustness Check: Without Weights				
Non-Accountability State	Yrs. Elapsed since SFR	0.0028	0.0002	0.0030
		(0.0042)	(0.0034)	(0.0034)
	Pre-trend	0.0231**	0.0078	0.0131
		(0.0109)	(0.0115)	(0.0127)
Accountability State	Yrs. Elapsed since SFR	0.0123**	0.0018	0.0110**
		(0.0058)	(0.0061)	(0.0052)
	Pre-trend	0.0020	0.0119	-0.0055
		(0.0096)	(0.0078)	(0.0086)
State fixed effects		Y	Y	Y
Subject-grade-year fixed effects		Y	Y	Y

Notes: Table presents results of the unweighted estimated effects of school finance reforms on student achievement. The dependent variable is the weighted mean score in low-income districts, high-income districts, and gaps between them. All specifications include state and subject-grade-year fixed effects. Note that these results do not report estimates for the 5 years and before dummies that are also included. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the state level in parentheses.