Title
Essays on public service delivery

Author
Sandholtz, Wayne Aaron

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on public service delivery

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Wayne Aaron Sandholtz

Committee in charge:

Professor Eli Berman, Co-Chair
Professor Karthik Muralidharan, Co-Chair
Professor Michael Callen
Professor Craig McIntosh
Professor Paul Niehaus
Professor Agustina Paglayan

2020

The dissertation of Wayne Aaron Sandholtz is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

University of California San Diego

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I gratefully acknowledge my committee members: Professor Eli Berman, Professor Karthik Muralidharan, Professor Michael Callen, Professor Craig McIntosh, Professor Paul Niehaus, and Professor Agustina Paglayan. Individually and collectively, their feedback improved the quality of this work.

Chapter 1, in full, is a reprint of the material as it appears in the American Economic Review 2020. Romero, Mauricio; Sandefur, Justin; Sandholtz, Wayne Aaron. I acknowledge with immense gratitude my coauthors Mauricio Romero and Justin Sandefur, whose support and encouragement was clutch.

Chapter 3, in part is currently being prepared for submission for publication of the material. Leff Yaffe, Daniel; Nakab, Alejandro; Sandholtz, Wayne Aaron. I very gratefully acknowledge my coauthors Daniel Leff Yaffe and Alejandro Nakab, who have taught me much.

VITA

2012	B. S. in Economics magna cum laude, Brigham Young University, Provo
2020	Ph. D. in Economics, University of California San Diego

PUBLICATIONS

Romero, Mauricio; Justin Sandefur; and Wayne Aaron Sandholtz. "Outsourcing Education: Experimental Evidence from Liberia", American Economic Review, 2020.

ABSTRACT OF THE DISSERTATION

Essays on public service delivery

by

Wayne Aaron Sandholtz

Doctor of Philosophy in Economics

University of California San Diego, 2020

Professor Eli Berman, Co-Chair
Professor Karthik Muralidharan, Co-Chair

This dissertation consists of three chapters which relate to public officials' capacity and incentives to improve public services. Chapter 1 examines a setting in which government seeks to augment its capacity by enlisting the private sector. It uses a randomized controlled trial (RCT) to measure the effect of a Liberian school reform which outsourced management of public primary schools to private providers. It finds that outsourced public schools saw learning gains of 0.18 standard deviations in both English and mathematics on average, but the effects varied a lot by private provider. Chapter 2 turns to the question of whether improved public services yield electoral rewards for the officials responsible. It leverages the random variation in learning gains provided by Chapter 1's randomized school reform in Liberia. On average, voters near treated schools

were 3 percentage points less likely to vote for the incumbent party's candidate than those near control schools. This negative average electoral effect, however, masks important heterogeneity. The negative electoral impact of the reform was concentrated in places where the reform reduced children's learning. In places where the reform significantly improved test scores, it also produced electoral rewards. Chapter 3 also seeks to measure the electoral gains to public good provision, focusing on the construction of the Interstate Highway System (IHS) in the United States of America. It uses a shift-share estimator to isolate exogenous variation in the timing of IHS construction by county. It finds that a mile of IHS construction in an election year increases county-level vote share for the incumbent governor's party by 0.6-2.2 percentage points during the period 1950-1972. These essays provide new empirical evidence of democratic accountability for public service provision, even as they illuminate directions for future research into the conditions in which democratic accountability binds.

Chapter 1

Outsourcing Education: Experimental Evidence from Liberia

with Mauricio Romero and Justin Sandefur

In 2016, the Liberian government delegated management of 93 randomly-selected public schools to private providers. Providers received USD 50 per pupil, on top of USD 50 per pupil annual expenditure in control schools. After one academic year, students in outsourced schools scored 0.18σ higher in English and mathematics. We do not find heterogeneity in learning gains or enrollment by student characteristics, but there is significant heterogeneity across providers. While outsourcing appears to be a cost-effective way to use new resources to improve test scores, some providers engaged in unforeseen and potentially harmful behavior, complicating any assessment of welfare gains.

1.1 Introduction

Governments often enter into public-private partnerships as a means to raise capital or to leverage the efficiency of the private sector (World Bank, 2015b). But contracts are inevitably incomplete, and thus contracting out the provision of public services to private providers will have theoretically ambiguous impacts on service quality (Hart, Shleifer, & Vishny, 1997; Holmstrom &

Milgrom, 1991). While private contractors may face stronger incentives for cost efficiency than civil servants, they may also cut costs through actions that are contractually permissible but not in the public interest.

In this paper we study the Partnership Schools for Liberia (PSL) program, which delegated management of 93 public schools (3.4% of all public primary schools, serving 8.6% of students enrolled in public primary or pre-school) to eight different private organizations. Providers received an additional USD 50 per pupil as part of the program, on top of the yearly USD 50 per pupil expenditure in control schools, and some providers independently raised and spent far more. PSL schools also negotiated successfully for more government teachers: They had an average of one teacher per grade, compared to 0.78 teachers per grade in traditional public schools. In exchange, providers were responsible for the daily management of the schools. These schools were to remain free and non-selective (i.e., providers were not allowed to charge fees or screen students based on ability or other characteristics). PSL school buildings remained under the ownership of the government. Teachers in PSL schools were civil servants, drawn from the existing pool of government teachers.

We study the impact of this program by randomly assigning existing public schools to be managed by a private provider. We paired schools (based on infrastructure and geography), then assigned pairs to providers, and subsequently randomly assigned treatment within each matched pair. Thus, we are able to estimate both the average impact of the PSL program as well as treatment effects across providers. Since treatment assignment may change the student composition across schools, we sampled students from pre-treatment enrollment records. We associate each student with their "original" school, regardless of what school (if any) they attend in later years. The combination of random assignment of treatment at the school level with sampling from a fixed and comparable pool of students allows us to provide clean estimates of the program's intention-to-treat (ITT) effect on test scores, uncontaminated by selection effects.

The ITT effect on student test scores after one year of the program is 0.18σ for English and 0.18σ for mathematics. These gains do not reflect teaching to the test, as they are also seen in new questions administered only at the end of the school year and in questions with a new format. Taking into account that some providers refused to work in some schools randomly assigned to them and some students moved schools, the treatment effect on the treated (ToT) after one year of the program is 0.21σ for English test scores and 0.22σ for mathematics.[1] We find no evidence of heterogeneity by students' socio-economic status, gender, or grade, suggesting that efficiency gains need not come at the expense of equity concerns. There is also no evidence that providers engaged in student selection: The probability of remaining in a treatment school is unrelated to age, gender, household wealth, or disability.

These gains in test scores reflect a combination of additional inputs and improved management. As a lower bound, the program spent an additional USD 50 per pupil, which was the government's budget target for PSL and the transfer made to operators. While some operators spent more than this, others reported spending near this amount. When the cost of additional teachers is included the cost rises to approximately USD 70 per student, and when the actual cost reported by providers for the first year is included the average increases to USD 238 (see Section 1.2.1 for details). The program also increased management quality, as proxied by teacher time on task. Teachers in PSL schools were 50% more likely to be in school during a spot check (a 20-percentage-point increase, from a base of 40%) and 43% more likely to be engaged in instruction during class time (a 15-percentage point increase, from a base of 35%). Teacher attendance and time on task improved for incumbent teachers, which we interpret as evidence of better management.

Since each provider was assigned schools in a matched-pair design, we are able to estimate (internally valid) treatment effects for each provider. While the assignment of treatment within matched pairs was random, the assignment of pairs to providers was not, resulting in non-random differences in schools and locations across providers. Therefore, the raw treatment effects for each individual provider are internally valid but they are not comparable without further assumptions

[1] Consistent with the design of the experiment, we focus on the ITT effect. The ToT is estimated using the assigned treatment as an instrument for whether the student is in fact enrolled in a PSL school during the 2016/2017 academic year. The percentage of students originally assigned to treatment schools who are actually in treatment schools at the end of the 2016/2017 schools year is 81%. The percentage of students assigned to control schools who are in treatment schools at the end of the 2016/2017 schools year is 0%.

(see Section 1.4 for more details). In the Appendix, we also present treatment effects adjusting for baseline differences and "shrinking" the estimates using a Bayesian hierarchical model — with qualitatively similar results. While the highest-performing providers generated increases in learning of over 0.36σ, the lowest-performing providers had no impact on learning. The group of highest-performing providers includes both the highest spender and some of the lowest-cost organizations. These results suggest that higher spending by itself is neither necessary nor sufficient for improving learning outcomes.[2]

Turning to whether PSL is a good use of scarce funds, we make two comparisons: a comparative cost-effectiveness calculation comparing PSL to business-as-usual expansion of Liberia's public school system, and a cost-benefit calculation based on the net present value of the Mincerian earnings returns to the education provided by PSL. Both calculations require strong assumptions (Dhaliwal, Duflo, Glennerster, & Tulloch, 2013), which we discuss in Section 1.5. While some providers incurred larger costs in the first year, assuming all providers will eventually reach the budget target of USD 50 per pupil implies that the program can increase test scores for treated students by 0.44σ per USD 100 spent. We estimate this yields a positive net present value for the program investment after considering the income gains associated with schooling, and is more cost-effective than additional spending under business-as-usual.

However, test score gains and expenditures fail to tell the entire story of the consequences of this public-private partnership. Some providers took unforeseen actions that may be socially undesirable. While the contract did not allow cream-skimming, it did not prohibit providers from capping enrollment in oversubscribed schools or from shifting underperforming teachers to other schools.[3] While most providers kept students in oversubscribed schools and retained existing teachers, one provider did not. This provider, Bridge International Academies, removed pupils after

[2]See Hanushek and Woessmann (2016) for a review on how school resources affect academic achievement.

[3]In principle, removing underperforming teachers could be positive for the school system. In practice, dismissed teachers ended up either teaching at other public schools or receiving pay without work (as firing public teachers was almost impossible). Reshuffling teachers is unlikely to raise average performance in the system as a whole, and Liberia already has a tight budget and short supply of teachers (the literacy rate is below 50%). Similarly, reducing class sizes may be good policy, but shifting students from PSL schools to other schools is unsustainable and may lead us to overstate the scalable impact of the program. While the experiment was designed to overcome any bias from student reallocation and we can track teacher reallocations, it is not designed to measure negative spillovers.

taking control of schools with large class sizes, and removed 74% of incumbent teachers from its schools.

More worryingly, news media have revealed serious sexual abuse scandals involving two of the private providers — one of them a US-based non-profit that was well regarded by the international community. Over the course of multiple years prior to the launch of the program and this study, a More than Me employee, who died of AIDS in 2016, raped over 30 girls in a More than Me school (Young, 2018).[4] In 2016, the Board Chair of the Liberian Youth Network (the previous name for the Youth Movement for Collective Action) was found guilty of raping a teenage boy (Baysah, 2016). It is possible that similar scandals take place in regular schools but that these were uncovered due to the heightened scrutiny of the public-private partnership. But at a minimum it shows that private providers are far from an obvious solution to sexual violence issues in public schools.

Some of these issues could arguably have been solved with more complete contracts or better partner selection. The first year was a pilot and a learning year, and the government deliberately tried to select "mission aligned" contractors and left the contracts quite open. However, some of the providers engaged in the worst behavior were considered some of the most promising. These events underscore the challenge of ensuring that private providers act in the public interest in a world of incomplete contracts. Thus, our results suggest that outsourcing has some promising features, but also presents its own set of difficulties.

We make several contributions to both research and policy. Proponents of outsourcing in education argue that combining public finance with private management has the potential to overcome a trade-off between efficiency and equity (Patrinos, Barrera-Osorio, & Guáqueta, 2009). On the efficiency side, private schools tend to be better managed than their public counterparts (N. Bloom, Lemos, Sadun, & Van Reenen, 2015; Muralidharan & Sundararaman, 2015). On the equity side, fee-charging private schools may increase inequality and induce socio-economic strati-

[4]Note that while these incidents occurred prior to the launch of the program, they were revealed in full only after the program launched, which enabled More than Me to dramatically expand its operations. The exhaustive investigation by Young (2018) exposes two wrongs. One is the systematic rape of Liberian children. The other is the refusal of More than Me's leadership to accept responsibility, and their (successful) efforts to conceal the case from public scrutiny.

fication in education (Hsieh & Urquiola, 2006; Lucas & Mbiti, 2012; Zhang, 2014). Thus, in theory, publicly-financed but privately-managed schools may increase efficiency without compromising equity. Most of the empirical evidence to date on outsourcing education comes from the U.S., where charter schools appear to improve learning outcomes when held accountable by a strong commissioning body (Cremata et al., 2013; Woodworth et al., 2017). However, there is limited evidence on whether private administration of public schools can improve learning outcomes in developing countries, where governments tend to have limited capacity to write complete contracts and enforce them. Two noteworthy studies which examine close analogs to PSL in the U.S. are Abdulkadiroğlu, Angrist, Hull, and Pathak (2016) who study charter takeovers (where traditional public schools are restarted as charter schools, similar to our setting) in Boston and New Orleans and Fryer (2014) who studies the implementation of a bundle of best practices from high-performing charter schools into low-performing, traditional public schools in Houston, Texas. In line with our results, both studies find increases in test scores. We provide some of the first experimental estimates on contracting out management of existing public schools in a developing country.[5]

An additional contribution is related to our experimental design and the treatment effects we are able to identify. Most U.S. studies use admission lotteries to overcome endogeneity issues (for a review see Chabrier, Cohodes, and Oreopoulos (2016); Betts and Tang (2014)). But oversubscribed charter schools are different (and likely better) than undersubscribed ones, truncating the distribution of estimated treatment effects (Tuttle, Gleason, & Clark, 2012). We provide treatment effects from across the distribution of outsourced schools, and across the distribution of students within a school. Relatedly, relying on school lotteries implies that the treatment estimates capture the joint impact of outsourcing *and* oversubscribed schools' providers. We provide treatment effects across a list of providers, vetted by the government, and show that the provider matters.

Finally, we contribute to the broader literature on outsourcing service delivery. Hart et al. (1997) argue that the bigger the adverse consequences of non-contractible quality shading, the

[5]For a review on the few existing non-experimental studies see Aslam, Rawal, and Saeed (2017). A related paper to ours increased the supply of schools through a public-private partnership in Pakistan (Barrera-Osorio et al., 2017). However, it is difficult to disentangle the effect of increasing the supply of schools from the effect of privately managed but publicly funded schools.

stronger the case for governments to provide services directly. Empirically, in cases where quality is easy to measure and to enforce, such as water services (Galiani, Gertler, & Schargrodsky, 2005) or food distribution (Banerjee, Hanna, Kyle, Olken, & Sumarto, 2019), outsourcing seems to work. Similarly, for primary health care, where quality is measurable (e.g., immunization and antenatal care coverage), outsourcing improves outcomes in general (Loevinsohn & Harding, 2005; E. Bloom et al., 2007). In contrast, for services whose quality is difficult to measure, such as prisons (Useem & Goldstone, 2002; Cabral, Lazzarini, & de Azevedo, 2013), outsourcing seems to be detrimental. In contrast to primary health care, there is some evidence that contracting out advanced care (where quality is harder to measure) increases expenditure without increasing quality (Duggan, 2004). Some quality aspects of education are easy to measure (e.g., enrollment and basic learning metrics), but others are harder (e.g., socialization and selection). In our setting, while outsourcing management improves most indices of school quality on average, the effect varies across providers. In addition, some providers' actions had negative unintended consequences and may have generated negative spillovers for the broader education system, underscoring the importance of robust contracting and monitoring for this type of program.

1.2 Research design

1.2.1 The program

Context

The PSL program breaked new ground in Liberia by delegating management of government schools and employees to private providers. Nonetheless, private actors — such as NGOs and USAID contractors — are already common in government schools. Over the past decade, Liberia's basic education budget has been roughly USD 40 million per year (about 2-3% of GDP), while external donors contribute about USD 30 million. This distinguishes Liberia from most other low-income countries in Africa, which finance the vast bulk of education spending through domestic tax revenue (UNESCO, 2016). The Ministry spends roughly 80% of its budget on teacher salaries (Ministry of Education - Republic of Liberia, 2017a), while almost all the aid money bypasses the Ministry, flowing instead through an array of donor contractors and NGO programs covering non-salary expenditures. For instance, in 2017 USAID tendered a USD 28 million education program to be implemented by a U.S. contractor in public schools over a five year period (USAID, 2017). The net result is that many "public" education services in Liberia, beyond teacher salaries, are provided by non-state actors. On top of that, more than half of children enrolled in preschool and primary attend private schools (Ministry of Education - Republic of Liberia, 2016a).

A second broad feature of Liberia's education system, relevant for the PSL program, is its performance — not only are learning levels low, but access to basic education and progression through school remains inadequate. The Minister of Education has cited the perception that "Liberia's education system is in crisis" as the core justification for the PSL program (Werner, 2017). While the world has made great progress towards universal primary education in the past three decades (worldwide net enrollment was almost 90% in 2015), Liberia has been left behind. Net primary enrollment stood at only 38% in 2014 (World Bank, 2014). Low *net* enrollment is partially explained by an extraordinary backlog of over-age children due to the civil war (see Figure 1.4 in Appendix 1.7.1): The median student in early childhood education is eight years old and over 60% of 15 years olds are still enrolled in early childhood or primary education (Liberia Institute of Statistics and Geo-Information Services, 2016). Learning levels are low: Only 25% of adult women (there is no information for men) who finish elementary school can read a complete sentence (Liberia Institute of Statistics and Geo-Information Services, 2014).

Intervention

The Partnership Schools for Liberia (PSL) program is a public-private partnership (PPP) for school *management*. The Government of Liberia contracted multiple non-state providers to run ninety-three existing public primary and pre-primary schools. There are nine grades per school: three early childhood education grades (Nursery, K1, and K2) and six primary grades (grade 1 -

grade 6). Providers receive funding on a per-pupil basis. In exchange they are responsible for the daily management of the schools.

The government allocated rights to eight providers to manage public schools under the PSL program. The organizations are as follows: Bridge International Academies (23 schools), BRAC (20 schools), Omega Schools (19 schools), Street Child (12 schools), More than Me (6 schools), Rising Academies (5 schools), Youth Movement for Collective Action (4 schools), and Stella Maris (4 schools). See Appendix 1.7.5 for more details about each organization.

Rather than attempting to write a complete contract specifying private providers' full responsibilities, the government opted instead to select organizations it deemed aligned with its mission of raising learning levels (i.e., "mission-matching" à la Besley and Ghatak (2005); Akerlof and Kranton (2005)). After an open and competitive bidding process led by the Ministry of Education with the support of the Ark Education Partnerships Group, the Liberian government selected seven of the eight organizations listed above, of which six passed financial due diligence. Stella Maris did not complete this step and, although included in our sample, was never paid. While Stella Maris never actually took control of their assigned schools, the government still considers them part of the program (e.g., they were allocated more schools in an expansion of the program not studied in this paper (Ministry of Education - Republic of Liberia, 2017b)). The government made a separate agreement with Bridge International Academies (not based on a competitive tender), but also considers Bridge part of the PSL program.

PSL schools remain public schools and all grades are required to be free of charge and non-selective (i.e., providers are not allowed to charge fees or to discriminate in admissions). In contrast, traditional public schools are not free for all grades. Public primary education is nominally free starting in Grade 1, but tuition for early childhood education in traditional public schools is stipulated at LBD 3,500 per year (about USD 38).

PSL school buildings remain under the ownership of the government. Teachers in PSL schools are civil servants, drawn from the existing pool of government teachers. The Ministry of Education's financial obligation to PSL schools is the same as all government-run schools: It provides teachers and maintenance, valued at about USD 50 per student. A noteworthy feature of PSL is that providers receive *additional* funding of USD 50 per student (with a maximum of USD 3,250 or 65 students per grade). Donors paid for the transfers made to providers in the first year. Donor money was attached to the PSL program and would not have been available to the government otherwise. Neither Bridge International Academies nor Stella Maris received the extra USD 50 per pupil. As mentioned above, Stella Maris did not complete financial due diligence. Bridge International Academies had a separate agreement with the Ministry of Education and relied entirely on direct grants from donors. Providers have complete autonomy over the use of these funds (e.g., they can be used for teacher training, school inputs, or management personnel).[6] On top of that, providers may raise more funds on their own.

Providers must teach the Liberian national curriculum, but may supplement it with remedial programs, prioritization of subjects, longer school days, and non-academic activities. They are welcome to provide more inputs such as extra teachers, books or uniforms, as long as they pay for them.

The intended differences between treated (PSL) and control (traditional public) schools are summarized in Table 1.1. First, PSL schools are managed by private organizations. Second, PSL schools are theoretically guaranteed (as per the contract) one teacher per grade in each school, plus extra funding. Third, private providers are authorized to cap class sizes. Finally, while both PSL and traditional public schools are free for primary students starting in first grade, public schools charge early-childhood education (ECE) fees.

[6]Providers may spend funds hiring more teachers (or other school staff); thus is possible that some of the teachers in PSL schools are not civil servants. However, this rarely occurred. Only 8% of teachers in PSL schools were paid by providers at the end of the school year. Informal interviews with providers indicate that in most cases providers are paying these salaries while awaiting placement of the teachers on the government payroll. Providers expect to be reimbursed by the government once this occurs.

Table 1.1: Policy differences between treatment and control schools

	Control schools	PSL treatment schools
Management		
Who owns school building?	Government	Government
Who employs and pays teachers?	Government	Government
Who manages the school and teachers?	Government	Provider
Who sets curriculum?	Government	Government + provider supplement
Funding		
Primary user fees (annual USD)	Zero	Zero
ECE user fees (annual USD)	USD 38	Zero
Extra funding per pupil (annual USD)	NA	USD 50[a]+ independent fund-raising
Staffing		
Pupil-teacher ratios	NA	Promised one teacher per grade, allowed to cap class sizes at 45-65 pupils[b]
New teacher hiring	NA	First pick of new teacher-training graduates[c]

[a] Neither Bridge International Academies nor Stella Maris received the extra USD 50 per pupil.
[b] Bridge International Academies was authorized to cap class sizes at 55 (but in practice capped them at 45 in most cases as this was allowed by the MOU), while other providers were authorized to cap class sizes at 65.
[c] Bridge International Academies had first pick, before other providers, of the new teacher-training graduates.

What do providers do?

Providers enjoy considerable flexibility in defining the intervention. They are free to choose their preferred mix of, say, new teaching materials, teacher training, and managerial oversight of the schools' day-to-day operations. Rather than relying on providers' own description of their model — where there may be incentives to exaggerate and activities may be defined in non-comparable ways across providers — we administered a survey module to teachers in treatment schools, asking if they had heard of the provider, and if so, what activities the provider had engaged in. We summarize teachers' responses in Figure 1.1, which shows considerable variation in the specific activities and the total activity level of providers.

For instance, teachers reported that two providers (Omega and Bridge) provided computers to schools, which fits with the stated approach of these two international, for-profit firms. Other providers, such as BRAC and Street Child, put more focus on teacher training and observing teachers in the classroom, though these differences were not dramatic. In general, providers such as

More than Me and Rising Academies showed high activity levels across dimensions, while teacher surveys confirmed administrative reports that Stella Maris conducted almost no activities in its assigned schools.

		Provider							
		Stella M	YMCA	Omega	BRAC	Bridge	Rising	St. Child	MtM
Provider Support	Provider staff visits at least once a week(%)	0	54	13	93	76	94	91	96
	Heard of PSL(%)	42	85	61	42	87	90	68	85
	Heard of (provider)(%)	46	96	100	95	100	100	100	100
	Has anyone from (provider) been to this school?(%)	42	88	100	94	100	100	99	100
Ever provided	Textbooks(%)	12	96	73	94	99	71	94	96
	Teacher training(%)	0	77	62	85	87	97	93	96
	Teacher received training since Aug 2016(%)	23	46	58	45	50	81	58	37
	Teacher guides (or teacher manuals)(%)	0	69	75	54	97	94	68	98
	School repairs(%)	0	12	25	24	53	52	13	93
	Paper(%)	0	92	30	86	70	97	88	98
	Organization of community meetings(%)	0	54	27	69	73	87	83	91
	Food programs(%)	0	8	2	1	1	10	0	17
	Copybooks(%)	4	65	30	92	18	97	94	91
	Computers, tablets, electronics(%)	0	0	94	0	99	3	3	2
Most recent visit	Provide/deliver educational materials(%)	0	4	45	17	18	26	29	50
	Observe teaching practices and give suggestions(%)	0	19	45	81	65	45	74	85
	Monitor/observe PSL program(%)	0	12	23	11	13	13	35	65
	Monitor other school−based government programs(%)	0	0	7	5	10	6	18	9
	Monitor health/sanitation issues(%)	0	8	9	2	5	0	10	28
	Meet with PTA committee(%)	0	12	8	10	7	0	21	41
	Meet with principal(%)	0	12	54	36	38	6	51	63
	Deliver information(%)	0	12	36	16	8	6	16	35
	Check attendance and collect records(%)	42	23	43	56	39	19	66	70
	Ask students questions to test learning(%)	4	4	24	33	18	58	44	43

Figure 1.1: What did providers do?

The figure reports simple proportions (not treatment effects) of teachers surveyed in PSL schools who reported whether or not the provider responsible for their school had engaged in each of the activities listed. The sample size, *n*, of teachers interviewed with respect to each provider is: Stella Maris, 26; Omega, 141; YMCA, 26; BRAC, 170; Bridge, 157; Street Child, 80; Rising Academy, 31; More than Me, 46. This sample only includes compliant treatment schools.

Cost data and assumptions

The government designed the PSL program based on the estimate that it spends roughly USD 50 per child in all public schools (mostly on teacher salaries), and it planned to continue to do so in PSL schools (Werner, 2017). As shown in Section 2.4, PSL led to reallocation of additional teaching staff to treatment schools and reduced pupil-teacher ratios in treatment schools, raising the Ministry's per-pupil cost to close to USD 70. On top of this, providers were offered a USD 50 per-pupil payment to cover their costs. As noted above, neither Bridge International Academies nor Stella Maris received the extra USD 50 per pupil. This cost figure was chosen because USD 100 was deemed a realistic medium-term goal for public expenditure on primary education nationwide (Werner, 2017).

In the first year, some providers spent far more than this amount. *Ex ante* per-pupil budgets submitted to the program secretariat before the school year started (on top of the Ministry's costs) ranged from a low of approximately USD 57 for Youth Movement for Collective Action to a high of USD 1,050 for Bridge International Academies (see Figure 1.2a). *Ex post* per-pupil expenditure submitted to the evaluation team at the end of the school year (on top of the Ministry's costs) ranged from a low of approximately USD 48 for Street Child to a high of USD 663 for Bridge International Academies (see Figure 1.2b). These differences in costs are large relative to differences in treatment effects on learning, implying that cost-effectiveness may be driven largely by cost assumptions.

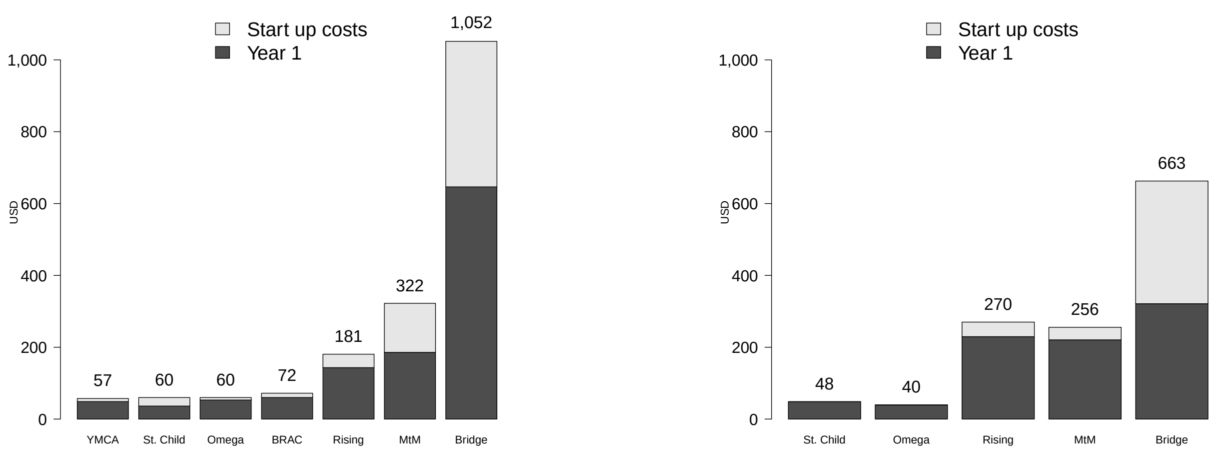
In principle, the costs incurred by private providers would be irrelevant for policy evaluation in a public-private partnership with this structure. If the providers are willing to make an agreement in which the government pays USD 50 per pupil, providers' losses are inconsequential to the government (philanthropic donors have stepped in to fund some providers' high costs under PSL).[7] Thus we present analyses using both the Ministry's USD 50 long-term cost target and providers' actual budgets.[8]

Providers' budgets for the first year of the program are likely a naïve measure of program

[7]These costs matter to the government under at least two scenarios. First, if providers are spending more during the first years of the program to prove effectiveness, they may lower expenditure (and quality) once they have locked in long-term contracts. Second, if private providers are not financially sustainable, they may close schools and disrupt student learning.

[8]While some providers relied almost exclusively on the USD 50 per child subsidy from the PSL pool fund, others have raised more money from donors. Bridge International Academies relied entirely on direct grants from donors and opted not to take part in the competitive bidding process for the USD 50 per pupil subsidy which closed in June 2016. Bridge did subsequently submit an application for this funding in January 2017, which was not approved, but allows us access to their budget data.

cost, as they combine start-up costs, fixed costs, and variable costs. It is possible to distinguish start-up costs from other costs as shown in Figure 1.2, and these make up a small share of the first-year totals for most providers. It is not possible to distinguish fixed from variable costs in the budget data. In informal interviews, some providers (e.g., Street Child) profess operating a variable-cost model, implying that each additional school costs roughly the same amount to operate. Others (e.g., Bridge) report that their costs are almost entirely fixed, and unit costs would fall if scaled; however, we have no direct evidence of this. Our estimate is that Bridge's international operating cost, at scale, is between USD 191 and USD 220 per pupil annually.[9]



(a) Ex ante budget per pupil

(b) Ex post cost per pupil

Figure 1.2: Budget and costs as reported by providers

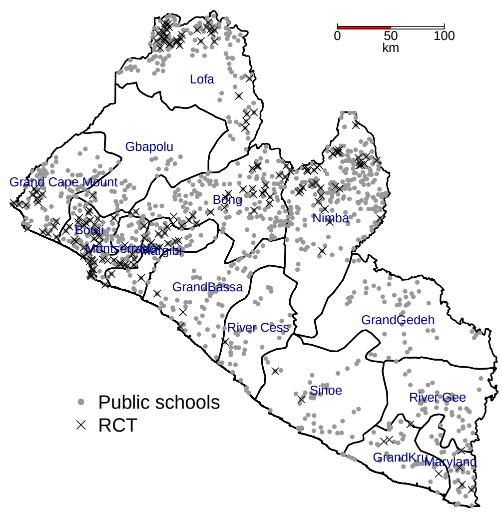
Note: Numbers in 1.2a are based on ex-ante budgets submitted to the program secretariat in a uniform template (inclusive of both fixed and variable costs). Stella Maris did not provide budget data. Numbers in 1.2b are based on self-reports on ex-post expenditures (inclusive of both fixed and variable costs) submitted to the evaluation team by five providers in various formats. Numbers do not include the cost of teaching staff borne by the Ministry of Education.

[9] In written testimony to the UK House of Commons, Bridge stated that its fees were between USD 78 and USD 110 per annum in private schools, and that it had approximately 100,000 students in both private and PPP schools (Bridge International Academies, 2017; Kwauk & Robinson, 2016). Of these, roughly 9,000 are in PPP schools and pay no fees. In sworn oral testimony, co-founder Shannon May stated that Bridge had supplemented its fee revenue with more than USD 12 million in the previous year (May, 2017). This is equal to an additional USD 120 per pupil, and implies Bridge spends between USD 191 and USD 220 per pupil at its current global scale.

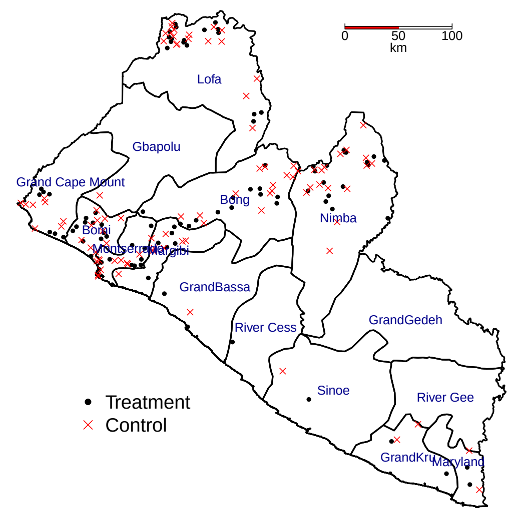
1.2.2 Experimental design

Sampling and random assignment

Liberia has 2,619 public primary schools. Private providers and the government agreed that potential PSL schools should have at least six classrooms and six teachers, good road access, a single shift, and should not contain a secondary school on their premises. A few schools were added to the list at the request of Bridge International Academies. Some of these schools had double shifts. Only 299 schools satisfied all the criteria, although some of these are "soft" constraints that can be addressed if the program expands. For example, the government can build more classrooms and add more teachers to the school staff. On average, schools in the experiment are closer to the capital (Monrovia), have more students, greater resources, and better infrastructure. While schools in the RCT generally have better facilities and infrastructure than most schools in the country, they still have deficiencies. For example, the average school in Liberia has 1.8 permanent classrooms — the median school has zero permanent classrooms — while the average school in the RCT has 3.16 classrooms. Figure 1.3a shows all public schools in Liberia and those within our sample. Table 1.11 in Appendix 1.7.1 has details on the differences between schools in the experiment and other public schools.



(a) Geographical distribution of all public schools in Liberia and those within the RCT.



(b) Geographical distribution of treatment and control schools.

Figure 1.3: Public primary schools in Liberia

Note: Data on school location is from Ministry of Education - Republic of Liberia (2015-2016) data. Geographical information from the administrative areas of Liberia comes from DIVA-GIS (2016) .

Two providers, Omega Schools and Bridge International Academies, required schools with 2G connectivity. Each provider submitted to the government a list of the regions they were willing to work in (Bridge International Academies had first pick of schools). Based on preferences and requirements the list of eligible schools was partitioned across providers. We paired schools in the experiment sample within each district according to a principal component analysis (PCA) index of school resources.[10] This pairing stratified treatment by school resources within each private provider, but not across providers. We gave a list of pairs to each provider based on their location preferences and requirements, so that each list had twice the number of schools they were to operate. Once each provider approved this list, we randomized the treatment assignment within each pair. There is one triplet due to logistical constraints in the assignment across counties, which resulted in one extra treatment school. In short, schools are assigned to a provider, then paired, and then randomly assigned to treatment or control.

[10]We calculated the index using the first eigenvector of a principal component analysis that included the following variables: students per teacher; students per classroom; students per chair; students per desk; students per bench; students per chalkboard; students per book; whether the school has a permanent building; whether the school has piped water, a pump or a well; whether the school has a toilet; whether the school has a staff room; whether the school has a generator; and the number of enrolled students.

Private providers did not manage all the schools originally assigned to treatment and we treat these schools as non-compliant, presenting results in an intention-to-treat framework. After providers visited their assigned schools to start preparing for the upcoming school year, two treatment schools turned out to be private schools that were incorrectly labeled in the government data as public schools. Two other schools had only two classrooms each. Of these four schools, two had originally been assigned to More Than Me and two had been assigned to Street Child. Omega Academies opted not to operate two of their assigned schools and Rising Academies opted not to operate one of their assigned schools. In total, there are 7 non-compliant treatment schools.[11] Figure 1.3b shows the treatment assignment.

Treatment assignment may change the student composition across schools. To prevent differences in the composition of students from driving differences in test scores, we sampled 20 students per school (from K1 to grade 5) from enrollment logs from 2015/2016, the year before the treatment was introduced. We associate each student with his or her "original" school, regardless of what school (if any) he or she attended in subsequent years. The combination of random treatment assignment at the school level with measuring outcomes of a fixed and comparable pool of students allows us to provide clean estimates of the program's intention-to-treat (ITT) effect on test scores within the student population originally attending study schools, uncontaminated by selection.

Timeline of research and intervention activities

We collected data in schools twice: At the beginning of the school year in September/October 2016 and at the end of the school year in May/June 2017.[12] We collected the first round of data 2

[11] More than Me and Street Child were provided with replacement schools, presenting them with a new list of counterparts and informing them, as before, that they would operate one of each pair of schools (but not which one). Providers approved the list before we randomly assigned replacement schools from it. However, we do not use this list as our main sample since it is not fully experimental. We analyzed results for this "final" treatment and control school list, and they are almost identical to the results for the "original" list. Results for this final list of treatment and control schools are available upon request. Bridge International Academies is managing two extra demonstration schools that were not randomized and are not part of our sample. Rising Academies was given one non-randomly assigned school, which is not part of our sample either. Thus, the set of schools in our analysis is not identical to the set of schools actually managed by PSL providers. Table 1.12 summarizes the overlap between schools in our main sample and the set of schools actually managed by PSL providers.

[12] A third round of data collection will take place in March/April 2019 conditional on continuation of the project and preservation of the control group (see Figure 1.5 in Appendix 1.7.1 for a detailed timeline of intervention and research

to 8 weeks after the beginning of treatment. While we intended the first survey wave to serve as a baseline, logistical delays led it to take place shortly after the beginning of the school year. We see evidence of treatment effects within this 1-2 month time frame and treat this early wave as a very short-term outcome survey. Hence, we do not control for test scores collected during the first wave of data collection.[13] We focus on time-invariant covariates and administrative data collected before the program began when checking balance between treatment and control schools (see Section 1.2.2).

Test design

In our sample, literacy cannot be assumed at any grade level, precluding the possibility of written tests. We opted to conduct one-on-one tests in which an enumerator sits with the student, asks questions, and records the answers. In addition, purely school-based tests would be contaminated by shifts in enrollment and attendance due to treatment. For the math part of the test we provided students with scratch paper and a pencil. We designed the tests to capture a wide range of student abilities. To make the test scores comparable across grades, we constructed a single adaptive test for all students. The test has stop rules that skip higher-order skills if the student is not able to answer questions related to more basic skills. Appendix 1.7.3 has details on the construction of the test.

We estimate an item response theory (IRT) model for each round of data collection. IRT models are the standard in the assessments literature for generating comparative test scores.[14]

activities).

[13]Our pre-analysis plan was written on the assumption we would be able to collect baseline data (Romero, Sandefur, & Sandholtz, 2017). Hence, the pre-analysis plan includes a specification that controls for test scores collected during the first wave of data collection along with the main specifications used in this paper. We report these results in Table 1.14 in Appendix 1.7.1. We view the differences in short-term outcomes as treatment effects rather than "chance bias" in randomization for the following reasons. First, time-invariant student characteristics are balanced across treatment and control (see Table 1.2). Second, the effects on English and math test scores appear to materialize in the later weeks of the fieldwork, as shown in Figure 1.6. Third, there is no significant effect on abstract reasoning, which is arguably less amenable to short-term improvements through teaching (although the difference between a significant English/math effect and an insignificant abstract reasoning effect here is not itself significant).

[14]For example, IRT models are used to estimate students' ability in the Graduate Record Examinations (GRE), the Scholastic Assessment Test (SAT), the Program for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS) assessments. The use of IRT models in the development and education literature in economics is less prevalent, but becoming common: For example, see Das and Zajonc (2010); Andrabi, Das, Khwaja, and Zajonc (2011); Andrabi, Das, and Khwaja (2017); Singh (2015, 2016); Muralidharan, Singh, and Ganimian (2016); Mbiti et al. (2019). Das and

There are two relevant characteristics of IRT models in this setting: First, they simultaneously estimate the test taker's ability and the difficulty of the questions, which allows the contribution of "correct answers" to the ability measure to vary from question to question. Second, they provide a comparable measure of student ability across different grades and survey rounds, even if the question overlap is imperfect. A common scale across grades allows us to estimate treatment effects as additional years of schooling. Following standard practice, we normalize the IRT scores with respect to the control group.

Additional data

We surveyed all the teachers in each school and conducted in-depth surveys with those teaching math and English. We asked teachers about their time use and teaching strategies. We also obtained teacher opinions on the PSL program. For a randomly selected class within each school, we conducted a classroom observation using the Stallings Classroom Observation Tool (World Bank, 2015a). Furthermore, we conducted school-level surveys to collect information about school facilities, the teacher roster, input availability (e.g., textbooks), and expenditures.

Enumerators collected information on some school practices. Specifically, enumerators recorded whether the school has an enrollment log and what information it stores; whether the school has an official time table and whether it is posted; whether the school has a parent-teacher association (PTA) and if the principal knows the PTA head's contact information (or where to find it); and whether the school has a written budget and keeps a record (and receipts) of past expenditures.[15] Additionally, we asked principals to complete two commonly used human resource instruments to measure their "intuitive score" (Agor, 1989) and "time management profile" (Schermerhorn, Osborn, Uhl-Bien, & Hunt, 2011).

For the second wave of data collection, we surveyed a random subset of households from

Zajonc (2010) provide a nice introduction to IRT models, while van der Linden (2017) provides a full treatment of IRT models.

[15]While management practices are difficult to measure, previous work has constructed detailed instruments to measure them in schools (e.g., see N. Bloom et al. (2015); Crawfurd (2017); Lemos and Scur (2016)). Due to budget constraints, we only checked easily observable differences in school management.

our student sample, recording household characteristics and attitudes of household members. We also gathered data on school enrollment and learning levels for all children 4-8 years old living in these households.

Balance and attrition

As mentioned above, the first wave of data was collected 2 to 8 weeks after the beginning of treatment; hence, we focus on time-invariant characteristics when checking balance across treatment and control. Observable (time-invariant) characteristics of students and schools are balanced across treatment and control (see Table 1.2). Eighty percent of schools in our sample are in rural areas, over an hour away from the nearest bank (which is usually located in the nearest urban center), and over 10% need to hold some classes outside due to insufficient classrooms. Boys make up 55% of our students and the students' average age is 12. According to pre-treatment administrative data (Ministry of Education - Republic of Liberia, 2015-2016), the number of students, infrastructure, and resources available to students were not statistically different across treatment and control schools (for details, see Table 1.13 in Appendix 1.7.1).

We took great care to avoid attrition: enumerators conducting student assessments participated in extra training on tracking and its importance, and dedicated generous time to tracking. Students were tracked to their homes and tested there when not available at school. Attrition in the second wave of data collection from our original sample is balanced between treatment and control and is below 4% (see Panel C). Appendix 1.7.2 has more details on the tracking and attrition that took place during data collection.

Table 1.2: Balance: Observable, time-invariant school and student characteristics

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Panel A: School characteristics (N = 185)				
Facilities (PCA)	-0.080	-0.003	-0.077	-0.070
	(1.504)	(1.621)	(0.230)	(0.232)
% holds some classes outside	13.978	14.130	-0.152	-0.000
	(34.864)	(35.024)	(5.138)	(5.094)
% rural	79.570	80.435	-0.865	-0.361
	(40.538)	(39.888)	(5.913)	(4.705)
Travel time to nearest bank (mins)	75.129	68.043	7.086	7.079
	(69.099)	(60.509)	(9.547)	(8.774)
Panel B: Student characteristics (N = 3,508)				
Age in years	12.394	12.291	0.104	0.059
	(2.848)	(2.935)	(0.169)	(0.112)
% male	54.949	56.146	-1.197	-1.459
	(49.769)	(49.635)	(2.041)	(1.247)
Wealth index	-0.006	0.024	-0.030	0.011
	(1.529)	(1.536)	(0.140)	(0.060)
% in top wealth quartile	0.199	0.219	-0.020	-0.018
	(0.399)	(0.413)	(0.026)	(0.014)
% in bottom wealth quartile	0.267	0.284	-0.017	-0.011
	(0.442)	(0.451)	(0.039)	(0.019)
ECE before grade 1	0.832	0.818	0.014	0.013
	(0.374)	(0.386)	(0.025)	(0.017)
Panel C: Attrition in the second wave of data collection (N = 3,511)				
% interviewed	95.60	95.74	-0.14	-0.35
	(20.52)	(20.20)	(0.64)	(0.44)

The first wave of data was collected 2 to 8 weeks after the beginning of treatment; hence, the focus here is on time-invariant characteristics (some of these characteristics may vary in response to the program in the long run, but are time-invariant given the duration of our study). This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2), as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Panel A has two measures of school infrastructure: The first is a school infrastructure index made up of the first component in a principal component analysis of indicator variables for classrooms, staff room, student and adult latrines, library, playground, and an improved water source. The second is whether the school ever needs to hold classes outside due to lack of classrooms. There are two measures of school rurality: First, a binary variable and second, the time it takes to travel by motorcycle to the nearest bank. Panel B has student characteristics. The wealth index is the first component of a principal component analysis of indicator variables for whether the student's household has a television, radio, electricity, a refrigerator, a mattress, a motorcycle, a fan, and a phone. Panel C shows the attrition rate (proportion of students interviewed at the first round of data collection who we were unable to interview in the second wave). Standard errors are clustered at the school level.

1.3 Experimental results

In this section, we first explore how the PSL program affected access to and quality of education. We then turn to mechanisms, looking at changes in material inputs, staffing, and school management. Replication data is available at Romero, Sandefur, and Sandholtz (2018).

1.3.1 Test scores

Following our pre-analysis plan (Romero et al., 2017), we report treatment-effect estimates from two specifications:

$$Y_{isg} = \alpha_g + \beta_{1.1} treat_s + \varepsilon_{isg} \tag{1.1}$$

$$Y_{isg} = \alpha_g + \beta_{1.2} treat_s + \gamma_{1.2} X_i + \delta_{1.2} Z_s + \varepsilon_{isg} \tag{1.2}$$

The first specification amounts to a simple comparison of post-treatment outcomes for treatment and control individuals, in which Y_{isg} is the outcome of interest for student i in school s and group g (denoting the matched pairs used for randomization); α_g is a matched-pair fixed effect (i.e., stratification-level dummies); $treat_s$ is an indicator for whether school s was randomly chosen for treatment; and ε_{isg} is an error term. The second specification adds controls for time-invariant characteristics measured at the individual level (X_i) and school level (Z_s).[16] We estimate both specifications via ordinary least squares, clustering the standard errors at the school level.

Table 1.3 shows results from student tests. The first three columns show differences between control and treatment schools' test scores after 1-2 months of treatment (September/October 2016), while the last three columns show the difference after 9-10 months of treatment (May/June 2017).

[16] These controls were specified in the pre-analysis plan and are listed in Table 1.15 (Romero et al., 2017). We had committed in the pre-analysis plan to a specification that controlled for pre-treatment individual outcomes:

$$Y_{isg} = \alpha_g + \beta_{1.3} treat_s + \gamma_{1.3} X_i + \delta_{1.3} Z_s + \zeta_{1.3} Y_{isg,-1} + \varepsilon_{isg} \tag{1.3}$$

However, as mentioned before, the first wave of data was collected after the beginning of treatment, so we lack a true baseline of student test scores. We report this specification in Table 1.14 in Appendix 1.7.1. The results are still statistically significant, but mechanically downward biased.

Columns 1, 2, 4, and 5 show intention-to-treat (ITT) treatment estimates, while Columns 3 and 6 show treatment-on-the-treated (ToT) estimates (i.e., the treatment effect for students that actually attended a PSL school in 2016/2017). The ToT is estimated using the assigned treatment as an instrument for whether the student is in fact enrolled in a PSL school during the 2016/2017 academic year.[17]

After 1-2 months of treatment, student test scores increase by 0.05σ in math (p-value=0.09) and 0.07σ in English (p-value=0.04). Part of these short-term improvements can be explained by the fact that most providers started the school year on time, while most traditional public schools began classes 1-4 weeks later. Hence, most students were already attending classes on a regular basis in treatment schools during our field visit, while their counterparts in control schools were not. We estimate the treatment effect separately for students tested during the first and the second half of the first round of data collection (see Figure 1.6 in Appendix 1.7.1), and show that the treatment effects fade in during the course of field work — further supporting our conclusion that these results represent early treatment effects as opposed to baseline imbalance.

In our preferred specification (Column 5), the treatment effect of PSL after one academic year is $.18\sigma$ for English (p-value < 0.001) and $.18\sigma$ for math (p-value < 0.001). We focus on the ITT effect, but the ToT effect is $.21\sigma$ for English (p-value < 0.001) and $.22\sigma$ for math (p-value < 0.001). Our results are robust to different measures of student ability (see Table 1.16 in Appendix 1.7.1 for details).

[17]The percentage of students originally assigned to treatment schools who are actually in treatment schools at the end of the 2016/2017 schools year is 81%. The percentage of students assigned to control schools who are in treatment schools at the end of the 2016/2017 schools year is 0%.

Table 1.3: ITT treatment effects on learning

	First wave (1-2 months after treatment)			Second wave (9-10 months after treatment)		
	ITT		ToT	ITT		ToT
	(1)	(2)	(3)	(4)	(5)	(6)
English	0.09	0.07	0.08	0.17	0.18	0.21
	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
Math	0.07	0.05	0.06	0.19	0.18	0.22
	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
Abstract	0.05	0.03	0.04	0.05	0.05	0.06
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Composite	0.08	0.06	0.07	0.19	0.18	0.22
	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
New modules				0.20	0.19	0.23
				(0.04)	(0.04)	(0.04)
Conceptual				0.13	0.12	0.15
				(0.04)	(0.04)	(0.05)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	3,508	3,508	3,508	3,492	3,492	3,492

Columns 1-3 are based on the first wave of data and show the difference between treatment and control schools taking into account the randomization design — i.e., including "pair" fixed effects — (Column 1), the difference taking into account other student and school controls (Column 2), and the treatment-on-the-treated (ToT) estimates (Column 3). Columns 4-6 are based on the second wave of data and show the difference between treatment and control taking into account the randomization design — i.e., including "pair" fixed effects — (Column 4), the difference taking into account other student and school controls (Column 5), and the treatment-on-the-treated (ToT) estimates (Column 6). The treatment-on-the-treated effects are estimated using the assigned treatment as an instrument for whether the student is in fact enrolled in a PSL school at the time of data collection. Standard errors are clustered at the school level.

An important concern when interpreting these results is whether they represent real gains in learning or better test-taking skills resulting from "teaching to the test". We show suggestive evidence that these results represent real gains. First, the treatment effect is significant (.19σ, p-value < 0.001) for new modules that were not in the first wave test (and unknown to the providers or the teachers), and statistically indistinguishable from the treatment effect over all the items (.18σ, p-value < 0.001). Second, the treatment effect is positive and significant (.12σ, p-value .0014) for the conceptual questions (which do not resemble the format of standard textbook exercises). We cannot rule out that providers narrowed the curriculum by focusing on English and mathematics or, conversely, that they generated additional learning gains in other subjects that we did not test.[18] We find no evidence of heterogeneous treatment effects by students' socio-economic status, gender, or grade (see Table 1.18 in Appendix 1.7.1).

1.3.2 Enrollment, attendance, and student selection

The previous section showed that education quality, measured using test scores in an ITT framework, increases in PSL schools. We now ask whether the PSL program increases access to education. To explore this question we focus on three outcomes which were committed to in the pre-analysis plan: Enrollment, student attendance, and student selection. PSL increased enrollment overall, but in schools where enrollment was already high and classes were large, the program led to a significant decline in enrollment (Romero et al., 2017). This does not appear to be driven by selection of "better" students, but by providers capping class sizes and eliminating double shifts.[19] As shown in Section 1.7.5, almost the entirety of this phenomenon is explained by Bridge International Academies.

Enrollment changes across treatment and control schools are shown in Panel A of Table 1.4. There are a few noteworthy items. First, treatment schools are slightly larger before treatment: They have 34 (p-value .095) more students on average before treatment. Table 1.13 uses administrative data, while Table 1.4 uses data independently collected by our survey teams. While the difference in enrollment in the 2015/2016 academic year is only significant in the latter, the point estimates are similar across both tables. Second, PSL schools on average have 57 (p-value < 0.001) more students than control schools in the 2016/2017 academic year, which results in a net increase (after controlling for pre-treatment differences) of 25 (p-value .09) students per school.

[18] As shown in Table 1.7 PSL schools have longer school days. As a result, treatment schools spend about 45 minutes per week more in both English and math. However, they do not spend a larger fraction of the school day in English or math (see Table 1.17). More broadly, we cannot rule out that PSL spent disproportionately more resources improving English and Math instruction.

[19] Three Bridge International Academies treatment schools (representing 28% of total enrollment in Bridge treatment schools) had double shifts in 2015/2016, but not in 2016/2017. One Omega Schools treatment school (representing 7.2% of total enrollment in Omega treatment schools) had double shifts in 2015/2016, but not in 2016/2017. The MOU between Bridge and the Ministry of Education authorized eliminating double shifts (Ministry of Education - Republic of Liberia, 2016b).

Since provider compensation is based on the number of students enrolled rather than the number of students actively attending school, increases in enrollment may not translate into increases in student attendance. An independent measure of student attendance conducted by our enumerators during a spot check shows that students in treatment schools are 16 (p-value < 0.001) percentage points more likely to be in school during class time (see Panel A, Table 1.4).

Turning to the question of student selection, we find no evidence that any group of students is systematically excluded from PSL schools. The proportion of students with disabilities is not statistically different in PSL schools and control schools (Panel A, Table 1.4).[20] Among our sample of students (i.e., students sampled from the 2015/2016 enrollment log), students are equally likely across treatment and control to be enrolled in the same school in the 2016/2017 academic year as they were in 2015/2016, and equally likely to be enrolled in any school (see Panel B, Table 1.4). Finally, selection analysis using student-level data on wealth, gender, and age finds no evidence of systematic exclusions (see Table 1.19 in Appendix 1.7.1).

[20] However, the fraction of students identified as disabled in our sample is an order of magnitude lower than estimates for the percentage of disabled students in the U.S and worldwide using roughly the same criteria (both about 5%) (Brault, 2011; UNICEF, 2013).

Table 1.4: ITT treatment effects on enrollment, attendance, and selection

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Panel A: School level data (N = 175)				
Enrollment 2015/2016	298.45	264.11	34.34	34.18
	(169.74)	(109.91)	(21.00)	(20.28)
Enrollment 2016/2017	309.71	252.75	56.96	56.89
	(118.96)	(123.41)	(18.07)	(16.29)
15/16 to 16/17 enrollment change	11.55	-6.06	17.61	24.60
	(141.30)	(82.25)	(17.19)	(14.35)
Attendance % (spot check)	48.02	32.83	15.19	15.57
	(24.52)	(26.55)	(3.81)	(3.13)
% of students with disabilities	0.59	0.39	0.20	0.21
	(1.16)	(0.67)	(0.14)	(0.15)
Panel B: Student level data (N = 3,639)				
% enrolled in the same school	80.50	83.16	-2.66	0.71
	(39.63)	(37.43)	(3.66)	(2.06)
% enrolled in school	94.13	93.99	0.14	1.23
	(23.52)	(23.77)	(1.33)	(0.87)
Days missed, previous week	0.85	0.85	-0.00	-0.06
	(1.41)	(1.40)	(0.10)	(0.07)

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Panel A presents school level data including enrollment (taken from enrollment logs) and student attendance measure by our enumerators during a spot check in the middle of a school day. If the school was not in session during a regular school day we mark all students as absent. Panel B presents student level data including whether the student is still enrolled in the same schools, whether he is enrolled in school at all, and whether it missed school in the previous week (conditional on being enrolled in school). Standard errors are clustered at the school level.

Providers are authorized to cap class sizes, which could lead to students being excluded from their previous school (and either transferred to another school or to no school at all). We estimate whether the caps are binding for each student by comparing the average enrollment before treatment in her grade cohort and the two adjacent grade cohorts (i.e., one grade above and below) to the theoretical class-size cap under PSL. We average over three cohorts because some providers used placement tests to reassign students across grade levels. Thus the "constrained" indicator is defined by the number of students enrolled in the student's 2016/2017 "expected grade" (as predicted based on normal progression from their 2015/2016 grade) and adjacent grades, divided by the "maximum capacity" in those three grades in 2016/2017 (as specified in our pre-analysis plan (Romero et al., 2017)):

$$c_{igso} = \frac{Enrollment_{is,g-1} + Enrollment_{is,g} + Enrollment_{is,g+1}}{3 * Maximum_o},$$

where c_{igso} is our "constrained" measure for student i, expected to be in grade g in 2016/2017, at school s, in a "pair" assigned to provider o. $Enrollment_{is,g-1}$ is enrollment in the grade below the student's expected grade, $Enrollment_{is,g}$ is enrollment in the student's expected grade, and $Enrollment_{is,g+1}$ is enrollment in the grade above the student's expected grade. $Maximum_o$ is the class cap approved for provider o. We label a student's grade-school combination as "constrained" if $c_{igso} > 1$.

Enrollment in constrained school-grades decreases, while enrollment in unconstrained school-grades increases (see Column 1 in Table 1.5). Thus, schools far below the cap have positive treatment effects on enrollment and schools near or above the cap offset it with declining enrollment. Our student data reveal this pattern as well: Columns 2 and 3 in Table 1.5 show the ITT effect on enrollment depending on whether students were enrolled in a constrained class in 2015/2016. In unconstrained classes students are more likely to be enrolled in the same school (and in any school). But in constrained classes students are less likely to be enrolled in the same school. While there is no effect on overall school enrollment, switching schools may be disruptive for children (Hanushek, Kain, & Rivkin, 2004). Finally, test-scores improve for students in constrained classes. This result is difficult to interpret as it includes the positive treatment effect over students who did not change schools (compounded by smaller class sizes) with the effect over students removed from their schools. These results are robust to excluding adjacent grades from the "constrained" measure (see Table 1.20 in Appendix 1.7.1).

Table 1.5: ITT treatment effects, by whether class size caps are binding

	(1) Δ enrollment	(2) % same school	(3) % in school	(4) Test scores
Constrained=0 × Treatment	5.30	3.90	1.65	0.15
	(1.11)	(1.40)	(0.73)	(0.034)
Constrained=1 × Treatment	-11.7	-12.5	0.085	0.35
	(6.47)	(7.72)	(4.12)	(0.11)
No. of obs.	1,635	3,637	3,485	3,490
Mean control (Unconstrained)	-0.75	81.89	93.38	0.13
Mean control (Constrained)	-7.73	83.85	94.81	-0.08
α_0 = Constrained - Unconstrained	-17.05	-16.36	-1.56	0.20
p-value ($H_0 : \alpha_0 = 0$)	0.01	0.04	0.71	0.07

Column 1 uses school-grade level data and the outcome is the change in enrollment (between 2015/2016 and 2016/2017) at the grade level. Columns 2 - 4 use student level data. The outcomes are whether the student is in the same school or not (Column 2), whether the student is still enrolled in any school (Column 3), and the composite test score (Column 4). Standard errors are clustered at the school level. There were 194 constrained classes before treatment (holding 30% of students), and 1,468 unconstrained classes before treatment (holding 70% of students). Standard errors are clustered at the school level.

1.3.3 Intermediate inputs

In this section we explore the effect of the PSL program on school inputs (including teachers), school management (with a special focus on teacher behavior and pedagogy), and parental behavior.

Inputs and resources

Teachers, one of the most important inputs of education, change in several ways in treatment schools (see Panels A/B in Table 1.6). PSL schools have 2.6 more teachers on average (p-value < 0.001), but this is not merely the result of operators hiring more teachers. Rather, the Ministry of Education agreed to release some underperforming teachers from PSL schools, replace those teachers, and provide additional ones. Ultimately, the extra teachers result in lower pupil-teacher ratios (despite increased student enrollment). This re-shuffling of teachers means that PSL schools have younger and less-experienced teachers, who are more likely to have worked in private schools in the past and have higher test scores (we conducted simple memory, math, word association, and abstract thinking tests). Replacement and extra teachers are recent graduates from the Rural Teacher Training Institutes (see King, Korda, Nordstrum, and Edwards (2015) for details on this program). While the program's contracts made no provisions to pay teachers differently in treatment and control schools, teachers in PSL schools report higher wages. A potential explanation, is that there are many teachers that are paid by the community in public schools (commonly known as 'volunteer' teachers). If higher salaries for teachers in PSL schools are conditional on them working in program schools, then this would create an incentive to perform well. However, we could not find an explanation for these higher salaries. Hence it is unclear whether higher salaries are tied to the program. But large unconditional increases in teacher salaries have been shown elsewhere to have no effect on student performance in the short run (de Ree, Muralidharan, Pradhan, & Rogers, 2018).

Our enumerators conducted a "materials" check during classroom observations (See Panels C - Table 1.6). Since we could not conduct classroom observations in schools that were out of session during our visit, Table 1.21 in Appendix 1.7.1 presents Lee (2009) bounds on these treatment effects (control schools are more likely to be out of session). Conditional on the school being in session during our visit, students in PSL schools are 23 percentage points (p-value < 0.001) more likely to have a textbook and 8.2 percentage points (p-value .051) more likely to have writing materials (both a pen and a copybook). However, we cannot rule out that there is no overall effect as zero is between the Lee (2009) bounds.

Table 1.6: ITT treatment effects on inputs and resources

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Panel A: School-level outcomes (N = 185)				
Number of teachers	9.62	7.02	2.60	2.61
	(2.82)	(3.12)	(0.44)	(0.37)
Pupil-teacher ratio (PTR)	32.20	39.95	-7.74	-7.82
	(12.29)	(18.27)	(2.31)	(2.12)
New teachers	4.81	1.77	3.03	3.01
	(2.56)	(2.03)	(0.34)	(0.35)
Teachers dismissed	3.27	2.12	1.15	1.13
	(3.81)	(2.62)	(0.48)	(0.47)
Panel B: Teacher-level outcomes (N = 1,167)				
Age in years	39.09	46.37	-7.28	-7.10
	(11.77)	(11.67)	(1.02)	(0.68)
Experience in years	10.59	15.79	-5.20	-5.26
	(9.20)	(10.77)	(0.76)	(0.51)
% has worked at a private school	47.12	37.50	9.62	10.20
	(49.95)	(48.46)	(3.76)	(2.42)
Test score in standard deviations	0.13	-0.01	0.14	0.14
	(1.02)	(0.99)	(0.07)	(0.06)
% certified (or tertiary education)	60.11	58.05	2.06	4.20
	(48.99)	(49.39)	(4.87)	(2.99)
Salary (USD/month)–Conditional on salary>0	121.36	104.54	16.82	13.90
	(44.42)	(60.15)	(6.56)	(4.53)
Panel C: Classroom observation (N = 143)				
Number of seats	20.64	20.58	0.06	0.58
	(13.33)	(13.57)	(2.21)	(1.90)
% with students sitting on the floor	2.41	4.23	-1.82	-1.51
	(15.43)	(20.26)	(2.94)	(2.61)
% with chalk	96.39	78.87	17.51	16.58
	(18.78)	(41.11)	(5.29)	(5.50)
% of students with textbooks	37.08	17.60	19.48	22.60
	(43.22)	(35.25)	(6.33)	(6.32)
% of students with pens/pencils	88.55	79.67	8.88	8.16
	(19.84)	(30.13)	(4.19)	(4.10)

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Panel A has school level outcomes. Panel B presents teacher-level outcomes including their score in tests conducted by our survey teams. Panel C presents data on inputs measured during classroom observations. Since we could not conduct classroom observations in schools that were out of session during our visit, Table 1.21 in Appendix 1.7.1 presents Lee (2009) bounds on these treatment effects (control schools are more likely to be out of session). Standard errors are clustered at the school level.

School management

Two important management changes are shown in Table 1.7: PSL schools are 8.7 percentage points more likely to be in session (i.e., the school is open, students and teachers are on campus, and classes are taking place) during a regular school day (p-value .058), and have a longer school day that translates into 3.2 more hours per week of instructional time (p-value .0011). Although principals in PSL schools have scores in the "intuitive" and "time management profile" scale that are almost identical to their counterparts in traditional public schools, they spend more of their time on management-related activities (e.g., supporting other teachers, monitoring student progress, meeting with parents) than actually teaching, suggesting a change in the role of the principal in these schools — perhaps as a result of additional teachers, principals in PSL schools did not have to double as teachers. Additionally, management practices (as measured by a "good practices" PCA index normalized to a mean of zero and standard deviation of one in the control group) are $.4\sigma$ (p-value .0011) higher in PSL schools.[21] This effect size can be viewed as a boost for the average treated school from the 50th to the 66th percentile in management practices.

[21]The index includes whether the school has an enrollment log and what information is in it, whether the school has an official time table and whether it is posted, whether the school has a parent-teacher association (PTA) and whether the principal has the PTA head's number at hand, and whether the school keeps a record of expenditures and a written budget. Table 1.22 has details on every component of the good practices index.

Table 1.7: ITT treatment effects on school management

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
% school in session at spot check	92.47	83.70	8.78	8.66
	(26.53)	(37.14)	(4.75)	(4.52)
Instruction time (hrs/week)	17.84	14.69	3.15	3.17
	(4.84)	(4.04)	(0.66)	(0.65)
Intuitive score (out of 12)	4.08	4.03	0.04	0.02
	(1.35)	(1.38)	(0.20)	(0.19)
Time management score (out of 12)	5.60	5.69	-0.09	-0.10
	(1.21)	(1.35)	(0.19)	(0.19)
Principal's working time (hrs/week)	21.43	20.60	0.83	0.84
	(11.83)	(14.45)	(1.94)	(1.88)
% of principle's time spent on management	74.06	53.64	20.42	20.09
	(27.18)	(27.74)	(4.12)	(3.75)
Index of good practices (PCA)	0.41	-0.00	0.41	0.40
	(0.64)	(1.00)	(0.12)	(0.12)
Observations	93	92	185	185

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Intuitive score is measured using Agor (1989)'s instrument and time management profile using Schermerhorn et al. (2011)'s instrument. The index of good practices is the first component of a principal component analysis of the variables in Table 1.22. The index is normalized to have mean zero and standard deviation of one in the control group. Standard errors are clustered at the school level.

Teacher behavior

An important component of school management is teacher accountability and its effects on teacher behavior. As mentioned above, teachers in PSL schools are drawn from the pool of unionized civil servants with lifetime appointments and are paid by the Liberian government. In theory, private providers have limited authority to request teacher reassignments and no authority to promote or dismiss civil service teachers. Thus, a central hypothesis underlying the PSL program is that providers can hold teachers accountable through monitoring and support, rather than rewards and threats.[22]

[22] As mentioned above, in practice the Ministry of Education agreed to release some underperforming teachers from PSL schools at the request of providers. While providers could have provided teachers with performance incentives, we have no evidence that any of them did.

To study teacher behavior, we conducted unannounced spot checks of teacher attendance and collected student reports of teacher behavior (see Panels A/B in Table 1.8). Also, during these spot checks we used the Stallings classroom observation instrument to study teacher time use and classroom management (see Panel C in Table 1.8).

Teachers in PSL schools are 20 percentage points (p-value < 0.001) more likely to be in school during a spot check (from a base of 40%) and the unconditional probability of a teacher being in a classroom increases by 15 percentage points (p-value < 0.001). Our spot checks align with student reports on teacher behavior. According to students, teachers in PSL schools are 7.5 percentage points (p-value < 0.001) less likely to have missed school the previous week. Students in PSL schools also report that teachers are 6.6 percentage points (p-value .011) less likely to hit them.

Classroom observations also show changes in teacher behavior and pedagogical practices. Teachers in PSL schools are 15 percentage points (p-value .0027) more likely to engage in either active instruction (e.g., teacher engaging students through lecture or discussion) or passive instruction (e.g., students working in their seat while the teacher monitors progress) and 25 percentage points (p-value < 0.001) less likely to be off-task.[23] Although these are considerable improvements, the treatment group is still far off the Stallings et al. (2014) good practice benchmark of 85 percent of total class time used for instruction, and below the average time spent on instruction across five countries in Latin America (Bruns & Luque, 2014).

[23]See Stallings, Knight, and Markham (2014) for more details on how active and passive instruction, as well as time off-task and student engagement, are coded.

Table 1.8: ITT treatment effects on teacher behavior

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Panel A: Spot checks (N = 185)				
% on schools campus	60.32	40.38	19.94	19.79
	(23.10)	(25.20)	(3.56)	(3.48)
% in classroom	47.02	31.42	15.60	15.37
	(26.65)	(25.04)	(3.80)	(3.62)
Panel B: Student reports (N = 185)				
Teacher missed school previous week (%)	17.72	25.12	-7.41	-7.53
	(10.79)	(14.93)	(1.92)	(1.95)
Teacher never hits students (%)	54.73	48.20	6.52	6.59
	(18.76)	(17.07)	(2.64)	(2.53)
Teacher helps outside the classroom (%)	50.02	46.59	3.42	3.56
	(18.24)	(18.01)	(2.67)	(2.28)
Panel C: Classroom observations (N = 185)				
Instruction (active + passive) (% of class time)	49.68	35.00	14.68	14.51
	(32.22)	(37.08)	(5.11)	(4.70)
Classroom management (% class time)	19.03	8.70	10.34	10.25
	(20.96)	(14.00)	(2.62)	(2.73)
Teacher off-task (% class time)	31.29	56.30	-25.01	-24.77
	(37.71)	(42.55)	(5.91)	(5.48)
Student off-task (% class time)	50.41	47.14	3.27	2.94
	(33.51)	(38.43)	(5.30)	(4.59)

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Panel A presents data from spot checks conducted by our survey teams in the middle of a school day. Panel B presents data from our panel of students where we asked them about their teachers' behavior. Panel C presents data from classroom observations. If the school was not in session during a regular school day we mark all teachers not on campus as absent and teachers and students as off-task in the classroom observation. Table 1.21 has the results without imputing values for schools not in session. Standard errors are clustered at the school level.

These estimates combine the effects on individual teacher behavior with changes to teacher composition. To estimate the treatment effect on teacher attendance over a fixed pool of teachers, we perform additional analyses in Appendix 1.7.1 using administrative data (EMIS) to restrict our sample to teachers who worked at the school the year before the intervention began (2015/2016). We treat teachers who no longer worked at the school in the 2016/2017 school year as (non-random) attriters and estimate Lee (2009) bounds on the treatment effect. Table 1.21 in Appendix 1.7.1 shows an ITT treatment effect of 14 percentage points (p-value < 0.001) on teacher attendance. Importantly, zero is not part of the Lee (2009) bounds for this effect. This aligns with previous findings showing that management practices have significant effects on worker performance (N. Bloom, Liang, Roberts, & Ying, 2014; N. Bloom, Eifert, Mahajan, McKenzie, & Roberts, 2013; Bennedsen, Nielsen, Pérez-González, & Wolfenzon, 2007).

1.3.4 Other outcomes

Student data (Table 1.9, Panel C) and household data (Table 1.9, Panel A) show that the program also increases student and parental satisfaction. Students in PSL are more likely to think going to school is fun, and parents with children in PSL schools (enrolled in 2015/2016) are 7.5 percentage points (p-value .022) more likely to be satisfied with the education their children are receiving.

Providers are not allowed to charge fees and PSL should be free at all levels, including early-childhood education (ECE) for which fees are permitted in government schools. We interviewed both parents and principals regarding fees. In both treatment and control schools parents are more likely to report paying fees than schools are to report charging them. The amount parents claim to pay in school fees is much higher than the amount schools claim to charge (see Panel A and Panel B in Table 1.9). Since principals may be reluctant to disclose the full amount they charge parents, especially in primary school (which is nominally free), this discrepancy is normal. While the likelihood of charging fees decreases in PSL schools by 26 percentage points according to parents and by 19 percentage points according to principals, 48% of parents still report paying some fees in PSL schools.

Providers often provide textbooks and uniforms free of charge to students (see Section 1.2.1). Indeed, household expenditures on fees, textbooks, and uniforms drop (see Table 1.23 for details).

In total, annual household expenditures on children's education decrease by 6.6 USD (p-value .11). A reduction in household expenditure in education reflects a crowding out response (i.e., parents decrease private investment in education as school investments increase). To explore whether crowding out goes beyond expenditure we ask parents about engagement in their child's education. However, we see no change in this margin (we summarize parental engagement using the first component from a principal component analysis across several measures of parental engagement; see Table 1.24 for the effect on each component).

To complement the effect of the program on cognitive skills, we also look for changes in student attitudes and opinions (see Table 1.9, Panel C). Some of the control group rates are noteworthy: 50% of children use what they learn in class outside school, 69% think that boys are smarter than girls, and 79% think that some tribes in Liberia are bad. Turning to treatment effects, children in PSL schools are more likely to think school is useful, more likely to think elections are the best way to choose a president, and less likely to think some tribes in Liberia are bad. The effect on tribe perceptions is particularly important in light of the recent conflict in Liberia and the ethnic tensions that sparked it. Our results also align with previous findings from Andrabi, Bau, Das, and Khwaja (2010), who show that children in private schools in Pakistan are more "pro-democratic" and exhibit lower gender biases (although we do not find any evidence of lower gender biases in this setting). Note, however, that our treatment effects are small in magnitude. It is also impossible to tease out the effect of who is providing education (private providers vs regular public schools) from the effect of better education, and the effect of younger and better teachers. Hence, our results show the net change in students' opinions, and cannot be attributed to providers per se but rather to the program as a whole.

Table 1.9: ITT treatment effects on household behavior, fees, and student attitudes

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Panel A: Household behavior (N = 1,115)				
% satisfied with school	74.90	67.41	7.49	7.51
	(19.18)	(23.99)	(3.20)	(3.23)
% paying any fees	48.08	73.59	-25.50	-25.74
	(50.00)	(44.13)	(4.73)	(3.27)
Fees (USD/year)	5.68	8.06	-2.38	-2.95
	(10.16)	(9.73)	(0.97)	(0.62)
Expenditure (USD/year)	65.57	73.53	-7.95	-6.60
	(74.84)	(79.32)	(6.95)	(4.11)
Engagement index (PCA)	-0.11	-0.09	-0.03	-0.03
	(0.84)	(0.91)	(0.07)	(0.06)
Panel B: Fees (N = 184)				
% with $>$ 0 ECE fees	11.83	30.77	-18.94	-18.98
	(32.47)	(46.41)	(5.92)	(5.42)
% with $>$ 0 primary fees	12.90	29.67	-16.77	-16.79
	(33.71)	(45.93)	(5.95)	(5.71)
ECE Fee (USD/year)	0.57	1.42	-0.85	-0.87
	(1.92)	(2.78)	(0.35)	(0.33)
Primary Fee (USD/year)	0.54	1.22	-0.68	-0.70
	(1.71)	(2.40)	(0.31)	(0.31)
Panel C: Student attitudes (N = 3,492)				
School is fun	0.58	0.53	0.05	0.05
	(0.49)	(0.50)	(0.02)	(0.02)
I use what I'm learning outside of school	0.52	0.49	0.04	0.04
	(0.50)	(0.50)	(0.02)	(0.02)
If I work hard, I will succeed.	0.60	0.55	0.05	0.04
	(0.49)	(0.50)	(0.03)	(0.02)
Elections are the best way to choose a president	0.90	0.88	0.03	0.03
	(0.30)	(0.33)	(0.01)	(0.01)
Boys are smarter than girls	0.69	0.69	0.00	0.01
	(0.46)	(0.46)	(0.02)	(0.01)
Some tribes in Liberia are bad	0.76	0.79	-0.03	-0.03
	(0.43)	(0.41)	(0.02)	(0.01)

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Panel A presents data from household surveys. The index for parent engagement is the first component from a principal component analysis across several measures of parental engagement; see Table 1.24 for details. Expenditure refers to the annual household expenditures on children's education. Panel B presents data from school principals on what fees schools charge. Panel C presents data on whether students agree or disagree with several statements. Standard errors are clustered at the school level.

1.4 Provider comparisons

1.4.1 Raw differences

As discussed in Section 1.2.2 and shown in Table 1.11, PSL schools are not a representative sample of public schools. Furthermore, there is heterogeneity in school characteristics across providers. This is unsurprising since providers stated different preferences for locations and some volunteered to manage schools in more remote and marginalized areas. Therefore, the raw treatment effects for each individual provider are internally valid, but not comparable with each other without further assumptions (see Section 1.4.2).

We show how the average school for each provider differs from the average public school in Liberia in Table 1.25. We reject the null that providers' schools have similar characteristics on at least three margins: number of students, pupil/teacher ratio, and the number of permanent classrooms. Bridge International Academies is managing schools that were considerably bigger (in 2015/2016) than the average public school in Liberia (by over 150 students), and these schools are larger than those of other providers by over 100 students. Most providers have schools with better infrastructure than the average public school in the country, except for Omega and Stella Maris. Finally, while all providers have schools that are closer to a paved road than other public schools, Bridge's and BRAC's schools are about 2 km closer than other providers' schools. Overall, these results confirm that some providers were more willing to work in average Liberian schools, while others preferred schools with easier access and better infrastructure.

We now turn to provider-by-provider outcomes. We focus on three margins: 1) *Learning*, as measured by test scores; 2) *Sustainability*, providers' willingness to improve the behavior and pedagogy of existing teachers (as opposed to having the worst-performing teachers transferred to other public schools, imposing a negative externality on the broader school system); and 3) *Equity*, or providers' commitment to improving access to quality education (rather than learning gains for a subset of pupils).

Table 1.10: Raw (fully experimental) treatment effects by provider

	(1) BRAC	(2) Bridge	(3) MtM	(4) Omega	(5) Rising	(6) St. Child	(7) Stella M	(8) YMCA
Panel A: Student test scores (ITT)								
English (standard deviations)	0.19	0.28	0.19	-0.07	0.36	0.23	-0.23	0.58
	(0.10)	(0.09)	(0.22)	(0.11)	(0.24)	(0.13)	(0.23)	(0.26)
Math (standard deviations)	0.09	0.39	0.18	-0.06	0.42	0.28	-0.17	0.27
	(0.09)	(0.09)	(0.22)	(0.11)	(0.23)	(0.13)	(0.22)	(0.26)
Composite (standard deviations)	0.14	0.36	0.18	-0.08	0.42	0.27	-0.19	0.38
	(0.09)	(0.09)	(0.22)	(0.11)	(0.23)	(0.13)	(0.22)	(0.26)
Panel B: Changes to the pool of teachers								
% teachers dismissed	-6.75	50.47	15.51	-8.58	-5.79	-3.18	-10.99	21.08
	(6.43)	(6.29)	(11.75)	(6.81)	(12.72)	(8.49)	(14.34)	(14.34)
% new teachers	39.53	63.17	70.88	24.44	24.30	41.14	-20.32	62.37
	(12.27)	(12.00)	(22.43)	(12.99)	(24.28)	(16.19)	(27.37)	(27.37)
Age in years (teachers)	-5.03	-10.92	-11.20	-5.46	-10.75	-5.79	-4.53	3.25
	(1.93)	(2.01)	(3.52)	(2.03)	(3.82)	(2.54)	(4.30)	(4.30)
Test score in standard deviations (teachers)	0.03	0.36	0.48	0.18	0.18	0.32	0.16	-0.59
	(0.17)	(0.17)	(0.31)	(0.17)	(0.33)	(0.22)	(0.38)	(0.38)
Panel C: Enrollment and access								
Δ enrollment	38.02	-13.26	-25.98	51.27	19.31	44.86	-15.92	45.38
	(34.33)	(33.60)	(62.76)	(35.26)	(67.84)	(45.21)	(76.59)	(76.53)
Δ enrollment (constrained grades)	0.00	-23.85	0.00	0.28	0.00	32.15	-1.00	-46.35
	(0.00)	(11.19)	(0.00)	(37.16)	(0.00)	(61.95)	(5.13)	(27.05)
Student attendance (%)	20.12	5.25	37.80	18.01	28.76	19.56	9.72	13.53
	(9.02)	(9.05)	(16.50)	(9.53)	(17.83)	(11.88)	(23.32)	(20.11)
% students still attending any school	1.27	5.19	-3.12	4.71	2.82	3.64	5.98	4.48
	(4.45)	(4.22)	(10.25)	(4.99)	(11.03)	(6.11)	(10.57)	(12.21)
% students still attending same school	0.80	4.42	0.65	1.56	3.81	-0.82	1.03	-0.81
	(2.20)	(2.09)	(5.07)	(2.46)	(5.45)	(3.02)	(5.23)	(6.04)
Panel D: Satisfaction								
% satisfied with school (parents)	11.72	13.22	0.75	0.21	4.95	-4.96	29.49	18.02
	(7.30)	(7.14)	(13.34)	(7.53)	(14.44)	(9.62)	(16.28)	(16.27)
% students who think school is fun	5.83	2.11	0.50	4.86	9.44	2.84	-17.50	20.92
	(4.89)	(4.63)	(11.25)	(5.47)	(12.11)	(6.71)	(11.60)	(13.40)
Observations	40	45	8	12	38	10	24	8

This table presents the raw treatment effect for each provider on different outcomes. The estimates for each provider are *not* comparable to each other without further assumptions, and thus we do not include a test of equality. Panel A presents data on students' test scores. Panel B presents data related to the pool of teachers in each school. Panel C presents data related to school enrollment. Δ enrollment measures the change in enrollment between the 2015/2016 and 2016/2017 school year. Panel D presents data from household surveys. Standard errors are shown in parentheses. Estimation is conducted on collapsed, school-level data.

The treatment effects on composite test scores are positive and significantly different from zero for three providers: Rising Academies, Bridge International Academies, and Street Child (Table 1.10 - Panel A). They are positive but statistically insignificant for Youth Movement for Collective Action, More Than Me, and BRAC. Non-compliance likely explains the negative (but statistically insignificant) effect for Stella Maris and Omega Schools. Stella Maris never took control of its assigned schools. Omega had not taken control of all its schools by the end of the school year. Our teacher interviews reflect these providers' absence. In 3 out of four Stella Maris schools, all the teachers reported that no one from Stella had been at the school in the previous week. In 6 out of 19 Omega schools all the teachers reported that no one from Omega had been at the school in the previous week. While we committed in the pre-analysis plan to compare for-profit to non-profit providers, this comparison yields no clear patterns (Romero et al., 2017).

To measure teacher selection, we study the number of teachers dismissed and the number of new teachers recruited (Table 1.10 - Panel B). As noted above, PSL led to the assignment of 2.6 additional teachers per school and 1.1 additional teachers exiting per school. However, large-scale dismissal of teachers was unique to one provider (Bridge International Academies), while successful lobbying for additional teachers was common across several providers. Although weeding out bad teachers is important, a reshuffling of teachers is unlikely to raise average performance in the system as a whole. We are unable to verify whether the teachers dismissed from PSL schools were reassigned to other public schools.

While enrollment increased across all providers, the smallest treatment effect on this margin is for Bridge, which is consistent with that provider being the only one enforcing class size caps (see Panel C in Table 1.10 and Figure 1.8 in Appendix 1.7.1 for more details). As shown in Section 1.3.2, in classes where class-size caps were binding (10% of all classes holding 30% of students at baseline), enrollment fell by 12 students per grade.

1.4.2 Comparable treatment estimates

There are two hurdles to comparing provider-specific treatment effects. First, while the assignment of schools within matched pairs was random, the assignment of pairs to providers was not, resulting in non-random differences in schools and locations across providers. Second, the sample sizes for most providers are too small to yield reliable estimates.

To mitigate the bias due to differences in locations and schools we control for a comprehensive set of school characteristics (to account for the fact that some providers' schools will score better than others for reasons unrelated to PSL), as well as interactions of those characteristics with a treatment dummy (to account for the possibility that raising scores through PSL relative to the control group will be easier in some contexts than others). We control for both student (age, gender, wealth, and grade) and school characteristics (pre-treatment enrollment, facilities, and rurality).

Because randomization occurred at the school level and some providers are managing only four or five treatment schools, the experiment is under-powered to estimate their effects. Additionally, since the "same program" was implemented by different providers, it would be naïve to treat providers' estimators as completely independent from each other. We take a Bayesian approach to this problem, estimating a hierarchical model (Rubin, 1981) (see Gelman, Carlin, Stern, and Rubin (2014) and Meager (2016) for a recent discussion). By allowing dependency across providers' treatment effects, the model "pools power" across providers, and in the process pulls estimates for smaller providers toward the overall average (a process known as "shrinkage"). The results of the Bayesian estimation are a weighted average of providers' own performance and average performance across all providers, and the proportions depend on the provider's sample size. We apply the Bayesian estimator after adjusting for baseline school differences and estimating the treatment effect of each provider on the average school in our sample.[24]

We show the full set of results across providers after adjusting for baseline differences and "shrinking" the estimates using the Bayesian hierarchical model in Table 1.26 in Appendix 1.7.1. While the comparable effects are useful for comparisons, the raw experimental estimates remain useful for non-comparative statements (e.g., whether a provider had an effect or not). Figure 1.7 in Appendix 1.7.1 shows the effects on learning after adjusting for differences in school characteristics (before the Bayesian hierarchical model) and the effects after applying a Bayesian hierarchical model (but without adjusting for school differences). Qualitatively, the results do not change. The learning gains remain positive for the same providers, and even after "shrinking" Bridge remains the only provider with a high (and statistically significant) percentage of teacher dismissal and the only one with a negative (and statistically significant) effect on enrollment in constrained grades.

[24]This model assumes that the true treatment effect for each provider is drawn from a normal distribution (with unknown mean and variance), and that the observed effect is sampled from a normal distribution with mean equal to the true effect. The "weight" given to the provider's own performance depends on the provider's sample size and the prior distribution for the standard deviation of the distribution of true effects. We assume a non-informative prior for the standard deviation. The results are robust to the choice of prior and are available upon request.

1.4.3 Excluding some providers

What will be the long-run impact of this program? The program was explicitly framed as a pilot, where the government would learn what works and what does not in the first year and adjust accordingly. Adjustments could be made on many different dimensions, but a unique feature of this program is the existence of eight independent operators offering competing services. This provides the opportunity for the PSL program to improve performance not only through learning by operators, but also through learning by the government about operators. Taking operator-specific performance as a fixed characteristic, we calculate how overall program performance could be improved in terms of both learning gains and non-learning outcomes through selective renewal or cancellation of operator contracts.

For example, setting aside any political economy considerations, the government could drop the two providers that did not make much effort to control their schools (Omega and Stella Maris). It could also drop any provider who is potentially generating negative externalities (Bridge) or who may fail on dimensions different from test scores such as protecting students from physical and sexual abuse (More than Me and YMCA). We estimate these potential outcomes by taking an inverse-variance weighted average across providers. We do this using both the raw estimates and comparable treatment estimates for completeness; however, given that the comparable treatment estimates are meant to inform about the treatment estimates from the operators in any school in the experiment we focus on those.[25] Dropping the worst performing providers (Omega and Stella Maris) increases the overall treatment effect to $.23\sigma$, while taking off the providers that may generate negative externalities (Bridge) reduces the treatment effect to $.16\sigma$ (see Table 1.28 in the Appendix for details). Dropping both the worst performers and Bridge increases the overall treatment effect to $.2\sigma$. Also dropping More than Me and YMCA, who have have allegedly failed to safeguard children in their schools from sexual abuse (Baysah, 2016; Young, 2018), results in an overall treatment

[25] An alternative is to estimate the overall treatment effect using a Bayesian framework among the experiments from the providers that are not dropped (akin to Bayesian meta-analysis). However, as we argue above, it would be naïve to treat providers' estimators as completely independent from each other and the treatment effect from the dropped providers is informative of the overall treatment effect (even in the absence of these providers), as well as the treatment effect in the average school in the sample.

effect of .19σ. While the political economy of provider selection is non-trivial, we see this as prima facie evidence that the program has the potential to improve outcomes further by selecting providers dynamically.

1.5 Was PSL worth the cost?

To attempt an answer to this question we make two comparisons: a comparative cost-effectiveness calculation comparing PSL to a business-as-usual expansion of Liberia's public school system, and a cost-benefit calculation based on the net present value of the Mincerian earnings returns to the education provided by PSL. Both calculations require strong assumptions (Dhaliwal et al., 2013), and we discuss a range of plausible alternatives. We focus on cost-effectiveness in this section, but our cost-benefit analysis suggest PSL is worth the investment under a fairly robust set of assumptions if we do not take into account the additional cost incurred by providers (see Appendix 1.7.4 for details).

Our data on operator costs are imperfect (see Section 1.2.1), and it is extremely difficult to predict the long term unit cost of the program. Therefore we take as a lower bound USD 50 per pupil, which was the government's budget target for PSL and the transfer made to operators. Computing the benefits is more straightforward. The ToT effect is .22σ, implying test scores increased at most by 0.44σ per USD 100 spent (assuming a linear-dose relationship).

The PSL program reflects a fairly holistic (and costly) overhaul of how public schools operate. Comparing the average costs and benefits of a large-scale reform to the literature measuring treatment effects of marginal improvements to existing school systems may be uninformative. Nevertheless, some of these reforms, particularly those focused on increasing accountability (e.g., teacher performance pay and school-based management) have generated equal or greater increases in learning in other contexts, at lower cost per child.[26] Further testing would be required to know

[26]For example, Glewwe, Ilias, and Kremer (2010) in Kenya and Mbiti et al. (2019) in Tanzania show that teacher performance pay increased test scores by 6.29σ and 4.58σ per $100 spent, respectively. In Indonesia, Pradhan et al. (2014) find that linking school committees to the village council increase test scores by 2.27σ per $100 spent. For a review of the most cost-effective school-level interventions in the developing world, see Kremer, Brannen, and

whether similar results could be achieved in Liberia.

Arguably, a more informative comparison is between PSL and a business-as-usual increase in expenditure on Liberian public schools. A useful benchmark for comparison is to assume that the government would follow its current pattern of spending almost exclusively on employing teachers. Thus, any increase in government expenditure would either increase teacher salaries or reduce pupil-teacher ratios.[27] Existing experimental estimates from the developing world suggest that either strategy would have, at best, modest impacts on test scores. In Indonesia, de Ree et al. (2018) show that large unconditional increases in teacher salaries have no effect on student performance in the short run. In Kenya, Duflo, Dupas, and Kremer (2015) find a reduction of the pupil-teacher ratio by ten increases test scores 0.06σ, and in India Banerjee, Cole, Duflo, and Linden (2007) find no significant effect (and a point estimate of the opposite sign). Likewise, using data from control schools, we estimate that the relationship between pupil-teacher ratios and student test scores is $-.0014\sigma$. Spending an extra USD 50 on hiring more teachers would cut in half pupil-teacher ratios (the average student faces a class size of 36) and increase test scores by $.026\sigma$, compared to $.22\sigma$ under PSL.

These estimates suggest that additional spending through PSL may be more cost-effective than additional spending (to increase the number of teachers) under business-as-usual. Indeed, increasing school resources without changing the incentives or the accountability structure has been shown to have little impact on learning outcomes in developing countries (Glewwe, Kremer, & Moulin, 2009; Das et al., 2013; Sabarwal, Evans, & Marshak, 2014; Mbiti et al., 2019).

Glennerster (2013).

[27]We do not present a cost-effectiveness comparison of the effect of the program on access to schooling since the overall treatment effect on enrollment is not statistically different from zero. However, an alternative policy of increasing the number of teachers may attract new students, particularly if those new teachers were placed in new or understaffed schools.

1.6 Conclusions

Public-private partnerships in education are controversial and receive a great deal of attention from policymakers. Yet, there is little evidence for or against them in developing countries (Aslam et al., 2017). A typical argument in favor is that privately provided but publicly funded education is a means to inject cost-efficiency into education without compromising equity. A typical argument against is that outsourcing will lead to student selection and other negative, unintended consequences.

We present empirical evidence to test both arguments. The Partnership Schools for Liberia program, a public-private partnership that delegated *management* of 93 public schools (3.4% of all public schools) to eight different private organizations, was an effective way to circumvent weak public-sector management and improve learning outcomes. The ITT treatment effects of private management on student test scores after one academic year of treatment are $.18\sigma$ for English (p-value < 0.001) and $.18\sigma$ for math (p-value < 0.001).

We find no evidence that providers engaged in student selection — the probability of remaining in a treatment school was unrelated to a student's age, gender, household wealth, or disability. However, costs were high, performance varied across providers, and the largest provider pushed excess pupils and under-performing teachers into other government schools or completely out of the system. In addition, while outside the scope of our experimental analysis, the program has been plagued by accusations some operators failed to prevent, or actively concealed, sexual abuse in schools they managed. Teachers or staff of two PSL providers (More than Me Academy and Youth Movement for Collective Action) have been accused of sexual abuse since the start of the program, and an investigative report published in late 2018 alleged that More than Me Academy employed a serial child rapist in its schools prior to the start of PSL (Baysah, 2016; Young, 2018).

One interpretation of our results is that contracting rules matter. Changing the details of the contract might improve the results of the program. For instance, contracts could forbid class-size caps or require that students previously enrolled in a school be guaranteed re-admission once a school joins the PSL program. Similarly, contracts could require prior permission from the Ministry of Education before releasing a public teacher from their place of work. Stricter government oversight of child protection and vetting of private operators on this basis also appears warranted.

However, fixing the contracts and procurement process is not just a question of technical tweaks; it reflects a key governance challenge for the program. Contract differences reflect political influence: The largest provider opted not to take part in the competitive bidding process and made a separate bilateral agreement with the government. Ultimately, this agreement allowed pushing excess pupils and under-performing teachers into other government schools. This underlines the importance of uniform contracting rules and competitive bidding in a public-private partnership.

To our knowledge, we provide the first experimental estimates of the intention-to-treat (ITT) effect of outsourcing the management of existing public schools to private providers in a developing country. In contrast to the U.S. charter school literature, which focuses on experimental effects for the subset of schools and private providers where excess demand necessitates an admissions lottery, we provide treatment effects from across the distribution of outsourced schools in this setting.

However, an assortment of questions remain open for future research. First, given the bundled nature of this program, more evidence is needed to isolate the effect of outsourcing management. Variations of outsourcing also need to be studied (e.g., not allowing any teacher re-assignments, or allowing providers to hire teachers directly).

Second, while we identify sources of possible externalities from the program — e.g., pushing excess pupils into nearby schools — we are unable to study the effect of these externalities (positive or negative). Another key potential negative externality for other public schools is the opportunity cost of the program: PSL may deprive other schools of scarce resources by garnering preferential allocations of teachers or funding. On the other hand, traditional public schools may learn better management or pedagogical practices from nearby PSL schools. In addition, the program may lead to changes within the Ministry of Education that improve the performance of the system as a whole. For example, the need to monitor private providers has spurred the Ministry to reform some of its administrative information systems for all schools. All of this points to the need for future research to study these system-level effects and assess the impact of potentially important externalities.

1.7 Appendix

1.7.1 Additional tables and figures

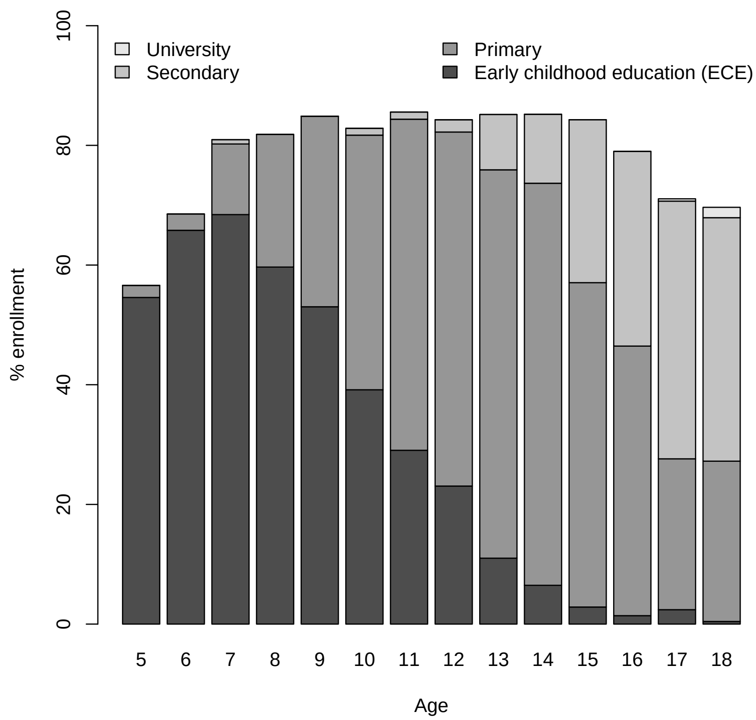
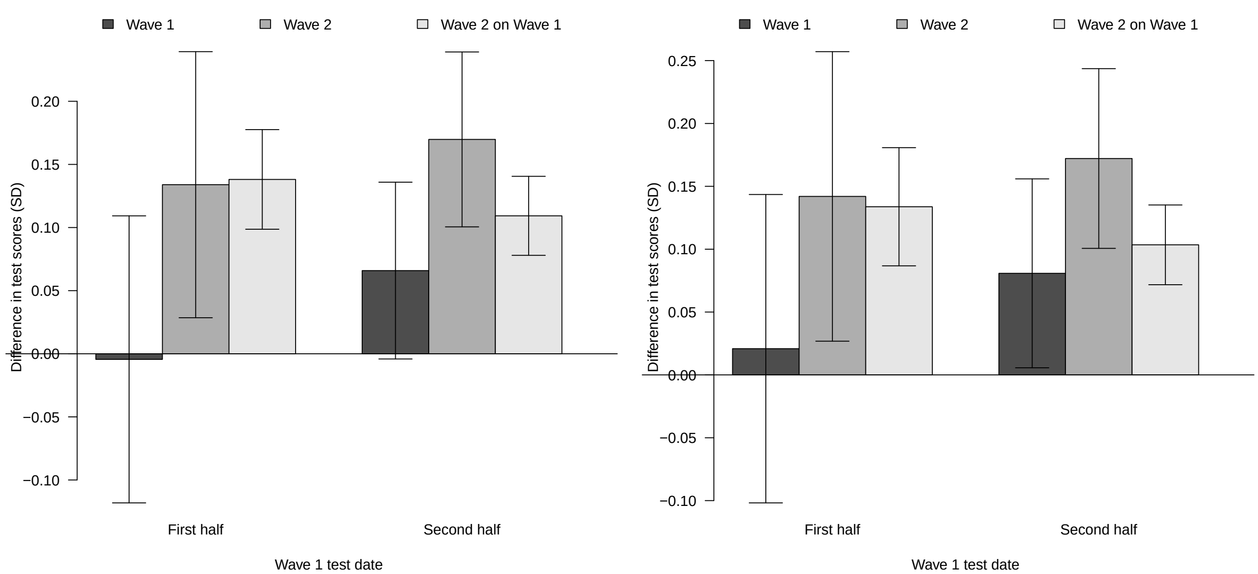


Figure 1.4: Enrollment by age
Note: This figure shows the percentage of students enrolled in early childhood education, primary schools, secondary schools, or universities by age. Authors' calculations based on 2014 Household Income and Expenditures Survey.

Research Activities	Year	Month	Intervention Activities
	2016	Jun	Operator selection
Randomization		Jul	
		Aug	
First Wave		Sep	School year begins
		Oct	
		Nov	
		Dec	
	2017	Jan	
		Feb	
		Mar	
		Apr	
Second Wave		May	
		Jun	
		Jul	School year ends

Figure 1.5: Timeline

Note:This figures provides a timeline for both research (left most column) and program (right most column) activities. Bridge signed its MOU with the Government of Liberia in March 2016, and thus started preparing for the program earlier than other providers.

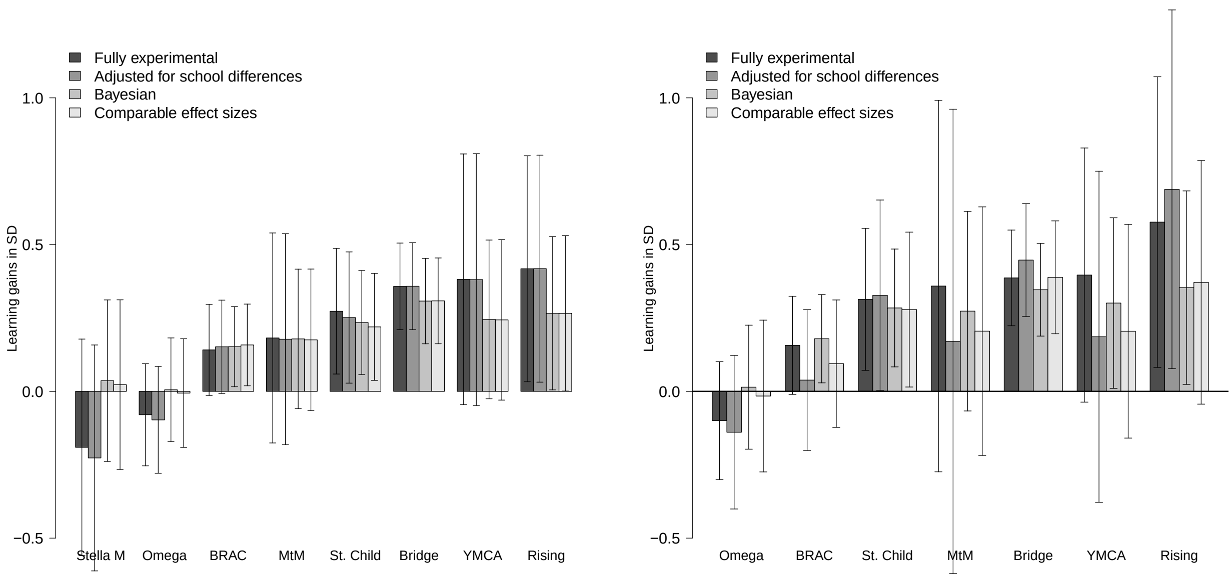


(a) Math

(b) English

Figure 1.6: Treatment effects by date tested during the first round of data collection

Note: These figures show the intention-to-treat treatment effects on test scores by whether during the first wave of data collection students were tested in the first or the second half of field work. Treatment effects are estimated using test scores from the first and the second wave of data collection. We also estimate the treatment effect during the second wave of data collection controlling for test scores during the first wave of data collection (Wave 2 on Wave 1). The panel on the left shows results for math test scores, while the panel on the right shows English test scores.

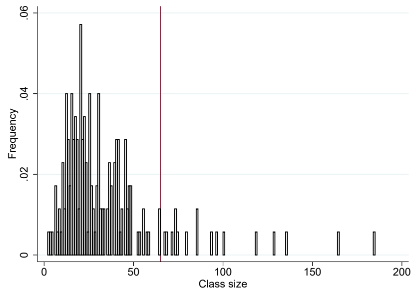


(a) Intention-to-treat (ITT) effect

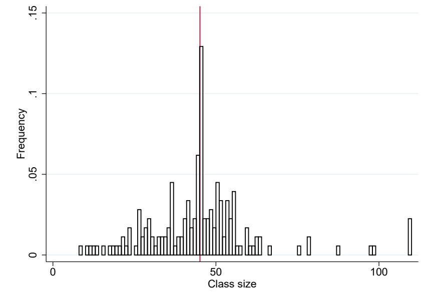
(b) Treatment-on-the-treated effect (ToT)

Figure 1.7: Treatment effects by provider

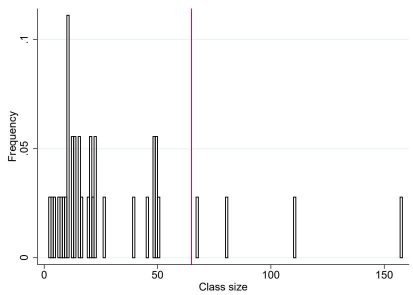
Note: These figures show the raw, fully experimental treatment effects, the effects after adjusting for differences in school characteristics (before the Bayesian hierarchical model), the effects after applying a Bayesian hierarchical model (but without adjusting for school differences), and the comparable treatment effects after adjusting for differences in school characteristics and applying a Bayesian hierarchical model. Figure 1.7a shows the intention-to-treat (ITT) effect, while Figure 1.7b shows the treatment-on-the-treated (ToT) effect. The ToT effects are larger than the ITT effects due to providers replacing schools that did not meet the eligibility criteria, providers refusing schools, or students leaving PSL schools. Stella Maris had full non-compliance at the school level and therefore there is no ToT effect for this provider.



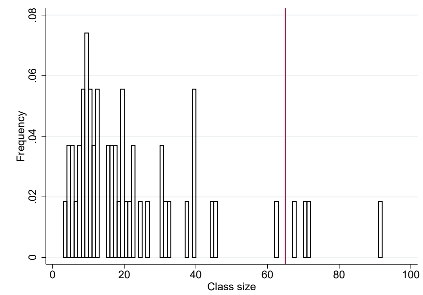
(a) BRAC



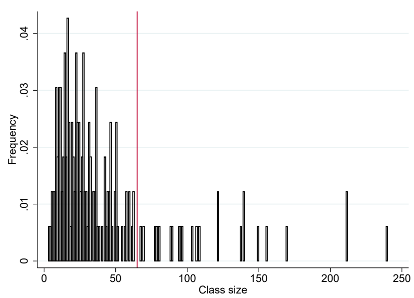
(b) Bridge



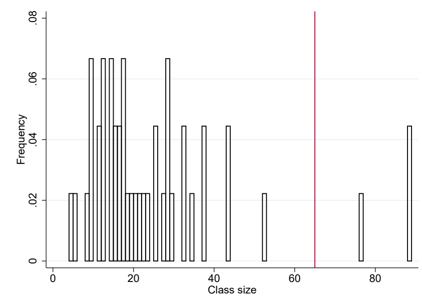
(c) YMCA



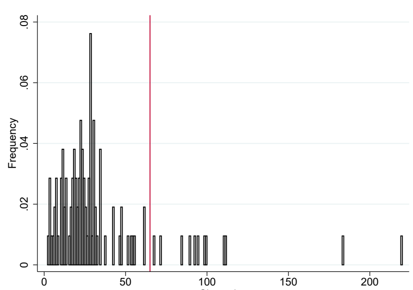
(d) More than Me



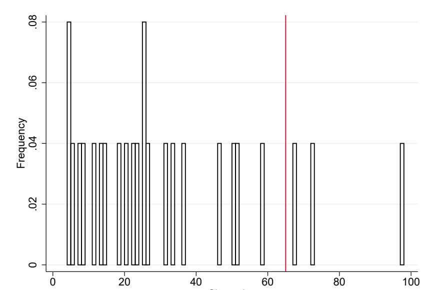
(e) Omega Schools



(f) Rising Academies



(g) St. Child



(h) Stella M

Figure 1.8: Class sizes and class caps

Note: These figures show the distribution of class sizes in treatment schools during the 2016/2017 academic year, as well as the class cap for each provider. The cap for all providers is 65 students, except for Bridge that has a cap of 45.

Table 1.11: External validity: Differences in characteristics of schools in the RCT (treatment and control) and other public schools (based on EMIS data)

	(1) RCT (Treatment and control)	(2) Other public schools	(3) Difference
Students: ECE	142.68	112.71	29.97
	(73.68)	(66.46)	(5.77)
Students: Primary	151.55	132.38	19.16
	(130.78)	(143.57)	(10.18)
Students	291.91	236.24	55.67
	(154.45)	(170.34)	(12.15)
Classrooms per 100 students	1.17	0.80	0.37
	(1.63)	(1.80)	(0.13)
Teachers per 100 students	3.04	3.62	-0.58
	(1.40)	(12.79)	(0.28)
Textbooks per 100 students	99.21	102.33	-3.12
	(96.34)	(168.91)	(7.88)
Chairs per 100 students	20.71	14.13	6.58
	(28.32)	(51.09)	(2.38)
Food from Gov or NGO	0.36	0.30	0.06
	(0.48)	(0.46)	(0.04)
Solid building	0.36	0.28	0.08
	(0.48)	(0.45)	(0.04)
Water pump	0.62	0.45	0.17
	(0.49)	(0.50)	(0.04)
Latrine/toilet	0.85	0.71	0.14
	(0.33)	(0.45)	(0.03)
Observations	185	2,420	2,605

This table presents the mean and standard error of the mean (in parentheses) for schools in the RCT (Column 1) and other public schools (Column 2), as well as the difference in means across both groups (Column 3). The sample of RCT schools is the original treatment and control allocation. ECE = Early childhood education. MOE= Ministry of Education. Authors' calculations based on Ministry of Education - Republic of Liberia (2015-2016) data. Standard errors are clustered at the school level.

Table 1.12: Number of schools by provider

	(1)	(2)	(3)	(4)	(5) (1)-(2)+(3)+(4)	(6) [(1)-(2)]/(1)
	Randomly assigned	Noncompliant	Replacement	Outside sample	Managed	% compliant in sample
BRAC	20	0	0	0	20	100%
Bridge	23	0	0	2	25	100%
YMCA	4	0	0	0	4	100%
MtM	6	2	2	0	6	67%
Omega	19	2	0	0	17	89%
Rising	5	1	0	1	5	80%
Stella	4	4	0	0	0	0%
St. Child	12	2	2	0	12	83%

The table shows the number of schools originally assigned to treatment (Column 1) and the schools that either did not meet Ministry of Education criteria or were rejected by providers (Column 2). The Ministry of Education provided replacement schools for those that did not meet the criteria, presenting each provider with a new list of paired schools and informing them, as before, that they would operate one of each pair (but not which one). Replacement schools are shown in Column 3. Column 4 contains non-randomly assigned schools given to some providers. Column 5 shows the final number of schools managed by each provider. Finally, the last column shows the percentage of schools actually managed by the provider that are in our main sample.

Table 1.13: Balance table: Differences in characteristics of treatment and control schools, pre-treatment year (2015/2016, EMIS data)

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Students: ECE	148.51	136.72	11.79	11.03
	(76.83)	(70.24)	(10.91)	(9.74)
Students: Primary	159.05	143.96	15.10	15.68
	(163.34)	(86.57)	(19.19)	(16.12)
Students	305.97	277.71	28.26	27.56
	(178.49)	(124.98)	(22.64)	(19.46)
Classrooms per 100 students	1.21	1.13	0.09	0.08
	(1.62)	(1.65)	(0.24)	(0.23)
Teachers per 100 students	3.08	2.99	0.09	0.09
	(1.49)	(1.30)	(0.21)	(0.18)
Textbooks per 100 students	102.69	95.69	7.00	7.45
	(97.66)	(95.40)	(14.19)	(13.74)
Chairs per 100 students	18.74	22.70	-3.96	-4.12
	(23.06)	(32.81)	(4.17)	(3.82)
Food from Gov or NGO	0.36	0.36	-0.01	-0.01
	(0.48)	(0.48)	(0.08)	(0.05)
Solid building	0.39	0.33	0.06	0.06
	(0.49)	(0.47)	(0.07)	(0.06)
Water pump	0.56	0.67	-0.11	-0.12
	(0.50)	(0.47)	(0.07)	(0.06)
Latrine/toilet	0.85	0.86	-0.01	-0.01
	(0.35)	(0.32)	(0.05)	(0.05)
Observations	93	92	185	185

This table presents the mean and standard error of the mean (in parenthesis) for the control (Column 1) and treatment (Column 2), as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Authors' calculations based on Ministry of Education - Republic of Liberia (2015-2016) data.

Table 1.14: ITT and ToT effect after one year

	Wave 2			Wave 2 on Wave 1		
	Math (1)	English (2)	Abstract (3)	Math (4)	English (5)	Abstract (6)
Panel A: ITT						
Treatment	0.18	0.18	0.046	0.14	0.13	0.032
	(0.034)	(0.031)	(0.038)	(0.023)	(0.021)	(0.036)
No. of obs.	3,492	3,492	3,492	3,492	3,492	3,492
Panel B: ToT						
Treatment	0.22	0.21	0.056	0.17	0.16	0.038
	(0.040)	(0.037)	(0.045)	(0.027)	(0.025)	(0.043)
No. of obs.	3,492	3,492	3,492	3,492	3,492	3,492

This table presents the treatment effect on test scores at the end of the schools year. Columns 1-3 use a specification that takes into account the randomization design — i.e., includes "pair" fixed effects — and includes student and school controls. Columns 4-6 also control for test scores collected during the first wave of data collection. Panel A has the intention-to-treat (ITT) effect, while Panel B has the treatment-on-the-treated (ToT) effect. The treatment-on-the-treated effect is estimated using the assigned treatment as an instrument for whether the student is in fact enrolled in a PSL school during the 2016/2017 academic year. Standard errors are clustered at the school level.

Table 1.15: Control variables

Student controls	Question	Questionnaire
Wealth index	A1-A7	Student
Age	B1	Student
Gender	B2	Student
Grade (2015/2016)	B6a	Student
School controls		
Enrollment (2015/2016)	C1	Principal
Infrastructure quality (2015/2016)	L1-L3	Principal
Travel time to nearest bank	L6	Principal
Rurality	L7	Principal
NGO programs in 2015/2016	M1-M4	Principal
Donations in 2015/2016	N1A-N3b_a_5	Principal

This table shows the control variables included in specification 1.2.

Table 1.16: Treatment effects across various measures of difference in student ability

	ITT (1)	ToT (2)
Panel A: Base IRT model		
English	0.18	0.21
	(0.03)	(0.04)
Math	0.18	0.22
	(0.03)	(0.04)
Panel B: Base IRT model standarized by grade		
English	0.23	0.28
	(0.04)	(0.05)
Math	0.23	0.27
	(0.04)	(0.05)
Panel C: PCA		
English	0.16	0.19
	(0.03)	(0.04)
Math	0.24	0.28
	(0.04)	(0.05)
Panel D: PCA standarized by grade		
English	0.19	0.23
	(0.04)	(0.05)
Math	0.28	0.33
	(0.05)	(0.06)
Panel E: % correct answers		
English	2.96	3.56
	(0.55)	(0.66)
Math	4.24	5.09
	(0.71)	(0.84)
Observations	3,492	3,492

Column 1 shows the intention-to-treat treatment effect estimated with a specification that takes into account the randomization design — i.e., includes "pair" fixed effects — and includes for student and school controls. The treatment-on-the-treated effect (Column 2) is estimated using the assigned treatment as an instrument for whether the student is in fact enrolled in a PSL school during the 2016/2017 academic year. Panel A uses our default IRT model and normalizes test scores using the same mean and standard deviation across all grades. Panel B uses the same IRT model as Panel A, but normalizes test scores using a different mean and standard deviation for each grade. Panel C estimates students' ability as the first component from a principal component analysis (PCA), and normalizes test scores using a common mean and standard deviation across all grades. Panel D uses the same model as Panel C but normalizes test scores using a different mean and standard deviation per grade. Panel E calculates the percentage of correct responses. Standard errors are clustered at the school level.

Table 1.17: Treatment effect on instruction time by subject

	Hours per week			% time per week		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.70	0.34	0.34	0.87	-0.38	-0.38
	(0.11)	(0.13)	(0.13)	(0.26)	(1.20)	(1.20)
Math or English	2.08	1.71		8.74	7.47	
	(0.13)	(0.16)		(1.10)	(1.60)	
Treatment × Math or English		0.71			2.50	
		(0.26)			(2.18)	
Math			0.34			-0.86
			(0.075)			(2.02)
English			3.08			15.8
			(0.30)			(2.09)
Treatment × Math			0.65			7.63
			(0.17)			(2.72)
Treatment × English			0.77			-2.63
			(0.47)			(2.83)
No. of obs.	4,299	4,299	4,299	4,299	4,299	4,299

This table presents the treatment effect on instruction time by subject. The outcome is hours per week in Columns 1-3. The outcome in Columns 4-6 is fraction of time per week. The unit of observation is at the grade-subject level. All regressions take into account the randomization design (i.e., include "pair" fixed effects). Standard errors are clustered at the school level.

Table 1.18: Heterogeneity by student characteristics

	Male (1)	Top wealth quartile (2)	Bottom wealth quartile (3)	Grade (4)
Treatment	0.20	0.18	0.17	0.16
	(0.047)	(0.035)	(0.035)	(0.10)
Treatment × covariate	-0.021	0.029	0.061	0.0050
	(0.068)	(0.066)	(0.050)	(0.020)
No. of obs.	3,492	3,492	3,492	3,492

The outcome variable is the test scores at the end of the school year. All regressions include "pair" fixed effects and include student and school controls. Each column shows the interaction of a different covariate with treatment. Standard errors are clustered at the school level.

Table 1.19: Student selection

	(1) Same school	(2) Same school	(3) Same school
Treatment	0.061	0.012	0.021
	(0.082)	(0.026)	(0.019)
Treatment × Age	-0.0042		
	(0.0064)		
Treatment × Male		-0.011	
		(0.028)	
Treatment × Asset Index (PCA)			-0.0059
			(0.011)
No. of obs.	3,487	3,487	3,428

The outcome variable is whether the student is enrolled at the end of the 2016/2017 school year in the same schools he or she was enrolled in the 2015/2016 school year. All regressions include "pair" fixed effects. Standard errors are clustered at the school level.

Table 1.20: ITT treatment effects, by whether class size caps are binding without including adjacent grades

	(1) Δ enrollment	(2) % same school	(3) % in school	(4) Test scores
Constrained=0 × Treatment	5.37	4.41	1.67	0.14
	(1.15)	(1.43)	(0.67)	(0.036)
Constrained=1 × Treatment	-8.92	-16.8	-0.051	0.41
	(6.26)	(8.01)	(4.15)	(0.14)
No. of obs.	1,635	3,637	3,485	3,490
Mean control (Unconstrained)	-0.70	81.63	93.45	0.12
Mean control (Constrained)	-7.62	87.14	94.12	-0.07
α_0 = Constrained - Unconstrained	-14.29	-21.20	-1.72	0.27
p-value ($H_0 : \alpha_0 = 0$)	0.04	0.01	0.68	0.07

This table mirrors Table 1.5, but adjacent grades are not included in the calculation of the constrained indicator. Column 1 uses school-grade level data. Columns 2 - 4 use student level data. The independent variable in Column 4 is the composite test score. The sample is the original treatment and control allocation. There were 216 constrained classes at baseline (holding 35% of students), and 1,448 unconstrained classes at baseline (holding 65% of students). Standard errors are clustered at the school level.

Table 1.21: Intensive margin effect on teacher attendance and classroom observation with Lee bounds

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)	(5) 90% CI (bounds)
Panel A: Spot check (N = 929)					
% on schools campus	68.15	52.40	15.75	14.17	2.67
	(46.64)	(50.00)	(4.45)	(3.75)	27.96
% in classroom	50.96	41.05	9.91	9.96	-1.21
	(50.04)	(49.25)	(4.78)	(3.86)	24.26
B: Classroom observation (N = 143)					
Active instruction (% class time)	38.12	30.13	7.98	7.62	-4.75
	(28.93)	(32.11)	(4.86)	(4.75)	19.92
Passive instruction (% class time)	16.24	12.80	3.44	4.72	-4.93
	(17.18)	(19.83)	(2.95)	(3.23)	9.62
Classroom management (% class time)	20.82	10.67	10.16	10.33	0.77
	(21.06)	(14.83)	(2.85)	(3.32)	16.99
Teacher off-task (% class time)	24.82	46.40	-21.58	-22.66	-40.24
	(32.65)	(41.09)	(5.92)	(6.26)	-10.32
Student off-task (% class time)	55.06	57.60	-2.54	-5.19	-16.05
	(31.23)	(34.87)	(5.26)	(4.88)	12.63
Panel C: Inputs (N = 143)					
Number of seats	20.64	20.58	0.06	0.58	-7.22
	(13.33)	(13.57)	(2.21)	(1.90)	5.36
% with students sitting on the floor	2.41	4.23	-1.82	-1.51	-7.48
	(15.43)	(20.26)	(2.94)	(2.61)	2.76
% with chalk	96.39	78.87	17.51	16.58	9.47
	(18.78)	(41.11)	(5.29)	(5.50)	27.85
% of students with textbooks	37.08	17.60	19.48	22.60	-1.21
	(43.22)	(35.25)	(6.33)	(6.32)	34.87
% of students with pens/pencils	88.55	79.67	8.88	8.16	1.36
	(19.84)	(30.13)	(4.19)	(4.10)	20.98

This table presents the mean and standard error of the mean (in parenthesis) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Column 5 shows the 90% confidence interval using Lee (2009) bounds. Panel A provides results from the spot check using the Ministry of Education - Republic of Liberia (2015-2016) data on teachers as a baseline, and treating teachers who no longer teach at school as attriters. Panel B and C provide the classroom observation information without imputing values for schools not in session during our visit, and treating the missing information as attrition. Standard errors are clustered at the school level.

Table 1.22: Treatment effect on schools' good practices

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Maintains an enrollment log	0.90	0.80	0.10	0.10
	(0.30)	(0.40)	(0.05)	(0.05)
Log contains student name	0.89	0.82	0.08	0.08
	(0.31)	(0.39)	(0.05)	(0.05)
Log contains student grade	0.94	0.84	0.10	0.10
	(0.25)	(0.37)	(0.05)	(0.05)
Log contains student age	0.65	0.64	0.00	0.00
	(0.48)	(0.48)	(0.07)	(0.07)
Log contains student gender	0.89	0.83	0.07	0.06
	(0.31)	(0.38)	(0.05)	(0.05)
Log contains student contact information	0.26	0.13	0.13	0.13
	(0.44)	(0.34)	(0.06)	(0.06)
Enrollment log is clean and neat	0.39	0.26	0.13	0.13
	(0.49)	(0.44)	(0.07)	(0.07)
Maintains official schedule	0.98	0.89	0.09	0.09
	(0.15)	(0.31)	(0.04)	(0.03)
Official schedule is posted	0.84	0.70	0.14	0.14
	(0.37)	(0.46)	(0.06)	(0.06)
Has a PTA	0.99	0.98	0.01	0.01
	(0.10)	(0.15)	(0.02)	(0.02)
Principal has PTA head's number at hand	0.41	0.26	0.15	0.15
	(0.49)	(0.44)	(0.07)	(0.06)
Maintains expenditure records	0.14	0.09	0.05	0.05
	(0.35)	(0.28)	(0.05)	(0.05)
Maintains a written budget	0.26	0.22	0.04	0.04
	(0.44)	(0.41)	(0.06)	(0.06)
Observations	93	92	185	185

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Standard errors are clustered at the school level.

Table 1.23: Treatment effect on household expenditure

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Fees (USD/year)	5.68	8.06	-2.38	-2.95
	(10.16)	(9.73)	(0.97)	(0.62)
Tutoring (USD/year)	0.34	0.38	-0.04	-0.04
	(1.20)	(1.34)	(0.09)	(0.08)
Textbooks (USD/year)	0.61	0.86	-0.25	-0.22
	(1.44)	(1.65)	(0.13)	(0.09)
Copy books (USD/year)	1.02	1.08	-0.06	-0.07
	(1.96)	(1.93)	(0.14)	(0.13)
Pencils (USD/year)	3.23	2.95	0.28	0.21
	(3.05)	(2.88)	(0.31)	(0.16)
Uniform (USD/year)	9.25	11.46	-2.20	-1.95
	(6.31)	(5.19)	(0.63)	(0.43)
Food (USD/year)	43.00	46.33	-3.33	-1.46
	(71.02)	(75.85)	(6.90)	(3.91)
Other (USD/year)	3.42	3.06	0.36	0.32
	(4.56)	(4.28)	(0.34)	(0.27)
Observations	595	520	1,115	1,115

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Standard errors are clustered at the school level.

Table 1.24: Treatment effect on household engagement

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Attended school meeting	0.76	0.77	-0.01	0.03
	(0.43)	(0.42)	(0.04)	(0.02)
Made cash donation	0.12	0.11	0.02	-0.00
	(0.33)	(0.31)	(0.02)	(0.02)
Made in-kind donation	0.03	0.04	-0.01	-0.02
	(0.17)	(0.20)	(0.01)	(0.01)
Donated work	0.13	0.15	-0.01	-0.00
	(0.34)	(0.35)	(0.03)	(0.02)
Helped with homework	0.58	0.61	-0.03	-0.04
	(0.49)	(0.49)	(0.04)	(0.03)
Observations	619	543	1,162	1,162

This table presents the mean and standard error of the mean (in parenthesis) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Standard errors are clustered at the school level.

Table 1.28: Simulated treatment effects without some providers

	(1) Raw	(2) Comparable
All operators	.18	.18
	(.032)	(.044)
Without Stella M and Omega	.27	.23
	(.035)	(.046)
Without Bridge	.12	.16
	(.037)	(.045)
Without Stella M, Omega, and Bridge	.21	.2
	(.042)	(.054)
Without Stella M, Omega, Bridge, MtM, and YMCA	.21	.19
	(.047)	(.063)

This table presents the average treatment effect of the program by taking an inverse-variance weighted average across providers. Column 1 presents the overall treatment effect from the raw treatment estimates. Column 2 presents the overall treatment effect from the comparable treatment effects after adjusting for differences in school characteristics and applying a Bayesian hierarchical model. The additional uncertainty for the comparable treatment effects (i.e., larger standard errors) comes from assuming a non-informative prior for the standard deviation of true effects.

Table 1.25: Baseline differences between treatment schools and average public schools, by provider

	(1) BRAC	(2) Bridge	(3) MtM	(4) Omega	(5) Rising	(6) St. Child	(7) Stella M	(8) YMCA	(9) p-value
Students	31.94	156.19	-23.03	35.49	-0.83	31.09	-19.16	-22.53	.00092
	(27.00)	(25.48)	(49.01)	(27.69)	(53.66)	(34.74)	(59.97)	(59.97)	
Teachers	1.23	2.72	1.42	1.70	1.16	0.59	1.13	0.76	.66
	(0.70)	(0.66)	(1.28)	(0.72)	(1.40)	(0.90)	(1.56)	(1.56)	
PTR	-4.57	5.77	-8.47	-5.45	-6.02	2.34	-10.62	-7.29	.079
	(3.27)	(3.09)	(5.94)	(3.36)	(6.50)	(4.21)	(7.27)	(7.27)	
Latrine/Toilet	0.18	0.28	0.26	0.25	0.23	0.22	0.06	0.18	.96
	(0.08)	(0.07)	(0.14)	(0.08)	(0.16)	(0.10)	(0.17)	(0.17)	
Solid classrooms	0.63	2.81	2.64	-0.11	1.85	1.59	-1.95	1.30	.055
	(0.75)	(0.71)	(1.36)	(0.77)	(1.49)	(0.97)	(1.67)	(1.67)	
Solid building	0.28	0.22	0.19	0.09	0.26	0.19	0.23	0.23	.84
	(0.08)	(0.07)	(0.14)	(0.08)	(0.15)	(0.10)	(0.17)	(0.17)	
Nearest paved road (KM)	-9.25	-10.86	-7.13	-8.22	-4.47	-7.13	-4.56	-7.79	.78
	(2.03)	(1.91)	(3.67)	(2.08)	(4.01)	(2.60)	(4.48)	(4.48)	

This table presents the difference between public schools and the schools operated by each provider. The information for all schools is taken from the Ministry of Education - Republic of Liberia (2015-2016) data, and therefore is pre-treatment information. Column 9 shows the p-value for testing $H_0 : \beta_{BRAC} = \beta_{Bridge} = \beta_{YMCA} = \beta_{MtM} = \beta_{Omega} = \beta_{Rising} = \beta_{St.Child} = \beta_{StellaM}$. Standard errors are clustered at the school level. The sample is the original treatment and control allocation. Since some providers had no schools with classes above the class caps, there is no data to estimate treatment effects over constrained classes. Standard errors are clustered at the school level.

Table 1.26: Comparable ITT treatment effects by provider

	(1) BRAC	(2) Bridge	(3) MtM	(4) Omega	(5) Rising	(6) St. Child	(7) Stella M	(8) YMCA	(9) p-value
Panel A: Student test scores									
English (standard deviations)	0.19	0.24	0.17	0.00	0.23	0.18	0.01	0.30	0.087
	(0.08)	(0.08)	(0.14)	(0.12)	(0.16)	(0.11)	(0.18)	(0.18)	
Math (standard deviations)	0.12	0.33	0.18	0.01	0.26	0.22	0.03	0.20	0.025
	(0.09)	(0.09)	(0.15)	(0.11)	(0.16)	(0.11)	(0.17)	(0.16)	
Composite (standard deviations)	0.16	0.31	0.18	-0.01	0.27	0.22	0.02	0.24	0.035
	(0.08)	(0.09)	(0.15)	(0.11)	(0.16)	(0.11)	(0.18)	(0.17)	
Panel B: Changes to the pool of teachers									
% teachers dismissed	-8.74	50.63	14.33	-5.87	0.78	-2.27	-7.52	11.81	<0.001
	(6.37)	(7.03)	(11.05)	(6.66)	(11.86)	(8.89)	(12.97)	(12.95)	
% new teachers	38.06	70.90	47.84	22.76	20.76	36.37	-8.23	35.94	0.0060
	(11.14)	(13.09)	(18.73)	(11.88)	(19.98)	(15.15)	(26.11)	(21.04)	
Age in years (teachers)	-5.50	-9.12	-7.80	-5.75	-8.07	-6.53	-5.99	-3.50	0.16
	(1.71)	(2.18)	(2.57)	(1.73)	(2.74)	(2.08)	(2.72)	(3.51)	
Test score in standard deviations (teachers)	0.12	0.24	0.23	0.17	0.17	0.23	0.17	0.05	0.46
	(0.13)	(0.14)	(0.18)	(0.13)	(0.18)	(0.16)	(0.18)	(0.23)	
Panel C: Enrollment and access									
Δ enrollment	31.91	7.80	12.63	28.93	16.24	25.38	15.79	27.52	0.49
	(25.45)	(26.78)	(32.57)	(25.01)	(32.75)	(28.81)	(33.86)	(34.02)	
Δ enrollment (constrained grades)	41.57	-29.66	41.64	-3.45	41.30	22.33	–	–	0.48
	(44.13)	(14.58)	(44.24)	(36.91)	(44.03)	(46.95)	(–)	(–)	
Student attendance (%)	18.45	12.77	20.72	17.55	19.01	19.36	16.65	17.44	0.48
	(6.59)	(7.54)	(9.14)	(6.67)	(8.98)	(7.95)	(9.49)	(9.09)	
% students still attending any school	-1.98	1.27	-4.78	-2.03	-3.82	-1.97	-3.21	-3.18	0.36
	(3.38)	(3.68)	(5.90)	(3.64)	(5.62)	(4.26)	(5.35)	(5.60)	
% students still attending same school	0.55	2.34	0.39	0.65	0.76	0.27	0.33	0.20	0.45
	(1.75)	(1.91)	(2.53)	(1.87)	(2.55)	(2.22)	(2.58)	(2.72)	
Panel D: Satisfaction									
% satisfied with school (parents)	11.75	10.97	3.72	1.72	2.59	-0.37	9.96	8.51	0.24
	(6.32)	(6.39)	(8.45)	(6.32)	(9.00)	(8.36)	(9.36)	(9.14)	
% students who think school is fun	4.08	2.73	2.49	3.25	3.47	2.59	0.01	4.75	0.59
	(3.90)	(3.66)	(5.42)	(4.05)	(5.59)	(4.65)	(6.67)	(6.12)	
heightObservations	40	45	8	12	38	10	24	8	

This table presents the ITT treatment effect for each provider, after adjusting for differences in baseline school characteristics, based on a Bayesian hierarchical model. Column 9 shows the p-value for testing $H_0 : \beta_{BRAC} = \beta_{Bridge} = \beta_{YMCA} = \beta_{MtM} = \beta_{Omega} = \beta_{Rising} = \beta_{St.Child} = \beta_{StellaM}$. Table 1.10 has the raw experimental treatment effects by provider. Standard errors are shown in parentheses. Estimation is conducted on collapsed, school-level data.

Table 1.27: Descriptive statistics by provider and treatment

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Provider	Treatment	Schools	Teachers				Enrollment			Enrollment in constrained classes		
			2015/2016	2016/2017	Dismissed	New	Classes	2015/2016	2016/2017	Constrained classes	2015/2016	2016/2017
BRAC	0	20	141	148	41	48	180	5,694	5,107	10	780	703
BRAC	1	20	141	209	33	101	180	5,684	5,872	11	1,130	1,138
Bridge	0	22	177	174	38	35	198	7,110	6,610	61	3,969	3,648
Bridge	1	23	236	212	174	150	207	9,788	8,282	72	6,909	3,475
YMCA	0	4	20	22	1	3	36	729	727	2	142	120
YMCA	1	4	27	40	6	19	36	908	1,068	2	217	238
MtM	0	6	52	41	21	10	54	1,140	1,312	2	155	167
MtM	1	6	46	64	20	38	54	1,145	1,223	2	171	159
Omega	0	19	132	130	33	31	171	4,895	5,200	12	1,255	1,232
Omega	1	19	151	196	26	71	171	5,764	6,841	19	1,953	2,446
Rising	0	5	47	43	23	19	45	1,209	1,308	2	202	185
Rising	1	5	36	47	11	22	45	918	1,134	1	87	89
St. Child	0	12	88	68	29	9	108	3,094	2,794	7	738	557
St. Child	1	12	81	100	22	41	108	3,351	3,506	9	877	797
Stella M	0	4	20	20	8	8	36	765	683	1	73	45
Stella M	1	4	31	27	9	5	36	958	978	3	213	192

This table shows the total number of teachers and students in treatment (=1 in Column 2) and control (=0 in Column 2) schools for each operator. Teachers in 2015/2016 are taken from the Ministry of Education - Republic of Liberia (2015-2016) data, while teachers in 2016/2017 are taken from our first-year follow-up data. “Dismissed” refers to the number of teachers in the Ministry of Education - Republic of Liberia (2015-2016) data who are not working at the school at the end of the 2016/2017 academic year. “New” is the number of teachers working at the school at the end of the 2016/2017 academic year who are not in the Ministry of Education - Republic of Liberia (2015-2016) data. “Constrained classes” are those with more students in 2015/2016 than the class size cap.

1.7.2 Tracking and attrition

A potential issue with our sampling strategy is differential attrition at each round of data collection. In the first round, enumerators were instructed to sample 20 students from the 2015/2016 enrollment logs, track them, and test them. However, if a student had moved to another village, had died, or was impossible to track, the enumerators were instructed to sample another student. Thus, even at the first round an endogenous sampling problem arises if treatment makes students easier or harder to track in combination with enumerator shrinkage. To mitigate this issue, enumerators participated in additional training on tracking and its importance and were provided with a generous amount of tracking time. Students were tracked to their homes and tested there when not available at school. As Table 1.29 shows, we have no reason to believe that this issue arose. The effort required to track students was different between treatment and control (it is easier to track students at the school), yet the total number of students sampled, to obtain a sample of 20 students, is balanced between treatment and control (see Table 1.29).

Table 1.29: Tracking and sampling in the first wave of data collection

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Number of students sampled	24.8	24.6	0.13	0.035
	(5.74)	(5.10)	(0.81)	(0.81)
Found at the school	18.2	16.7	1.49	1.555
	(2.30)	(4.70)	(0.55)	(0.54)
Found at home	1.73	2.91	-1.18	-1.223
	(2.12)	(3.97)	(0.48)	(0.47)
Interviewed	19.9	19.6	0.32	0.344
	(0.89)	(2.23)	(0.25)	(0.27)
Observations	88	90	178	171

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including “pair” fixed effects) in Column 4. The table shows the average number of students we sampled (and tried to track), the number of students we were able to track at the assigned school or at home, and the total number of students we tracked and found during the first round of data collection. Standard errors are clustered at the school level.

1.7.3 Test design

Most modules follow the Early Grade Reading Assessment (EGRA), Early Grade Mathematics Assessment (EGMA), Uwezo, and Trends in International Mathematics and Science Study (TIMSS) assessments. For the first wave of data collection the test contained a module for each of the following skills: object identification (like the Peabody Picture Vocabulary Test), letter reading (adapted from EGRA), word reading (adapted from EGRA), a preposition module, reading comprehension (adapted from Uwezo), listening comprehension (adapted from EGRA), counting (adapted from Uwezo), number discrimination (adapted from Uwezo), number identification (adapted from EGMA), addition (adapted from Uwezo and EGMA), subtraction (adapted from Uwezo and EGMA), multiplication (adapted from Uwezo and EGMA), division (adapted from Uwezo and EGMA), shape identification, fractions, and word problems in mathematics.

For the second round of data collection the test did not include the following modules: Prepositions, shape identification, and fractions. These modules were excluded given the low variation in responses in the first wave of data collection and to make space for new modules. Instead, new modules were introduced, including letter, word and number dictation, and a verb and a pronoun module. Additionally, we included some "conceptual" questions from TIMSS released items (items M031317 and M031316) that do not resemble the format of standard textbook exercises but rather test knowledge in an unfamiliar way. The number identification module remained exactly the same across rounds of data collection (to provide us with absolute learning curves on these two items), while every other module was different. In addition, the word and number identification modules were identical to the EGRA/EGMA assessments used in Liberia previously (for comparability with other impact evaluations taking place in Liberia, most notably USAID's reading program (Piper & Korda, 2011) and the LTTP program (King et al., 2015)), but during the first round of data collection they were different. Two of the reading comprehension questions were taken from the Pre-Pirls released items (L11L01C and L11L02M) and one of the word problems was taken from TIMSS released items (M031183). Finally, we added a Raven's style module to measure the students' abstract thinking abilities.

1.7.4 Cost-benefit analysis

To compute the net benefit of the PSL program, we must place a financial value on test score gains.[28] To map test score gains onto existing estimates of the Mincerian returns to schooling, we translate them into "equivalent years of schooling" (EYOS) following Evans and Yuan (2017). This calculation assumes the value of schooling is captured by test score gains. In the control group, an extra year of schooling is associated with average test score gains of $.3\sigma$ across subjects. Thus, the ToT treatment effect of $.22\sigma$ on test scores is roughly 0.74 EYOS.

We estimate whether the lifetime benefits accruing to students outweigh the costs of the program by estimating the net present value (NPV) of the investment. We make four key assumptions: 1) every extra year of schooling increases income by 8.6% (Millennium Challenge Corporation, 2013); 2) the discount rate is 10%;[29] 3) pupils' counterfactual adult earnings are equal to Liberia's GDP per capita of USD 694.3 in 2017 (World Bank, 2017); and 4) students work for 40 years (starting in ten years time). The NPV of investing USD 50 in the PSL program is positive, at $\sim$USD 110; if we take into account the cost of additional teachers, the NPV is $\sim$USD 90. NPV becomes negative when the investment is above $\sim$USD 160, which is the case for some providers in the short run.

How sensitive are these estimates to our assumptions? A lower discount rate of, say, 4%[30] would raise the minimum threshold cost to achieve positive NPV to USD 570 (still below the expenditure for some providers in the short run).

1.7.5 Provider's details

[28] We ignore non-pecuniary returns to education.

[29] This is the discount rate used by the World Bank to estimate Net ODA (World Bank, 2013).

[30] For reference, the deposit interest rate was 3.8 in 2016 (World Bank, 2016) and the latest auction of treasury bills from the Central Bank resulted in an average discount rate of 3.69% (see `https://www.cbl.org.lr/2content.php?sub=191&related=33&third=191&pg=sp&pt=Treasury%20Bills`)

Table 1.30: Provider's characteristics

Provider	For-profit	Local	Website/mission/vision/values
BRAC	No	No	`www.bracinternational.nl` Our Vision: A world free from all forms of exploitation and discrimination where everyone has the opportunity to realise their potential. Our Mission: Our mission is to empower people and communities in situations of poverty, illiteracy, disease and social injustice. Our interventions aim to achieve large scale, positive changes through economic and social programmes that enable men and women to realise their potential. Our Values: Integrity. Innovation. Inclusiveness. Effectiveness.
Bridge	Yes	No	`www.bridgeinternationalacademies.com` At Bridge, our mission is to provide millions of children with a life-changing education. We are committed to helping achieve United Nations Sustainable Development Goal 4: to ensure inclusive and quality education for all and promote lifelong learning. We believe every child has the right to education. To realise this, we work in partnership with governments, communities, teachers, and parents to deliver great schools and high-quality education for primary and pre-primary pupils. By providing a life-changing education to children in underserved communities, we help put children, their families, and their countries on a better development path. Education is one of the most effective ways to end the cycle of poverty; enabling growth, peace, and prosperity. It reduces conflict, enhances stability, and strengthens nations.
MtM	No	No	`www.morethanme.org` It All Started with a Girl: At 11 years old, she was selling herself for clean drinking water, when all she really wanted was to go to school. More Than Me was founded to meet her needs. We opened a school, the MTM Academy, for 150 girls and saw major progress. Then Ebola hit. The world didn't act quickly enough, and we knew we had to fight to save her life. After spending six months on the front lines fighting Ebola and seeing three students lose their families (and many other children lose their own lives), we realized that our girls would never be safe, never truly thrive, until Liberia thrives. The first step to rebuilding Liberia is education for all. Our Mission: More Than Me uses education as a catalyst for transformative social change for every girl in Liberia. Our Vision: Every girl empowered.

Continued on next page

Table 1.30 – continued from previous page

Provider	For-profit	Local	Website/mission/vision/values
Omega	Yes	No	`www.omega-schools.com` The enterprising poor, in huge numbers, are voting with their feet. They are sacrificing their meager earnings to put their wards into low-cost private schools. Omega Schools has responded with an innovative Pay-As-You-Learn model — a chain of low cost private schools with specialized curriculum, assessment, technology and management modules that are benefiting the poor and empowering aspirations of low income families and their communities. Founded by Ken & Lisa Donkoh, and James Tooley in 2008, and backed by Pearson's Affordable Learning Fund, Omega Schools is a social enterprise on a mission to deliver quality education at the lowest cost on a grand scale. The model has proven to be extremely attractive to parents, enabling its schools to be full within 10 days of opening. Currently the chain has 38 schools educating over 20,000 students and seeking to double that number in a year.
Rising	Yes	No	`www.risingacademies.com` Our mission is to create schools that open doors and change lives. We expect every student to leave Rising ready for further study, for a good job, and to become a role model in society. Founded in Sierra Leone in 2014, we provided emergency education to children kept out of school by the Ebola Epidemic before opening our first school in April 2015. Today, we run 29 government schools in rural Liberia under the LEAP (formerly the Partnership Schools for Liberia) initiative, and 10 schools in Sierra Leone serving families looking for a high quality education at an affordable cost.
St. Child	No	No	`www.street-child.co.uk` Every child deserves the chance to go to school and learn. Who We Are: 121 million school-aged children are currently out of education world-wide. Millions more children are in school but failing to learn. Street Child believes that achieving universal basic education is the single greatest step that can be taken towards the elimination of global poverty. Where We Work: Street Child prides itself on being willing to go to the world's toughest places where others won't, including remote, hard-to-reach areas and fragile, disaster-affected states. What We Do: We recognise that the barriers to education are complex and interlinked, and our projects focus on a combination of education, child protection and livelihood support to address the social, economic and structural issues that underpin today's education crisis. Wherever we work we partner with local organisations and communities and take an outcome-led approach. We use evidence to drive learning and the constant refinement and scale-up of programmes that create maximum impact for the most children at the lowest cost.

Continued on next page

Table 1.30 – continued from previous page

Provider	For-profit	Local	Website/mission/vision/values
Stella M	No	Yes	www.smp.edu.lr/index.html The Stella Maris Polytechnic is an Institution of higher learning own and operated by the Catholic Archdioces of Monrovia. With main campus situated on Capitol Hill in Monrovia the Polytechnic boasts of four colleges with over 2,500 enrollment. The almost 570 graduates of the 13th commencement convocation exercises in 2012 was the largest number of graduates ever produced by the polytechnic since its establishment. Moreover, it was delightful to see that there were a number of honored students as compared to previous classes. Stella Maris Polytechnic formerly Don Bosco Polytechnic received its Charter from the National Legislature on August 15, 1988. With the change of name to Stella Maris Polytechnic a Bill to amend the charter was presented to the National Transitional Legislative Assembly and was approved on November 18, 2005. Stella Maris Polytechnic is recognized by the Ministry of Education, Republic of Liberia, the Commission on Higher Education in the Republic of Liberia as a founding member of the Association of Liberian Universities (ALU) it accepts the credits, (within guidelines: please see Transfer students) from Cuttington University, United Methodist University, African Methodist Episcopal University, African Methodist Episcopal Zion University, and the University of Liberia. Due to the over-crowdedness of the undergraduate school, the University of Liberia presently accepts only students from Stella Maris Polytechnic for the graduate school. Internationally, Stella Maris Polytechnic (Mother Patern College of Health Sciences) partners with AGEH/Cologne Germany, Kyper College, Calvin College, Columbia University, USA and the Phillipan Women Universities System, Manila, and has been recognized by the Wolrd Education Service (WES) based in the United States of America. The Nursing Program is accredited by Liberia Board of Nursing and Mid-wifery Stella Maris Polytechnic (Bishop John Collins Teacher College partners with Ministry of Education in the production of textbooks and in the training of teachers.

Continued on next page

Table 1.30 – continued from previous page

Provider	For-profit	Local	Website/mission/vision/values
YMCA	No	Yes	`www.umovementliberia.org` Founded in 2011, Youth Movement for Collective Action is a registered non-governmental, non-political and non-for-profit organization, which strongly believed that a good education is the key to breaking the cycle of poverty and can create a better future for children and their families. That is why we promote girl's education by facilitating sponsorship for students who otherwise would not have a chance to attend. Our programs complement the effort of the Ministry of Education and the Sustainable Development Goals (SDGs) — 3: Good health and well-being for people 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all and 5: Achieve gender equality and empower all women and girls with equal access to education, health care, decent work, and representation in political and economic decision-making processes which fuel sustainable economies and benefit societies and humanity at large. The Core Programs of UMOVEMENT are Education, Health Promotion and Disease Prevention, Democracy and Livelihoods. Mission: Our mission is to increase positive change among children and young people through standardized and sustainable youth hands-on programs/projects for social change. Goal: Our goal is to ensure that children and youth 'empowered' as agents of change to secure a better future for themselves and their communities. Program Objectives: To advocate for children and women rights to be prompted, protected and respected by all persons. To improve the living standard of community youth through basic livelihoods activities in order to be self-employed and autonomy for a better living environment. To facilitate access to Sexual Reproductive Health (SRH) services, including sanitation and hygiene services for children and youth at the local community levels. To promote vocational training and credit mechanisms for entrepreneurship and employment for adolescents. To undertake community project for the creation of a healthy and safe community. To advance health, education and development for children and youth through sports and play activities. To increase the capacity of children and youth to prevent, reduce and cope with conflict and promote peace. To organize democratic initiatives (civic education, voter education, etc.) to enhance peace promotion across Liberia. To provide psychosocial support for less fortunate children and widows To foster peace and unity among young people and organize academic excellence programs among young people in various communities. Vision: Youth Movement for Collective Action envisages advance health, education and development for children and youth through collective action -Community engagement.

Note: Providers Mission/vision/values were taken from their website on Jan 28th, 2019. Youth Movement for Collective Action began the evaluation as "Liberian Youth Network". The group has since changed its name.

ACKNOWLEDGEMENTS

Chapter 1, in full, is a reprint of the material as it appears in the American Economic Review 2020. Romero, Mauricio; Sandefur, Justin; Sandholtz, Wayne Aaron. I acknowledge with immense gratitude my coauthors Mauricio Romero and Justin Sandefur, whose support and encouragement was clutch.

Chapter 2

Do voters reward service delivery? Experimental evidence from Liberia

Do voters provide incentives for politicians to focus on public service delivery rather than patronage? This paper examines the electoral effects of a randomized Liberian school reform, previously shown to have caused increased teacher attendance and student test scores. In the subsequent election, this seemingly successful policy caused a 3pp (10%) *reduction* in average vote share for the responsible party's presidential candidate. This may reflect disruptions to the patronage network: treated teachers were less likely to support the incumbent party, staff polling booths, or campaign for candidates. However, heterogeneity in the policy's effectiveness reveals that voters did perceive and reward school quality improvements. In places where the treatment caused large improvements (reductions) in test scores, it also caused large electoral gains (losses). Electoral effects appear only for presidential, not legislative, candidates – suggesting that sophisticated voters attributed credit or blame at the correct level of government. Survey experiments among both candidates and voters confirm the picture of an electorate well informed about the policy's effectiveness and provenance. This paper highlights the political risks of moving from an electoral strategy based on patronage to one based on public service delivery – but it also suggests that increased policy effectiveness has the potential to counteract the opposition of entrenched interests.

2.1 Introduction

In democracies, public officials may seek votes through patronage or public service delivery. Political agency models typically operate on the assumption that while patronage can prevent improvements to the quality of service delivery, voters would reward such improvements if given the chance. Testing this assumption is vital for understanding democratic accountability. But it is a difficult assumption to test: elections are blunt instruments of preference expression, exogenous shocks to public services are rare, and even measuring impacts on service delivery quality is hard.[1]

There are a number of ways voter accountability for service delivery may be diluted or crowded out by the patronage norms common to many developing democracies (Vicente & Wantchekon, 2009; Cruz, Keefer, Labonne, & Trebbi, 2018). Relatively low levels of education may imply a less-well informed electorate more susceptible to capture by special interests (Grossman & Helpman, 1996). Policymakers themselves may not know which policies improve service quality (Hjort, Moreira, Rao, & Santini, 2019). And even seemingly effective policies can be prevented from scale-up by civil servants in patronage jobs (Banerjee, Duflo, & Glennerster, 2008; Bold, Kimenyi, Mwabu, Ng, & Sandefur, 2018). Understanding the extent to which voters provide direct electoral incentives for public service delivery is necessary for characterizing the set of politically feasible policies.

This paper leverages a unique experimental setting to test how voters react to changes in public service quality, overcoming many of the obstacles that have hampered previous work. In 2016, the government of Liberia implemented a randomized school reform which improved school quality on average. Under the Partnership Schools for Liberia (PSL) program, the government outsourced the management of 93 public primary schools to one of several private school operators in a public private partnership. Crucially, the policy was evaluated via randomized control trial (RCT): treated schools were chosen randomly from within pairs of eligible schools matched on

[1] Cash transfers have been shown to be politically effective, but they represent a fundamentally different form of government spending than service provision. Also, a small but insightful quasi-experimental literature documents electoral effects of public goods, but these largely focus on visible public spending such as construction.

pre-program characteristics.[2] This randomization, unusual for a government policy, gives the researcher a great degree of confidence in the policy's effect on the quality of public services: the program improved average student test scores by .18 standard deviations, and teacher attendance by 50%, after one school year (Romero, Sandefur, & Sandholtz, 2019). The randomization also provides rare exogeneity in *which areas* received these improvements to service quality, allowing a comparison of voters' reactions near treatment versus control schools. Attribution of credit was relatively straightforward: education policy is set the executive branch, and Liberia has no elected local politicians. The program's funding came from external donors and was earmarked specifically for the program; this allows a direct test of the electoral effect of this policy, unconfounded by voters' preferences over any possible counterfactual use of the funds. Finally, the policy was implemented in an election year and garnered significant press attention, making it an unusually salient education reform – the evaluation's results were announced in a press conference in September 2017, one month before the October general election. PSL therefore provides a rare exogenous shock to school quality, and furthermore, the unusual matched-pair randomization provides a valid counterfactual all along the distribution of shocks to school quality.

This paper shows that the policy caused a 3-percentage-point (10%) *reduction* in vote share for the responsible party's presidential candidate. This may have been partly due to popular anger over the actions of the highest-profile private school operator, which dismissed students from schools with large class sizes and secured the dismissal of a large fraction of its existing teachers. These moves made headlines and provoked heated opposition; teacher unions inside and outside the country came out strongly against the program. Teachers are often key political brokers (Pierskalla & Sacks, 2019; Larreguy, Montiel Olea, & Querubin, 2017), and this paper shows that the program appears to have disrupted traditional forms of patronage-based electioneering. It caused a reduction in teachers' support for the incumbent government – especially among unionized teachers. It also caused a decline in participation in political activities: staffing of polling stations and campaigning

[2]The schools in the study were not a representative sample of all Liberia's schools. But they did constitute 3.4% of public primary schools, and 8.6% of public primary and early-childhood education students in the country. The sample included schools from 13 of the 15 counties in Liberia.

for candidates both went down by nearly a third among teachers in treated schools.

However, the negative average effect does not imply that voters are insensible to the quality of service delivery. Heterogeneity in the policy's effectiveness reveals that voters were attuned to school quality, and rewarded or punished the responsible party in proportion. The matched-pair randomization design of the program means that the local average treatment effect measured for each individual pair of schools is unbiased. While comparing these LATEs to each other is not experimental, it is illustrative. The pair-level treatment effect on incumbent party vote share follows the gradient of the pair-level treatment effect on test scores. In other words, the treatment increased or decreased the incumbent party's vote share in proportion to how much it increased or decreased test scores (see Figure 2.5). Where test scores improved more than about 0.5σ (around the 80th percentile in the school distribution), the policy caused significant gains for the incumbent party's candidate. Where test scores worsened by more than about 0.3 σ (around the 20th percentile), it caused significant losses. This suggests that voters were able to perceive changes to school quality and reward or punish them. It is notable that these electoral effects, both average and heterogeneous, appear only at the presidential level, suggesting voters attribute credit or blame to the appropriate political actor.

Survey experiments among both candidates and voters confirm the picture of a well-informed electorate. Two information experiments, conceived and implemented prior to the election, were designed to test whether information frictions inhibit accountability for public services. The first randomly provided evidence on the program's effectiveness and popularity to a pool of over 600 legislative candidates. The second randomly provided candidates' policy positions to the households of children from PSL study schools. However, baseline knowledge among both groups was so high as to leave little room for improvement, and the extra information does not seem to have shifted priors.

This paper highlights the political risks entailed in moving from an electoral strategy based on patronage to one based on public service delivery. Overall, a policy which improved test scores had a negative level effect on the electoral fortunes of the responsible candidate. The data suggests a loss of patronage support as a likely explanation – teachers became less likely to engage in electioneering. However, the paper also suggests that increased policy effectiveness has the potential to counteract the opposition of entrenched interests. In places where the policy worked well, it created positive electoral effects which completely counteracted the negative level effect. Insofar as continued research can reduce uncertainty on how to improve public services, it may empower politicians who seek to bypass patronage-driven interests and appeal directly to voters who value service quality.

This paper contributes to the literature on electoral rewards for public services by focusing on the intensive margin of service quality. Handing out money, whether clientelistically or programmatically, is a proven vote-winner (Wantchekon, 2003; Vicente & Wantchekon, 2009; Manacorda, Miguel, & Vigorito, 2011; De La O, 2013; Golden & Min, 2013). This direct redistribution is distinct from investment in public goods and services, about which creative studies have also been written (Harding, 2015; Litschig & Morrison, 2013; Marx, 2018; Samuels, 2002; Zimmermann, 2018). But this literature mostly focuses on visible inputs such as roads and school-building.[3] The current paper is unusual in being able to measure voters' reactions not to school inputs, but school quality. The most closely related paper is Dias and Ferraz (2019), which examines electoral returns to information about test scores in Brazil.

Another strand of literature this study adds to is that on the political economy of adoption and scale-up of reforms. From the theoretical literature, Majumdar and Mukand (2004) posit that policy experimentation sends a *negative* signal about politician ability to voters, which could lead to risk aversion in policy provision. This is consonant with the empirical findings of the current paper. A related paper is Bursztyn (2016), which shows that not all voters prefer public service spending to direct redistribution.

Finally, this paper contributes to the literature on voter information, by showing that even poorly-educated voters inform themselves about electorally consequential policies in sophisticated ways. A certain tension exists in the current literature between studies like Ferraz and Finan (2008),

[3]Closely related is the credit-claiming literature demonstrating that opportunistic politicians can sometimes receive unearned benefits from public good provision (Cruz & Schneider, 2017; Guiteras & Mobarak, 2015).

which shows that voters respond to information dispersed through known sources, and (Dunning et al., 2019), which finds no effect of information through new sources created by researchers.[4] The present paper reconciles these findings somewhat by showing how difficult it is for a researcher to shock voters' priors about policies they already care enough about to change their votes over.[5]

The rest of this article is structured as follows: Section 2.2 provides context about Liberia and the policy; Section 2.3 outlines the empirical strategy; Section 2.4 presents the main results; Section 2.5 describes the information experiments; and Section 2.6 concludes.

2.2 Context

Education often ranks among governments' central public service responsibilities. However, in many developing countries, the public school system fails to provide the level of education deemed basic by international institutions. Liberia's moribund public education system exemplifies this failure. The civil wars of 1999-2003 and the Ebola epidemic of 2014 left the Ministry of Education with little capacity to run a national school system. An effort to clean thousands of ghost teachers from Ministry payrolls was cut short (New York Times, 2016), and while systematic data is scarce, teacher absenteeism appears common (Mulkeen, 2009). Nearly two-thirds of primary aged children are not in school, including over 80% of children in the poorest quintile, placing Liberia in the lowest percentile of net enrollment rates in the world, and at the 7th percentile in youth (15-24) literacy (EPDC, 2014). Demographic and Health Surveys show that among adult women who did not go to secondary school, only six percent can read a complete sentence. In 2013, after all 25,000 high school graduates who sat the University of Liberia's college entrance exam failed, President Ellen Johnson Sirleaf said the education system was "a mess."

[4]Some examples exist of experimenter-provided information affecting electoral outcomes (De Figueiredo, Hidalgo, & Kasahara, 2011).

[5]However, this paper contrasts somewhat with studies which have found plenty of room to improve politicians' informedness about their constituents' preferences (Butler & Nickerson, 2011; Liaqat et al., 2018; Hjort et al., 2019).

2.2.1 The policy

In response, the Liberian Ministry of Education announced a pilot program – "Partnership Schools for Liberia" (PSL) – which contracts the management of 93 government primary schools to one of eight private school operators in a public-private partnership. The government (and donors) provide these operators with funding on a per-pupil level. The operators were given responsibility for (though not ownership of) the resources the government normally uses to provide education – schools, classrooms, materials, and teachers, and a grant equal to USD$50 per pupil on top of that. In exchange, operators are responsible for the daily management of the schools, and can be held accountable for results. The operators include high profile, for-profit chains with investors like Bill Gates, Mark Zuckerberg, the World Bank, and DFID. Other operators are non-profit NGOs based in Liberia and abroad, and one operator is a respected Liberian religious institution.

2.2.2 The evaluation

Based on criteria established by the evaluation team, the Ministry of Education, and operators, 185 PSL-eligible schools across 12 of Liberia's 15 counties were identified. These schools are not a representative sample of public schools in the country – they have better facilities, internet access, and road access than the average school in the country. The eligible schools were split into pairs matched on administrative data, and treatment was assigned randomly within matched pairs.

Romero et al. (2019) showed that the program increased test scores by around 60%, teacher attendance by 50%, and satisfaction of both students and parents by about 10%. The evaluation also showed that at least some of the critiques of program detractors were well-founded: one operator chose to enforce class size limits, forcing hundreds of students to leave their regular school (though nearly all enrolled elsewhere). The same operator also requested reassignment of 74% of the teachers in the schools it operated. The independent evaluation team's report of these results was released in a press conference about one month before nationwide elections for the presidency and the House of Representatives, which happened in October 2017.

A central finding of the evaluation was the wide heterogeneity in treatment effects on school quality. Because randomization was carried out within pairs of schools matched on pre-treatment characteristics, it is possible to recover an unbiased (if noisy) effect of treatment on test scores for each school pair. Those treatment effects are plotted in Figure 2.1, demonstrating the degree of variation in treatment effects experienced at the local level:

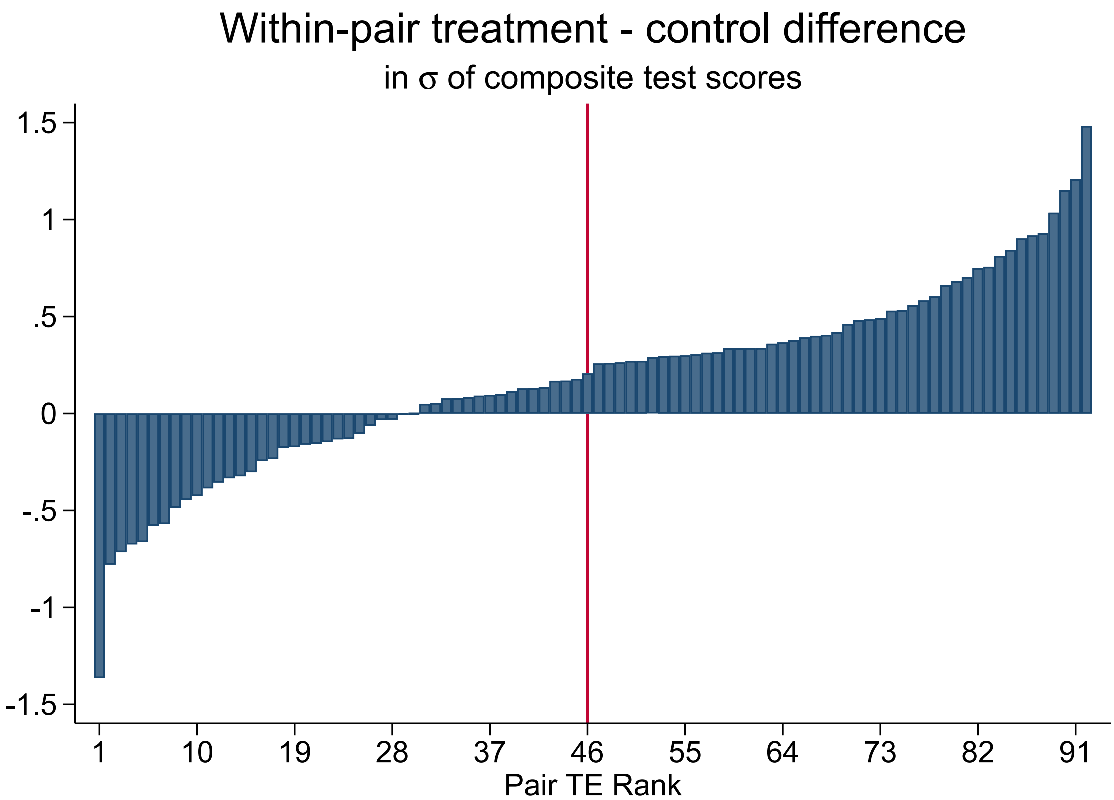


Figure 2.1: Heterogeneity in treatment effect on test scores

2.2.3 The political context

The program enjoyed a relatively high profile, garnering attention from local and international news outlets, and receiving condemnation from a UN Special Rapporteur, who wrote that Liberia was abrogating its responsibilities under international law. The National Teacher Association of Liberia staged a strike, calling for the resignation of the Minister of Education. In response, students blocked the main highway to the country's international airport, demanding that the government and the teachers' union send the teachers back to class.

The policy was championed by the incumbent president, Ellen Johnson Sirleaf of the Unity Party (UP). At the time of the policy's implementation, Sirleaf was nearing the end of her second (and constitutionally-mandated final) term as president. Her vice president, Joseph N. Boakai, was the UP's presidential candidate in the 2017 election. Reforming the education sector was a priority for the administration. Both the president and the vice president championed PSL and went out of their way to associate themselves with the program during its first year. (President Sirleaf attended a Flag Day celebration at one of the private operator's schools, and Vice President Boakai spoke at the graduation ceremony for the same operator's teacher training course.) However, as the election neared, the politically powerful National Teachers Association of Liberia (NTAL), which opposes the program, became more vocal in their opposition. Neither Boakai nor any of the other presidential candidates gave the policy a strong role in their campaign.

A few factors made it difficult to predict ex ante how voters would react to this school policy. On the one hand, the program brought more resources into communities and improved average test scores, which parents might have perceived. On the other hand, it could alienate voters who see the privatization of education provision as a dereliction of the government's duty. Voters may be inclined to reward politicians for acknowledging the inadequacy of public provision and offering other options. Or they may punish politicians for outsourcing a fundamental responsibility to private entities with limited accountability, many foreign and some for-profit.

2.3 Design

2.3.1 Timeline

This paper's main results leverage the randomization of the PSL program and use administrative voting data from the October 2017 general election as outcomes. The paper also presents the effects of the PSL program on households' political attitudes, using data collected at the end of the first school year in which the policy was in place: May/June 2017. Finally, it presents results from two information experiments carried out in the weeks leading up to the election. Figure 2.2 diagrams which data collection efforts happened at what time, and how they informed each other.

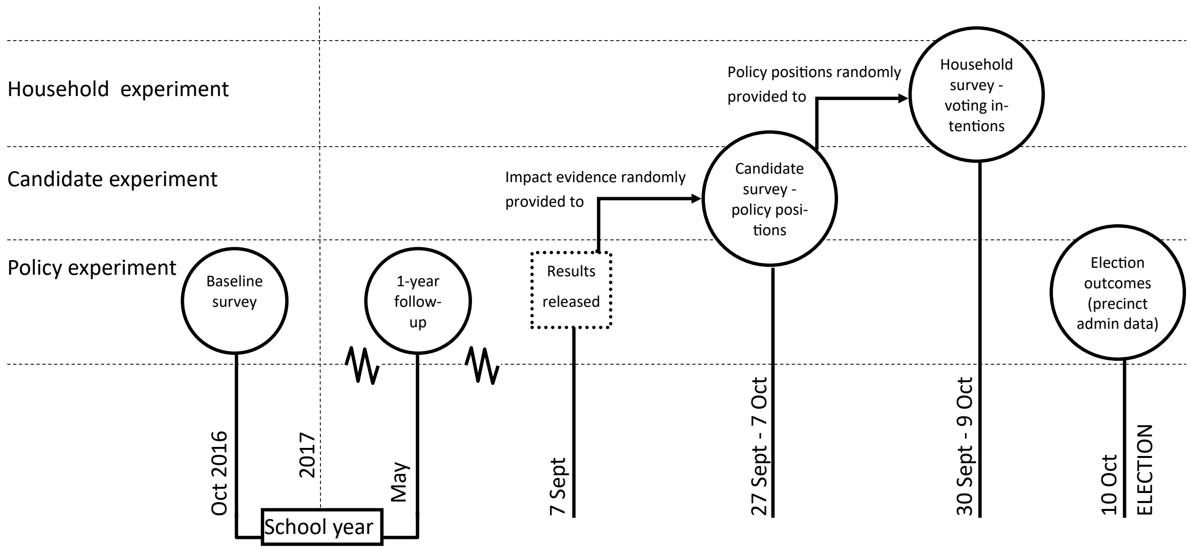


Figure 2.2: Timeline
Timeline not to scale. Circles indicate data collection efforts, and the height of the circle indicates the relevant experiment.

2.3.2 Data

There were 2,080 polling booths in Liberia in the 2017 election, with 637 votes cast in the median booth. Electoral data at the voting booth level, as well as booth GPS coordinates, were obtained from the National Elections Commission (NEC) of Liberia (http://www.necliberia.org).

Studying the effect of PSL on voting outcomes necessitates deciding how to map the unit of treatment assignment (schools) onto the unit of outcome measurement (polling booths). One possible scheme would assign all booths the treatment status of their nearest school. However, this raises the risk of contamination: consider a booth which is infinitesimally closer to a control school than a treatment school. The treatment school may be expected to exert at least as much influence on voters' choices as the control school, yet it would be classified as control.

Instead, this paper considers a certain radius around each booth, defining the booth's "fraction treated" as the number of treated schools divided by the number of total RCT schools within the radius.[6] This in turn leaves the question of how wide to draw the radius. Under the assumption that the strength of a school's impact on voters' choices decreases as some function of distance from the school, this implies a trade-off between the sample size and effect size; booths beyond the radius will be discarded.

The main specifications in this paper use a radius of 10km, for three reasons. 1) This radius is wide enough to embrace at least one polling booth for all 185 schools in the RCT (though in some cases only one). 2) A clear majority (58%) of booths lie within 10km of an RCT school, and the density of polling places drops off precipitously after this threshold (see Figure 2.3). 3) 97 % of students in the RCT live within 10km of their school. Figure shows a map of Liberia depicting the 185 schools from the RCT and the 1200 booths within 10km of at least one of them.

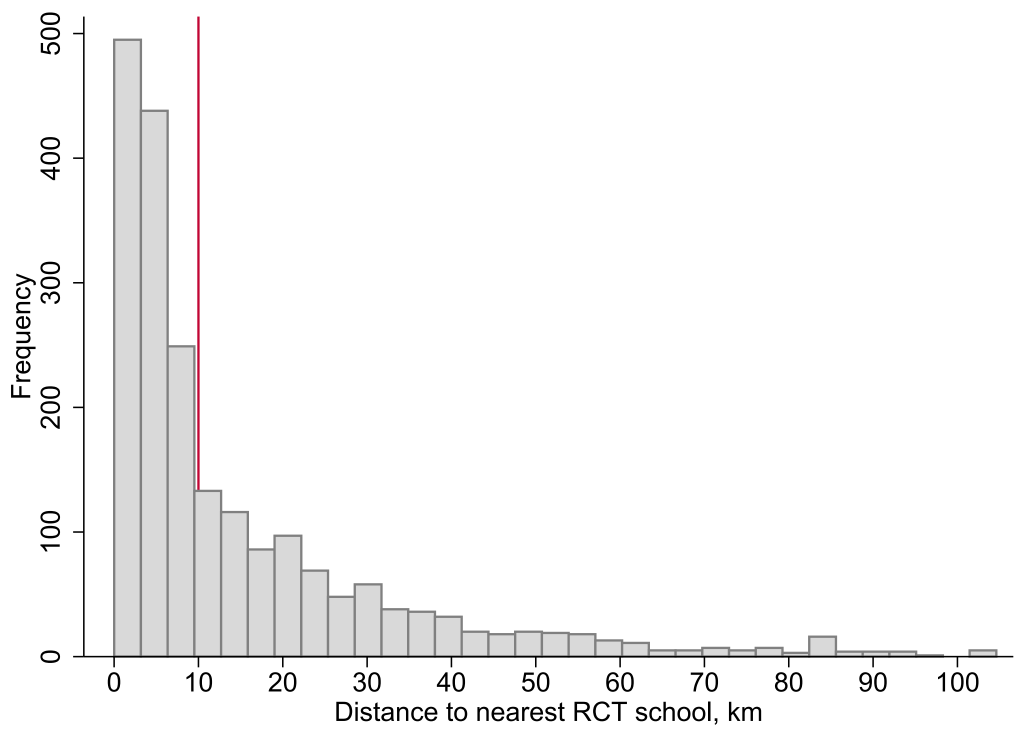


Figure 2.3: Histogram of polling booths by distance to nearest RCT school
Universe of 2080 polling booths from 2017 general election.

[6]Following Romero et al. (2019), this paper considers the original ITT treatment assignment of the schools, despite the fact that there was some noncompliance in which schools actually came under private administration.

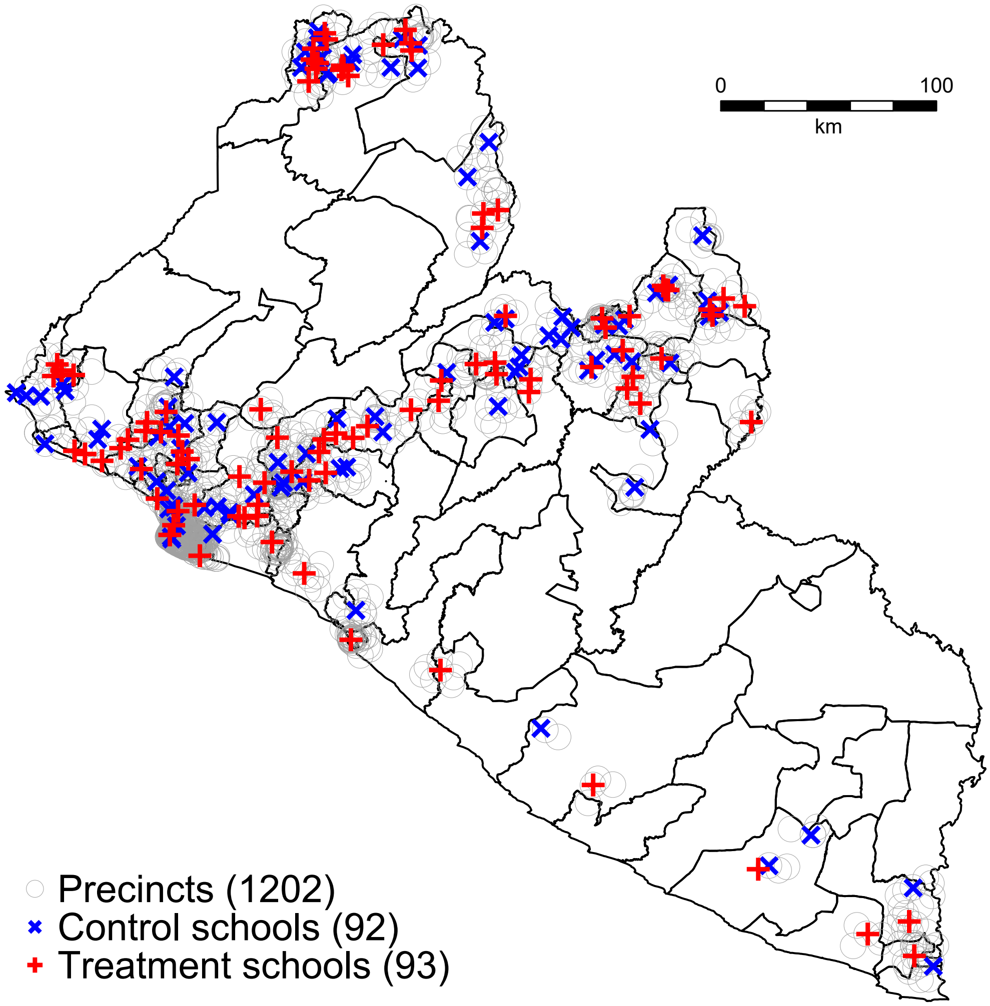


Figure 2.4: Map of treatment and control schools, and polling booths within 10km

2.3.3 Empirical Specification

The main results in this paper come from the following specification:

$$Y_{isp} \quad = \quad \alpha_p + \beta \mathit{FracTreated}_i + \gamma X_i \varepsilon_{isp} \tag{2.1}$$

Y_{isp} represents electoral outcomes for polling booth i near school s in pair p. α_p are matched-pair fixed effects (stratification dummies based on the pair p corresponding to the booth's nearest school s). $\mathit{FracTreated}_i$ is defined as the number of treated schools with 10km of booth i over the total number of RCT schools within the same radius. X_i are booth-level controls, consisting of the election outcome at that booth from the previous election in 2011. Standard errors are clustered at the level of the electoral district (J = 63).

2.3.4 Balance

Table 2.1 checks that this specification provides a set of polling places which are balanced on election outcomes from 2011 (the last nationwide election before the treatment), regressing treatment status on 2011 election measures as "outcomes." While not rising to the level of statistical significance at even the 10% level here, the point estimate on the difference in ruling party presidential vote share is somewhat large, and in some alternate specifications this difference shows up as significant. Therefore, all specifications include controls for 2011 election outcomes.

Table 2.1: Balance: pre-treatment outcomes (2011 election)

	Registered voters / pop.	Votes cast / pop.	Turnout (rep.)	Share invalid (rep.)	Ruling party (pres. 1st rd.)	Ruling party (pres. runoff)	Ruling party (rep.)	Share for 2017 incumbent
Fraction RCT schools treated	0.010	0.006	-0.008	-0.004	-0.023	-0.005	0.017	-0.019
	(0.011)	(0.008)	(0.008)	(0.005)	(0.017)	(0.010)	(0.028)	(0.051)
N	1200	1200	1200	1200	1200	1200	1200	1200
Mean (control)	0.124	0.091	0.724	0.073	0.429	0.842	0.179	0.337

Standard errors clustered by electoral district. School matched-pair fixed effects included for booth's nearest school. Regressions include precincts from the 2011 election located within 10km of any RCT school, with treatment of the precinct defined as fraction of RCT schools assigned to the PSL treatment. 134 precincts which are within 10km of a RCT school were newly created between 2011 and 2017 and hence have missing values for 2011 election variables. Missing values have been replaced with zero, and indicator varables for whether values are missing have been included in all regressions.
* p<0.10, ** p<0.05, *** p<0.01

2.4 Results

2.4.1 Main electoral results

The paper's main outcome of interest is how the school reform affected electoral outcomes in the October 2017 general election. Table 2.2 displays them.

Table 2.2: Average school policy effects on vote share

	Ruling party: president (1st round)		Ruling party: president (runoff)		Ruling party: legislative		Incumbent: legislative	
Fraction RCT schools treated	-0.032**	-0.029**	-0.033**	-0.028*	0.003	0.008	0.021	0.015
	(0.015)	(0.012)	(0.016)	(0.015)	(0.021)	(0.020)	(0.033)	(0.032)
N	1200	1200	1200	1200	1200	1200	1200	1200
Mean (control)	0.293	0.293	0.382	0.382	0.128	0.128	0.201	0.201
Controls		✓		✓		✓		✓

Standard errors clustered by electoral district. Matched-pair fixed effects corresponding to precinct's nearest school included. Regressions include precincts from the 2017 election located within 10km of any RCT school, with treatment of the precinct defined as fraction of RCT schools assigned to the PSL treatment. Missing values have been replaced with zero, and indicator variables for whether values are missing have been included in all regressions. Controls: number of registered voters in 2011; total votes cast in the first-round presidential election in 2011; 2011 UP presidential 1st-round vote share.
* p<0.10, ** p<0.05, *** p<0.01

The school policy reduced average vote share for the presidential candidate from the incumbent Unity Party, in both the first round and the runoff election held a month later. However, it appears to have had no effect on the electoral outcomes for federal legislative candidates, the only other politicians on the ballot. This is consistent with a picture of an electorate well enough informed to know that the policy was the product of the executive branch, and the legislative branch's ability to affect is was minimal.

2.4.2 Heterogeneity in election outcomes by policy impact

Because randomization happened within matched pairs, each pair can be considered a mini-experiment, and local *learning* treatment effects can be defined at the level of the pair as the difference between average test scores at the treatment and control school. The following table interacts the treatment variable with a dummy for whether the polling booth's nearest school is part of a treatment pair in which the treatment effect on test scores was above the median. That is, the coefficient on this interaction term represents the additive effect on electoral outcomes of treatment in places where test scores improved a lot. Table 2.3 tests whether treatment assignment had a differential effect on electoral outcomes in pairs of different local learning treatment effects.

Table 2.3: Effects on 2017 vote share, interacted with learning treatment effect

	Ruling party: president (1st round)		Ruling party: president (runoff)		Ruling party: legislative		Incumbent: legislative	
Fraction RCT schools treated	-0.056***	-0.047***	-0.059***	-0.048***	-0.015	-0.008	0.048	0.045
	(0.018)	(0.014)	(0.017)	(0.016)	(0.022)	(0.021)	(0.036)	(0.035)
Fraction RCT schools treated × TE > p50	0.077***	0.060***	0.086***	0.068**	0.060	0.053	-0.089	-0.098
	(0.021)	(0.020)	(0.025)	(0.028)	(0.052)	(0.050)	(0.073)	(0.073)
N	1200	1200	1200	1200	1200	1200	1200	1200
Mean (control)	0.293	0.293	0.382	0.382	0.128	0.128	0.201	0.201
Controls		✓		✓		✓		✓

Standard errors clustered by electoral district. Nearest school matched-pair fixed effects included. Regressions include precincts from the 2017 election located within 10km of any RCT school, with treatment of the precinct defined as fraction of RCT schools assigned to the PSL treatment. Missing values have been replaced with zero, and indicator variables for whether values are missing have been included in all regressions. Controls: polling-place level electoral outcomes from previous general election in 2011 (number of registered voters; total votes cast; presidential vote share for ruling Unity Party candidate).
* p<0.10, ** p<0.05, *** p<0.01

The treatment's effect on electoral outcomes was a function of how well the program worked at boosting school quality. Negative electoral effects were driven by places where the program caused small increases (or reductions) in test scores. Where the program caused big test score increases, treatment more than counteracted the negative average effect.

As with the main result, the policy only affected voters' choices in the presidential race. Legislative candidates had no influence on the policy, but they might have been expected to attempt to claim credit in their districts – especially incumbents (Cruz & Schneider, 2017; Guiteras & Mobarak, 2015). Instead, voters' choices in the legislative race were largely unaffected.

Figure 2.5 depicts a similar analysis non-parametrically, plotting mean electoral treatment effects (and 95% bootstrapped confidence intervals) along the distribution of local learning treatment effects.

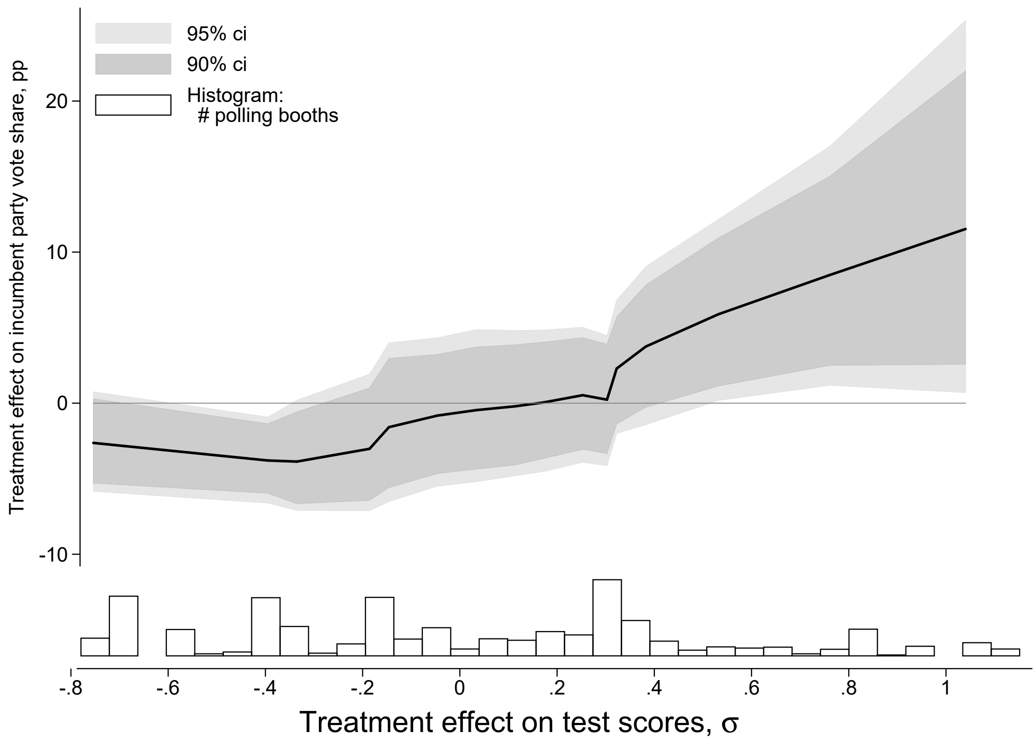


Figure 2.5: Effect of PSL on responsible party's presidential vote share, by treatment effect on learning

This figure plots the lowess-smoothed coefficients of the fraction of schools treated on UP presidential candidate's vote share (controlling for its 2011 pre-treatment value) in bins corresponding to 20iles of matched-pair-specific treatment effects on test scores. Constructing the bootstrapped confidence intervals consisted in calculating the same estimates from 1000 resamples of the original data, keeping the 2.5th, 5th, 95th, and 97.5th percentile of the distribution of the estimates from this resampling procedure. The histogram shows data winsorized at 1%/99%.

It may be surprising that voters are able to perceive and reward something like learning gains, which are difficult for PhD researchers to measure. That said, people living near these schools may observe important school quality variables which are invisible to researchers. In any case, some caution is in order in interpreting these results – learning treatment effects are not assigned randomly.[7] It's possible that learning treatment effects correlate with other, more easily observable dimensions of heterogeneity which are what voters really care about.

To check this, Table 2.4 reports results from similar analyses to Table 2.3, but with other observable dimensions of school quality. All columns in this table look at the same outcome variable:

[7]For another recent example of useful nonrandom variation combined with an experiment, see Balboni, Bandiera, Burgess, Ghatak, and Heil (2020).

vote share for the ruling party's presidential candidate. All include controls for polling-place-level electoral outcomes from the previous presidential election in 2011. Each column looks separately at a dimension of school quality which voters might plausibly observe – and which was improved by the reform.[8] Column 1 interacts treatment with a dummy for whether the voting booth's nearest school is part of a pair exhibiting above-median treatment effects on teacher attendance. Column 2's interaction is with treatment effects on student attendance rates. Column 3 interacts treatment with a dummy for whether the treatment school underwent any construction or major repairs and the control school didn't.[9] Finally, Column 4 includes both the construction interaction and the learning interaction from Table 2.3.

[8]Romero et al. (2019) found that the reform significantly improved both student and teacher attendance, and furthermore, that teacher attendance was one of the best predictors of learning gains. The reform also caused an improvement in construction, though this was not reported in Romero et al. (2019).

[9]This variable took a value of 1 if the principal reported *either* new construction *or* major repairs in the foregoing year to any of the following: classrooms, office/staff rooms, store rooms, toilets/latrines, staff housing, library, playground, water source. 62% of treated schools, and 48% of control schools, underwent any construction or repairs by this measure. "Treatment effect" at the pair level here simply means the difference in this indicator between the treatment and control schools in a pair. The distribution of pair-level treatment effects is 28% positive, 60% zero, and 12% negative. So while I have expressed the dummy here as "above the median treatment effect," this ends up being simply a dummy for a positive treatment effect. NB also that this variable is self-reported by principals. Alternatively, using an enumerator's measure of observed classroom quality in this analysis yields a similar and significant result.

Table 2.4: Effects on 2017 vote share, interacted with treatment effect on various dimensions of school quality

	Ruling party: president (1st round)			
	(1)	(2)	(3)	(4)
Fraction RCT schools treated	-0.029*	-0.029	-0.041***	-0.059***
	(0.014)	(0.021)	(0.013)	(0.014)
Fraction RCT schools treated × TE>p50: teacher attendance	-0.001			
	(0.028)			
Fraction RCT schools treated × TE>p50: student attendance		-0.001		
		(0.031)		
Fraction RCT schools treated × TE>p50: new construction or repairs			0.093**	0.138**
			(0.036)	(0.055)
Fraction RCT schools treated × TE>p50: learning				0.067***
				(0.020)
Fraction RCT schools treated × TE>p50: learning × TE>p50: new construction or repairs				-0.116
				(0.075)
N	1200	1200	1200	1200
Mean (control)	0.293	0.293	0.293	0.293
Controls	✓	✓	✓	✓

Standard errors clustered by electoral district. Nearest school matched-pair fixed effects included. Regressions include precincts from the 2017 election located within 10km of any RCT school, with treatment of the precinct defined as fraction of RCT schools assigned to the PSL treatment. Missing values have been replaced with zero, and indicator variables for whether values are missing have been included in all regressions. Controls: polling-place level electoral outcomes from previous general election in 2011 (number of registered voters; total votes cast; presidential vote share for ruling Unity Party candidate).
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Although the reform has previously been shown to have caused gains in both teacher and student attendance overall, big treatment effects in these variables did not predict treatment effects in relevant electoral outcomes. Electoral gains did appear wherever school construction happened in the pair's treatment school but not in its control school. Figure 2.6 shows the symmetry of this effect: treatment caused significant electoral gains in places where treatment schools had construction and control schools didn't; it caused very large electoral losses where control schools had construction and treatment schools didn't; and it had (a reasonably precise) zero electoral effect where there was no difference in construction between treatment and control schools.

Does this mean voters are simply observing treatment effects on construction, which happen to correlate with treatment effects on learning? No. At the pair level (N=92), the correlation between learning TE and and construction TE is only 0.16. Column 4 of Table 2.4 interacts treatment with dummies for high treatment effects in both learning and construction, showing that both interactions have predictive power. This suggests that voters observe and reward learning gains in some way independent of their perceptions and rewards of construction gains.

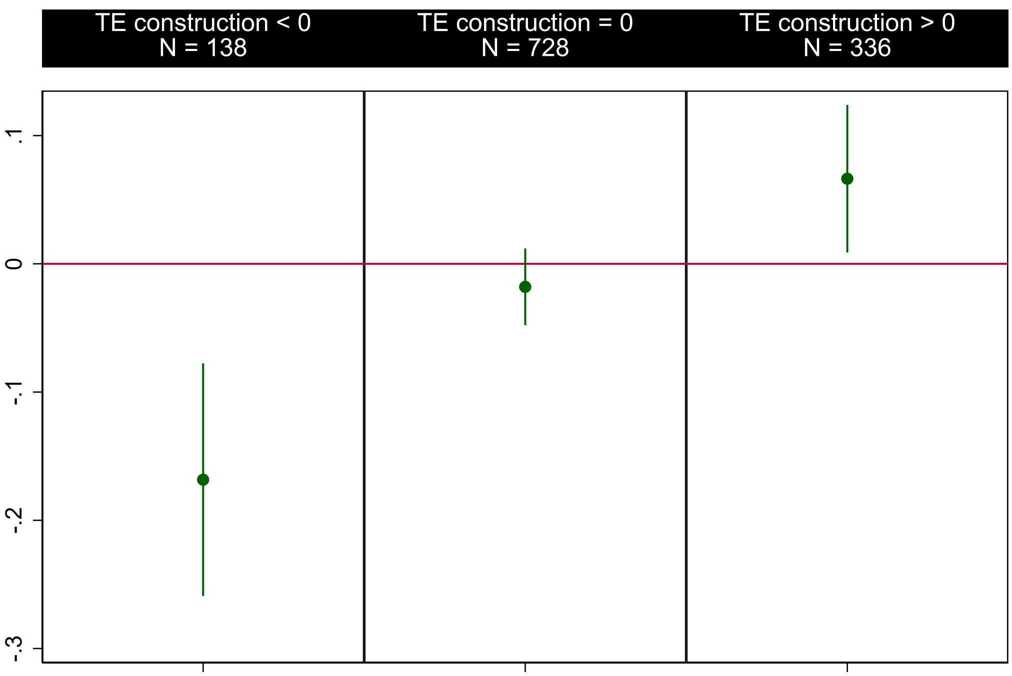


Figure 2.6: Effect of PSL on responsible party's presidential vote share, by treatment effect on construction and repairs

This figure plots the coefficient and 95% confidence intervals of the treatment effect of PSL on voting-booth-level vote share for the incumbent Unity Party's presidential candidate in the 2017 election, for three different groups of polling booths. The leftmost column shows the coefficient for the booths whose nearest school is part of a matched pair where the control school underwent any new construction or major repairs and the treatment school did not. The middle column shows the coefficient for booths whose nearest school was part of a pair where both treatment and control schools had the same construction and repair status. The rightmost column shows the coefficient for booths whose nearest school was part of a pair where the treatment school experienced new construction or repairs while the control school did not.

2.4.3 Effects on household attitudes

A policy as high-profile as PSL offers a lot of information about a government, even to voters who are not directly affected because their kids attend the pilot schools. In order to shed light on whether electoral results are driven by those directly affected, this section presents the effects of the policy on household members of students from the treatment and control school. These attitudes were measured in May/June 2017, about five months before the election. (The RCT measures of the program's popularity which were presented to candidates as part of the candidate information experiment in Section 2.5 came from this survey.)

Table 2.5: Effect of PSL on household and teacher survey outcomes (May 2017)

	Treatment	Control	Difference	Difference (F.E.)
Household midline survey (N = 1271)				
Considers childs school a gov school	0.917	0.937	-0.020	-0.011
	(0.276)	(0.244)	(0.020)	(0.015)
Satisfied w/ childs edu	0.743	0.689	0.054*	0.069***
	(0.437)	(0.463)	(0.028)	(0.020)
Gov performance on edu is good	0.566	0.549	0.017	0.034*
	(0.496)	(0.498)	(0.032)	(0.020)
Schools top priority for gov spending	0.811	0.739	0.072***	0.068***
	(0.392)	(0.440)	(0.026)	(0.020)
Liberia is moving forward	0.577	0.507	0.070**	0.071***
	(0.494)	(0.500)	(0.032)	(0.019)
Satisfied with president	0.651	0.632	0.019	0.020
	(0.477)	(0.483)	(0.032)	(0.023)
Satisfied with legislator	0.545	0.535	0.009	0.004
	(0.498)	(0.499)	(0.036)	(0.024)
Plans to vote for UP	0.178	0.198	-0.020	-0.021
	(0.383)	(0.399)	(0.028)	(0.017)

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Standard errors are clustered at the school level. The sample is the original (intent-to-treat) treatment and control allocation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5 shows that treated households became more satisfied with their children's education,[10] more impressed with the government's performance on schools, and more likely to say schools were their top priority for government spending. Although the PSL treatment consisted in outsourcing school management to private school operators, nearly all households in both treated and control schools still saw their children's schools as government schools. Treated households were more likely to feel sanguine about Liberia's future. However, they were no more likely to be satisfied with the performance of the president (whose administration created the policy), or the legislator representing them in Congress. Finally, they were no more likely to report planning to vote for the ruling Unity Party (UP) which had created the policy (although very few people were willing to divulge voting intentions before the election).

The fact that households directly affected by the policy were no more likely to report differential satisfaction with the current administration, or intentions to vote for its continuation, suggests that the aggregate electoral results were not driven by the households of students in the schools themselves.

2.4.4 Effects on teachers' political participation

In many parts of the world, public sector teaching jobs function as patronage, with the expectation that those in them will help turn out people to vote for the politicians who provided the job.(Larreguy et al., 2017; Pierskalla & Sacks, 2019) If PSL succeeded in professionalizing the teaching force, it might have caused teachers to engage less in political activities. On the other hand, given reports in the press that the national teachers' union opposed PSL, it might have been the case that the program galvanized opposition and caused more teachers to organize and participate in the political process. Indeed, Romero et al. (2019) reported that teachers in treated schools were significantly more likely to be dismissed.

A May/June 2017 teacher survey, a few months before the election, asked teachers about some of their political attitudes and voting intentions. The effects of treatment on these attitudes is summarized in Table 2.6.

[10]This particular result from this was previous reported in (Romero et al., 2019).

Table 2.6: Effect of PSL on teachers' political participation

	Willing to state voting intention	Satisfied w/ incumbent pres.		Intends to vote for ruling party pres. candidate	
Treatment	0.022	-0.043	0.018	-0.077*	-0.007
	(0.035)	(0.031)	(0.039)	(0.044)	(0.053)
Union member			0.188***		0.156**
			(0.056)		(0.071)
Treatment × Union member			-0.208***		-0.250***
			(0.080)		(0.095)
N	948	782	782	473	473
Mean (control)	0.524	0.771	0.771	0.642	0.642

Standard errors clustered by school. School matched-pair fixed effects included. Outcomes from a May/June 2017 survey. Columns 3 and 5 interact treatment with an (endogenous) dummy for union membership.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

During the election season, not all teachers were willing to share political opinions, but this does not appear to have differed by school treatment status (column 1). These smaller sample sizes make estimates on teacher attitudes a bit imprecise, but treatment appears to have reduced both satisfaction with the incumbent government and intentions to vote for its party in the next election (columns 2 and 4). Columns 3 and 5 interact treatment with a dummy for the teacher's union membership. Treatment made unionized teachers *much* less likely to express satisfaction with the incumbent government or to want to vote for its candidate.

A follow-up survey of teachers at PSL treatment and control schools in June-July 2019 asked teachers about their political activities during the 2017 election, including whether they had staffed registration booths and/or polling stations, encouraged participation in general, and campaigned for a particular party or candidate. Table 2.7 shows the impact of the policy on teachers' reported political activities.

Table 2.7: Effect of PSL on teachers' political participation

	Support ruling party	Registration booths	Polling booths	Encourage participation	Campaign for a party or candidate	Involved in any
Treatment	-0.067	-0.035***	-0.054**	-0.022	-0.044**	-0.103***
	(0.043)	(0.012)	(0.023)	(0.020)	(0.022)	(0.034)
N	385	847	847	847	847	847
Mean (control)	0.732	0.059	0.174	0.149	0.152	0.396

Standard errors clustered by school. School matched-pair fixed effects included. Outcome in column 1 is whether the teacher said in a May/June 2017 survey that they intended to vote for presidential candidate from the outgoing administration in the upcoming Oct 2017 presidential election; sample limited to those willing to state a voting intention for any candidate. Other outcomes come from a June/July 2019 follow-up survey asking teachers to recall their political activities from the election.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

In the absence of treatment, sizable minorities of teachers were politically involved – 40% reported participating in at least one of the political activities asked about. However, treatment appears to have disrupted traditional patronage activities: teachers were much less likely to report being politically involved in the election.

2.5 Information experiments

The administrative voting results provide evidence that education policy was an important election issue for many voters. But accountability also requires that politicians have enough awareness of voters' preferences to respond to them. Politicians may not know how much voters care about service delivery, or they may not know which policies are likely to work to improve service quality. Politicians may not know how much voters care about service delivery, or they may not know how to hold the bureaucracy accountable to actually improve service quality. The next two experiments were designed to test whether politicians were constrained by a lack of information about either policy effectiveness or voters' preferences – and whether voters were constrained by a lack of information on politicians' policy positions.[11]

[11]The experiments described here were pre-registered along with pre-analysis plans at `https://www.socialscienceregistry.org/trials/1501` (policy impact on political attitudes) and `https://www`

2.5.1 Candidate experiment

The candidate information experiment was conducted through a phone survey in which survey enumerators attempted to call all 992 candidates running for seats in the House of Representatives.[12] The sample consists of the 681 candidates reached (69%). These candidates received fewer votes on average than non-participating candidates, but they were not uniformly inconsequential: the sample includes 112 "veteran" candidates who ran for Congress in the previous election of 2011; 25 of the 73 incumbent House incumbents (of whom 22 were standing for reelection); and 32 of the 73 eventual winners. 13% of the sample ended up as the winner or the runner-up in their district. Figure 2.7 plots the density of vote shares for candidates who did and did not participate in the survey.

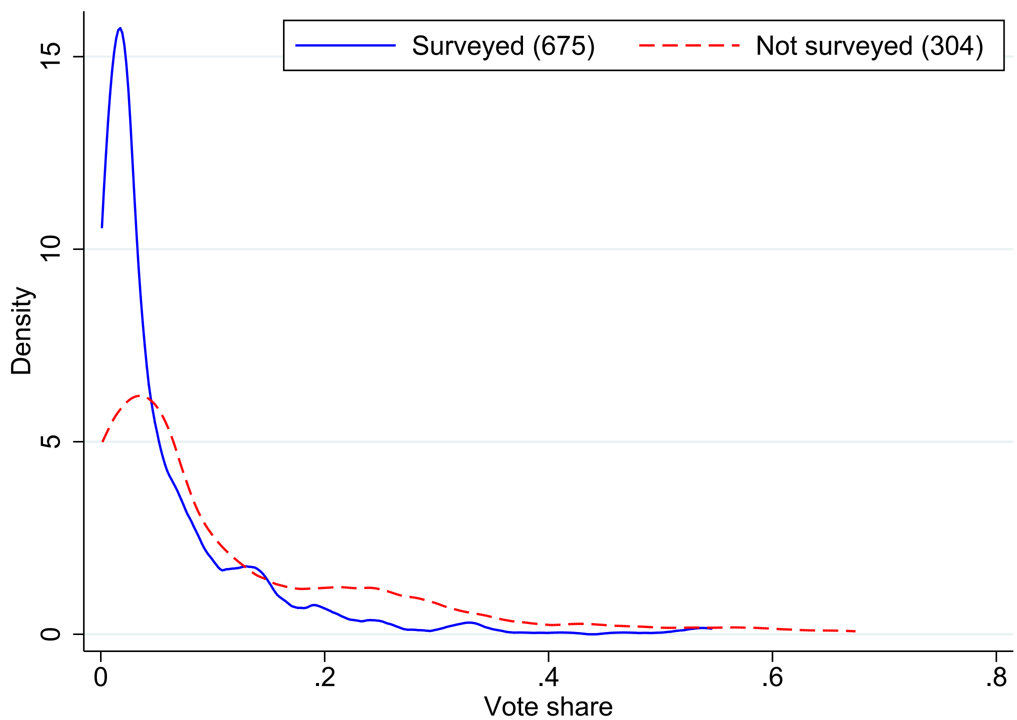


Figure 2.7: Distribution of vote shares for surveyed and unsurveyed candidates
Note: Figure excludes about 20 surveyed candidates from two districts for which election results were still pending as of the time of writing

.socialscienceregistry.org/trials/2506 (information experiments for candidates and households).

[12]A randomized controlled trial registry entry and the pre-analysis plan for both the candidate and household experiments are available at: https://www.socialscienceregistry.org/trials/2506.

The treatment consisted of random provision of the RCT evidence of the program's treatment effects on a) learning outcomes and/or b) the the program's popularity among affected households.[13] Conditional on being reached by phone a candidate was randomized into one of four treatment arms:

1. "Control:" basic description of the school policy, and one sentence about what supporters and opponents of the policy said about it;
2. "Impact information:" control language plus a brief summary of the findings of the independent evaluation, including positive effects on test scores and teacher attendance, as well as student and teacher dismissals;
3. "Popularity information:" control language plus a brief summary of effects on political attitudes (those seen in this paper in Table 2.5);
4. "Both:" control condition, impact information, and popularity information.

The exact text of these information interventions is in Appendix 2.7.

Candidate balance and summary statistics

The balance check on candidates' characteristics and pre-treatment survey responses is in Table 2.8. For simplicity of comparison, it pools all information treatments into a single "any information" treatment.

[13] Learning outcome RCT evidence took the form of a very concise synthesis of the main results from Romero et al. (2019). Popularity RCT evidence took the form of the results on household attitudes presented in Subsection 2.4.3, Table 2.5 in this paper.

Table 2.8: Balance – Candidate experiment

Variable	Control	Any info	Difference
Incumbent	0.024	0.041	0.017
	(0.154)	(0.198)	(0.017)
Eventual winner or runner-up	0.128	0.094	-0.034
	(0.335)	(0.292)	(0.027)
UP (incumbent pres. party)	0.036	0.050	0.014
	(0.187)	(0.219)	(0.019)
CDC (main opposition)	0.072	0.060	-0.012
	(0.260)	(0.238)	(0.022)
Number of attempts necessary to interview	2.554	2.417	-0.137
	(1.949)	(1.813)	(0.165)
Has own children in primary	0.601	0.596	-0.005
	(0.491)	(0.491)	(0.044)
Candidate has Univ. degree	0.717	0.726	0.009
	(0.452)	(0.446)	(0.040)
It's good for gov't to work w/ private companies to provide edu.	0.931	0.905	-0.026
	(0.254)	(0.293)	(0.026)
Heard of PSL	0.596	0.642	0.046
	(0.492)	(0.480)	(0.043)
Heard of any PSL operator	0.892	0.882	-0.010
	(0.312)	(0.323)	(0.029)
'Strongly' or 'Somewhat' approve of teachers' union	0.981	0.978	-0.003
	(0.136)	(0.146)	(0.013)
Believes voters hold exec. branch responsible for education	0.842	0.858	0.015
	(0.365)	(0.350)	(0.032)
Believes voters hold exec. branch responsible for PSL	0.722	0.721	-0.001
	(0.450)	(0.449)	(0.044)
Believes more than 'a few' voters have heard of PSL	0.193	0.175	-0.018
	(0.396)	(0.380)	(0.034)
Observations	166	515	681

Candidates were randomized to receive information about PSL's popularity, its effectiveness, or both. For simplicity, this table compares the group who received no information with the pooled group of those who received any information, but comparisons among all interactions are available upon request.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

It is rare to survey such a large body of politicians; even the descriptive statistics are illuminating, and consistent with politicians who are reasonably well-informed about their constituents.

Most candidates had heard of the school policy's name ("PSL"), and nearly all had heard of at least one of the operators associated with it. Nearly all supported the idea of public-private partnerships in education, but nearly all also approved of the teachers' union (which officially opposes the PSL program), perhaps evincing ideological flexibility. Although over 80% of candidates themselves agreed that the government should work with private providers of education, and that the PSL program had increased learning, only 57% of candidates said they thought voters supported the program on average – qualitatively consistent with the divergent electoral outcomes measured from the administrative voting data. Candidates estimates' of their voters' support of the program correlated with the average treatment effect of the program within their constituency. 72% said their voters (correctly) credited the executive branch with responsibility for the program (compared to only 12% who thought their voters credited the legislative branch with the program).

The main outcome of interest was the candidate's position on the PSL policy. In order to elicit "public" policy positions that went beyond cheap talk, survey enumerators offered to communicate the candidate's PSL policy position of choice to voters as part of a sensitization campaign (see next section). Candidates were asked to select the statement that most closely aligned with their view:

- The PSL program should be expanded and paid for with tax revenues;
- The PSL program should be tested before any significant expansion;
- The PSL program should be immediately discontinued;
- No position.

52% of candidates interviewed said the program should be expanded; 30% said it should be tested further first; and 7% said it should be discontinued. The other 11% gave no position.

Candidate results

Table 2.9 shows the results of the experiment; however, at least as important as the treamtent estimates are the summary means at the bottom of the table. An overwhelming majority of candidates already believed the policy was successful at improving test scores, and a large majority also thought their constituents supported the policy (which, while contradicted by the eventual election outcome, was consistent with the household survey data provided in the information experiment). Hardly any candidate opposed the idea that governments should work with private companies to provide education. The independent evaluation of the PSL program had become public a few weeks earlier, and was reported in local press; candidates may have read about the results already.

Table 2.9: Policy information's effect on average candidate survey outcomes

	Expand PSL	Fund PSL w/ taxes	Gov should work w/ pvt. edu.	Too much foreign control	PSL $\Rightarrow\uparrow$ learning	Voters support PSL
Popularity info	0.075	0.008	0.013	-0.038	-0.079**	-0.077
	(0.056)	(0.052)	(0.035)	(0.055)	(0.035)	(0.054)
Impact info	0.050	0.034	0.013	-0.048	-0.012	0.003
	(0.055)	(0.050)	(0.034)	(0.053)	(0.029)	(0.052)
Popularity info × Impact info	-0.063	-0.014	-0.014	0.065	0.079*	0.086
	(0.077)	(0.071)	(0.047)	(0.075)	(0.046)	(0.074)
N	681	681	681	681	681	681
DV Mean	0.476	0.650	0.883	0.578	0.921	0.614
Controls	N	N	N	N	N	N
District FE	N	N	N	N	N	N

Robust standard errors in parentheses. 'Expand PSL' means the candidate asked us to tell their voters they support expanding the school policy. Other columns are candidate survey outcomes.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Given that candidates' priors were reasonably in tune with the information provided, it is perhaps unsurprising that the policy information failed to shift them much; neither the learning impact information nor the popularity information had a significant effect on the average candidates' answers to key survey questions. The first column shows the main outcome – whether candidates would go on record as supporting the policy. Candidates who received information about the school policy were, on average, no more likely to tell voters they supported expanding it. Nor did the information change their opinions on the (non-public) survey outcomes in the other columns: whether the program should be funded with taxpayer money, whether the government should work with the private sector in education, or whether foreign NGOs have too much control in Liberia. Perhaps most strikingly, neither the impact information nor the popularity information had any effect on candidates' likelihood of thinking that voters support the program. (If anything, popularity information seems to have decreased candidates' belief that the program boosted test scores.)[14]

2.5.2 Household experiment

A separate information experiment sought to measure whether households' approval ratings and voting intentions responded to information about candidates' positions on the PSL program. Conducted in October 2017, this survey re-contacted by phone a subset of the households who had been previously interviewed (in person) as part of the 1-year evaluation of the school policy (those interviewed in Table 2.5). 489 households participated, of the 833 for whom at least one unique phone number was available (59%). While this subset of households likely differs from the May 2017 sample (households with phones are likely to be wealthier), Table 2.10 shows within-sample balance in terms of the randomization into the information treatment.

All participants in this household follow-up survey – information treatment and control – received a brief summary of the PSL program's impacts on test scores. Treatment consisted in informing the household about presidential and legislative *candidates'* PSL policy positions. Legislative candidates' positions came from the candidate survey; only the positions of candidates in the household's legislative district were provided. The median (interquartile range) household in the information experiment received information about the positions of 4 (2,6) legislative candidates, representing 30 (24,46)% of the legislative candidates on the ballot in their district. Presidential candidates' positions came from a presidential debate held a few weeks before the election, which included a question about PSL. Only three candidates participated, none of whom would go on to make the runoff election. One (Cummings) had a broadly supportive position about PSL; the other two (Cooper and Jones) were more skeptical.[15] The text of the treatment conditions can be found in

[14]This nonresult stands somewhat in contrast to Hjort et al. (2019)'s recent experimental results showing that Brazilian mayors demand rigorous policy evidence. However, in the Brazilian context, the mayors who were the experimental subjects of the study do in fact have direct jurisdiction and autonomy over the policies about which evidence was given. The route to impact on education policy for federal legislators in Liberia is more roundabout.

[15]This 26 September debate seems to have been the only presidential debate in which moderators asked about the PSL program. The three mentioned candidates were the only ones who attended the debate; as mentioned, none of them were front-runners. The eventual vote shares received by Cummings, Jones, and Cooper in the general election were

Appendix 2.7.

Household balance and summary statistics

Table 2.10 shows balance of the October 2017 household sample on pre-information-treatment characteristics.

7.2%, 0.8%, and 0.7% respectively.

Table 2.10: Balance – Household experiment (N = 494)

Variable	Control	Candidate Info	Difference	N
In PSL treatment group	0.543	0.514	-0.028	494
	(0.499)	(0.501)	(0.045)	
Heard of PSL	0.644	0.615	-0.028	494
	(0.480)	(0.487)	(0.044)	
Heard of any operator	0.862	0.842	-0.020	494
	(0.345)	(0.365)	(0.032)	
Legislature created PSL	0.008	0.008	-0.000	494
	(0.090)	(0.090)	(0.008)	
Correctly identifies responsible party for PSL	0.441	0.417	-0.024	494
	(0.498)	(0.494)	(0.045)	
It's good for gov to work w/ private companies to provide sch	0.962	0.942	-0.021	480
	(0.190)	(0.235)	(0.020)	
PSL should be expanded and funded through the national budget.	0.803	0.782	-0.021	478
	(0.398)	(0.413)	(0.037)	
Children learn more in PSL schools	0.827	0.850	0.022	446
	(0.379)	(0.358)	(0.035)	
Knows current Representative's name	0.822	0.866	0.045	494
	(0.383)	(0.341)	(0.033)	
Satisfied with Representative	0.342	0.308	-0.034	477
	(0.475)	(0.463)	(0.043)	
Related to a rep. candidate	0.170	0.184	0.014	492
	(0.376)	(0.388)	(0.034)	
Related to member of teachers' union	0.333	0.277	-0.056	485
	(0.472)	(0.448)	(0.042)	
Attended any campaign event	0.358	0.329	-0.029	489
	(0.480)	(0.471)	(0.043)	
Any candidate has talked about PSL	0.146	0.160	0.014	444
	(0.354)	(0.367)	(0.034)	
Fraction of district's candidates who provided info to candidate survey	0.354	0.317	-0.037**	467
	(0.165)	(0.148)	(0.015)	

Notes
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The assignment of the Candidate Information treatment is mostly balanced in terms of the original PSL treatment. The one measurable difference between the Candidate Information treatment and the control group is that households randomly assigned to receive candidate information also happened to be in districts for which policy position information was available for slightly fewer legislative candidates (35% vs 32% of the total candidates on the ballot).

As with the candidate experiment, summary statistics here are highliy informative. Households in the sample are reasonably well-informed. Almost all respondents to the household survey could correctly name their current Representative, but only 34% are satisfied with him or her. 36% have attended at least one campaign event for a Representative candidate, and 17% are related to someone running for office. A third are related to a member of the teachers' union, but almost everyone supports the idea of public-private partnerships in education, and 4 in 5 think children learn more in PSL schools and that the policy should be expanded. Nearly everyone (correctly) does *not* credit the legislature with creating the policy. 44% gave an accurate answer about who was responsible for it (either the executive branch, private school companies, or foreign NGOs). Most of the rest said "don't know." 15% said they had heard at least one candidate mention PSL.

Household results

Table 2.11 presents the effects of information about candidates' positions on approval ratings and voting intentions for presidential candidates.

Table 2.11: Candidate information's effect on households' approval and voting intentions of presidential candidates

	Pro-PSL candidate (Cummings)		Anti-PSL candidate 1 (Cooper)		Anti-PSL candidate 2 (Jones)	
	Approve	Vote	Approve	Vote	Approve	Vote
Candidate info	0.056	0.030	-0.056*	0.004	-0.108***	-0.011
	(0.037)	(0.027)	(0.029)	(0.004)	(0.034)	(0.013)
N	494	494	494	494	494	494
DV Mean	0.309	0.147	0.167	0.000	0.276	0.043
District FE	N	N	N	N	N	N

Robust standard errors in parentheses clustered at the school. Sample consists of a subset of households originally contacted as part of PSL midline evaluation, reached by phone for this follow-up survey about one week before the election on 10 October 2017. Treatment consists of informing households of candidates' positions regarding PSL.

* p<0.10, ** p<0.05, *** p<0.01

These results show that the information had little impact on voters' views regarding these presidential candidates. Columns 1 and 2 correspond to Alexander Cummings, whose debate statement about PSL was broadly favorable. Columns 3 and 4 correspond to MacDella Cooper, whose statement on PSL was skeptical. Columns 5 and 6 correspond to Mills Jones, whose position was stridently opposed to PSL. Candidate information reduced approval rates for candidates who opposed PSL in the presidential debate, and the effect sizes are sizable compared to the mean approval. This might be considered a manipulation check, as this is a context where experimenter demand effects are operative. Passing this manipulation check, then, shows that households did indeed listen to the survey and take it seriously. That gives the non-effect of information on voting intentions more weight. Voters appear not to have been swayed in their voting intentions by the information provided.[16]

Table 2.12 shows the effects of information on candidates' policy positions on electoral outcomes for *legislative* candidates.

Table 2.12: Candidate information's effect on household voting intentions: Representative

	Vote Rep. PSL Yes	Vote Rep. PSL No Info	Vote Rep. UP	Vote Rep. Incumbent	Vote Any Rep.
Candidate info	0.001	0.041	0.026	0.003	-0.053
	(0.024)	(0.033)	(0.023)	(0.026)	(0.045)
N	494	494	494	494	494
DV Mean	0.138	0.677	0.147	0.273	0.632
Controls	N	N	N	N	N
FE	N	N	N	N	N

Robust standard errors in parentheses clustered at the school. Sample consists of a subset of households originally contacted as part of PSL midline evaluation, reached by phone for this follow-up survey about one week before the election on 10 October 2017.

* p<0.10, ** p<0.05, *** p<0.01

Candidate information has no discernible effect on voting patterns for representatives. Households receiving the information were no more likely to plan to vote for representatives who

[16]To be fair, baseline voting intentions were already very low for these candidates, leaving little room for downward movement anyway. Sample nonresponse to these questions also likely attenuates any potential result here; only about 2/3 of respondents chose to express a voting preference for either presidential or legislative candidates.

supported (or opposed) the expansion of the policy, nor for candidates about whom they had received no information. They were no more likely to vote for representatives from the ruling party or for incumbents. They were also no more or less likely to express a voting preference at all.

While somewhat imprecise, these null results are consistent with the administrative voting data results in Section2.4. Those results showed that the PSL program affected voters' choices for presidential candidates, but not for legislative candidates. This household survey shows that households *knew* legislators were not responsible for the program. They may have seen information on legislative candidates' PSL policy positions as immaterial. The large electoral results from before also imply that people were willing to invest enough research to form their own opinions without the help of information provided by researchers. This highlights a paradox of voter information interventions: people already have incentives to learn about what they want to know, so if they don't know it already it's because they don't care about it. This makes it difficult to shock their information set.

2.6 Conclusion

Understanding whether and how voters create electoral incentives for politicians to improve public services is a fundamental question of political economy. Its urgency may be even greater in Liberia, a post-conflict country and one of the poorest countries in the world, where democracy is young and literacy is around 50%. This paper shows that Liberian voters are sophisticated consumers of public services; both politicians and voters are well-informed about the effects of a controversial school reform. Electoral rewards for the policy were contingent on its effectiveness: voters rewarded the responsible politician where the policy worked well, and punished him where it worked poorly. Overall, however, this worked out to a negative average electoral effect for the responsible politician, likely at least in part due to the opposition of organized groups who perceived the reform as a threat to the status quo.

This highlights the risks of policy experimentation (Majumdar & Mukand, 2004). Policymakers who seek to improve public service delivery often face the unenviable task of shaking up entrenched systems with motivated supporters. They often lack credible evidence on how a given intervention is likely to work in their context (Pritchett & Sandefur, 2014). Meanwhile, they can be confident that any change will provoke opposition from those who benefit under the status quo (Acemoglu, 2010). In these circumstances, clientelism or vote-buying may provide a less risky path to electoral victory (Wantchekon, 2003; Cruz et al., 2018). It is just possible, however, that better research can reduce policymakers' uncertainty on forecasts of policy effectiveness and thereby improve the odds that voter gratitude for a given policy's improved services outweighs opposition from those who stand to lose from change. Further research, of course, is necessary to help answer the question of whether further research can ease the transition from patronage- to public service-based politics.

2.7 Appendix: Information experiments detail

2.7.1 Text of candidate information treatments

CONTROL CONDITION:

> *In the Partnership Schools program, 93 government primary schools became Partnership Schools, managed by one of eight private and NGO school providers. [Sentence describing which providers operated in the candidate's county] Teachers in Partnership Schools remain on government payroll, and buildings remain the property of the government and free to students. These schools also received extra resources from foreign donors: 50 US per student.*
>
> *Supporters of Partnership Schools believe that private management can bring innovation and improvement to Liberia's schools. Opponents of Partnership Schools believe that the resources would be better spent within the public system, without private contractors.*

IMPACT INFORMATION CONDITION: [Control condition language, plus:]

> *The Ministry of Education commissioned an independent scientific evaluation of Partnership Schools using state-of-the-art methodology. The study was carried out by academics at institutions based in the United States: The Center for Global Development and the University of California.*
>
> *The evaluation showed how the outcomes for students and teachers were different in Partnership Schools. The children in Partnership Schools learned 60% more math and English than children in the traditional public schools. That means that children in a Partnership School learned more in 6 months than children in a traditional public school learned in a whole school year. The evaluation also found that teachers in Partnership Schools were twice as likely to attend school.*[17] *The evaluation also identified some problems: in some schools run by Bridge International Academies, some students were kicked out and had to transfer to different schools, and over half of the teachers were removed. The program is also expensive: the partnership schools cost at least twice as much to run as government schools, and in some cases much more.*

POPULARITY INFORMATION CONDITION: [Control condition language, plus:]

> *The Ministry of Education commissioned an independent scientific evaluation of Partnership Schools using state-of-the-art methodology. The study was carried out by academics at institutions based in the United States: The Center for Global Development and the University of California.*

[17] This was an inadvertent error. Teachers were in fact 50% more likely to attend, not twice as likely.

The researchers interviewed voters whose children went to Partnership Schools and traditional government schools, as well as teachers in these schools.

They found that voters whose children went to Partnership Schools were: 10% MORE satisfied with their children's education, 7% MORE likely to say the government's performance on education was good, 11% MORE likely to say education is their top priority for government, and 14% MORE likely to say Liberia is moving forward .

Teachers in Partnership Schools were: 21% LESS likely to be satisfied with the teachers' union (NTAL), and 7% MORE likely to say Liberia is moving forward.

The fourth condition contained the control language, the impact information, and the popularity information.

2.7.2 Text of household information treatment

The control condition consisted of this brief mention of the three presidential candidates who took part in a debate:

Thank you. We are near the end of the survey. Now I just want to give you some information about the candidates.

Liberia's last presidential debate was on September 26th. The three candidates who attended the debate were: MacDella Cooper from LRP, Alexander Cummings from ANC, and Mills Jones from MOVEE.

The treatment condition included that prelude as well as the candidate's words regarding the school policy from that debate:

In that debate, each candidate made a statement about Partnership Schools or PSL. I'm going to read you a part of each candidate's statement. Please listen:

MacDella Cooper said: "It's a test project. Maybe at the end of the test, we'll see . . . Putting the Liberian public school in the hands of a private organization, I don't see the benefit yet."

Alexander Cummings said "We should also be open to different solutions. And we can't be fixated on only one traditional way of doing things. We got to be creative. We got to be bold."

Mills Jones said: "We are not going to do it. It suggests to me that we have given up on our own capacity to solve our problems and so we must look outside for help. We're not going to do that."

The treatment condition also included a list of the representative candidates who had participated in the candidate survey, who had asked survey enumerators to let their voters know their position on the school policy:

Some of the candidates for Representative in YOUR district also have made statements about PSL, which they wanted us to share with you. Please listen carefully:

These candidates say PSL should be taken into more schools, and supported by the national budget: [names]

These candidates say PSL needs to be tested more before making a decision: [names]

These candidates say PSL should be stopped immediately, and normal government schools should get that support: [names]

Chapter 3

Infrastructure, connectivity, and elections

with Daniel Leff Yaffe and Alejandro Nakab

Do voters provide incentives for public good provision by rewarding it at the ballot box? Much of the political theory which rationalizes democracy depends upon the answer to this question, but empirical evidence is sparse, partly because exogenous variation is rare. In this paper, we measure the effect of construction of the US Interstate Highway System (IHS) on county-level vote shares for the party of incumbent governors. To recover exogenous variation in county-year-level highway construction, we a) include county and year fixed effects and b) construct a shift-share instrument which predicts actual county-year-level construction by interacting state-level yearly congressional apportionments with county shares according to the original 1947 federal plan. We find that one extra highway mile causes a 0.6-pp increase in vote share for the incumbent governor's party. We also demonstrate that when adjacent counties are connected to the interstate, they see a reduction in the difference between their levels of public good provision from both local and state authorities. This is suggestive of yardstick competition as an incentive for public good provision. We propose a new theory of reference-dependent retrospective voting, in which voters take into consideration not only the level but the distribution of local public goods between counties.

3.1 Introduction

Elections are the mechanism by which democracies translate popular will into practical governance. According to the theory of retrospective voting, voters reward politicians who perform well and punish those who perform poorly (Key, 1966). Some scholars have conjectured that this feature of democracies leads them to "invest more in broad-based public goods" (Acemoglu, Naidu, Restrepo, & Robinson, 2019; Acemoglu, 2008).

But the empirical evidence that voters reward public good provision is limited. Persson and Tabellini (2002), in their seminal work *Political Economics: Explaining economic policy*, note that in this realm of inquiry "the bridge linking theory with data is way too fragile," an assessment which despite some progress remains broadly true two decades later. One reason for the dearth of empirical evidence is the difficulty of credible identification. Governments rarely assign the provision of public goods randomly. In fact, there is good evidence that they target spending to places and times they think will be electorally advantageous, although scholars disagree about the nature of this targeting. [1] This means that not only is public good provision endogenous to electoral outcomes – we don't even know the direction of the bias.

This paper identifies the electoral effects of a canonical infrastructure program: the USA's Interstate Highway System (IHS). Specifically, we measure the effect of new IHS construction on county-level vote share for incumbent state governors from 1950-1972.[2] We focus our analysis on transportation infrastructure because its visibility makes it a natural place to look for political rewards(Mani & Mukand, 2007). Transportation infrastructure also produces the kind of positive externalities that mean government provision is necessary to reach a socially optimum allocation. It

[1]See Rogoff (1990); Rogoff and Sibert (1988); Nadeau and Blais (1992); Peltzman (1992); Katsimi and Sarantides (2012); Potrafke (2010) for evidence of political budget cycles in the US, and Jones, Meloni, and Tommasi (2012) for an example from Argentina. Not all these studies sign the correlation between the electoral cycle and government spending the same way. Meanwhile, political scientists debate when and why politicians target "core" vs. "swing" voters (Dixit & Londregan, 1996; Stokes, 2005).

[2]We consider the US to be a democracy for the purposes of this analysis. For what it's worth, the US scored a "perfect" 10 Polity score during most of the period we consider (along with the likes of Western Europe, Costa Rica, Japan and Israel), dipping to 8 for the stretch from 1967-1973 (in the company of Gambia, India, Turkey, and Uruguay). For a contemporary comparison, the US regained a score of 10 in 1974 and maintained it until dipping again to 8 for the years 2016-2018. http://www.systemicpeace.org/inscrdata.html

has been shown to increase productivity and growth (Donaldson, 2018; Donaldson & Hornbeck, 2016; Fernald, 1999). Transportation infrastructure projects account for almost 20% of World Bank lending (World Bank, 2007). In the US, highway spending alone constitutes 28% of gross government investment (Leff-Yaffe, 2020). Transportation infrastructure can be characterized as one of the archetypal public good investments in which democracies are hypothesized to invest more (Acemoglu et al., 2019).

Identifying this effect necessitates solving the problem of selection. We show that highway construction is concentrated in the two years preceding elections, consistent with models of the political business business cycle.[3] Politicians may manipulate construction of the IHS both temporally and spatially, choosing to build in the most electorally crucial counties at the most electoral crucial times. To address this, we employ two complementary identification strategies. First we use county and year fixed effects, comparing within counties the vote share of the incumbent governor's party in election years where the county saw high vs. low amounts of IHS construction. We complement this strategy with a Bartik-style shift-share instrumental variable (IV) based on the original 1947 road network plan. We multiply 1) the number of total IHS miles constructed (or apportioned) in a state in a year by 2) the fraction of the state's total allotted miles represented by a given county's total allotted miles (Bartik, 1991; Goldsmith-Pinkham, Sorkin, & Swift, 2020). The instrument predicts actual new IHS miles constructed in a county-year, with a very high F-statistic.

Our IV estimate shows that opening one additional highway mile causes the incumbent party's vote share in that county to increase by 0.6 percentage points. Fixed-effects regressions produce smaller estimates with the same sign, suggesting that incumbent governors target infrastructure construction in counties where they are less popular.

We then turn to the role highways play not only as public goods, but as connections. We establish a new empirical fact: connecting adjacent counties via the IHS prompts a homogenization of other public good provision. We show that county-pair-level differences in state-level revenues,

[3]Nordhaus (1975) and MacRae (1977) pioneered the political business cycle literature by positing that voters punish unemployment. Like us, Bercoff and Meloni (2009); Jones et al. (2012); Katsimi and Sarantides (2012), and Potrafke (2010) all provide more micro-level evidence of mechanisms governments may use to achieve these aims or similar. Peltzman (1992), by contrast, finds that US government expenditure decreases in electoral years.

direct expenditures, and hospital spending all diminish after a pair of adjacent counties is connected via the IHS. This empirical finding relates to existing models of yardstick competition, in which voters evaluate incumbents by making comparisons with neighboring jurisdictions (Besley & Case, 1995). To rationalize this result, we create a model of reference-dependent voter preferences with asymmetric information (building on the framework of Kőszegi and Rabin (2006)). We propose that highway connectivity lowers the cost of moving between counties, leading to an increase of public information, and allowing voters to compare spending in their county to that of neighboring counties. The model assumes that voters' satisfaction with government services is relative to the level of service they perceive in other counties, forcing incumbents to compete with other county governments through public good provision or reduced taxes. This accords with what we observe in the data.

Our paper provides empirical backing for the theory of retrospective voting (Key, 1966; Ferejohn, 1986), which posits that government policy has electoral effects, creating incentives for politicians. The existing empirical work in this vein is somewhat sparse, but Levitt and Snyder (1997) is also based on the US context and provides a useful benchmark. They estimate that an additional $100 federal spending per capita in the two years preceding an election cause a 2-percentage-point increase in vote share for the congressional candidate from the incumbent's party. Other research finds that the relationship between government services and electoral rewards is more nuanced. Sandholtz (2019) shows that Liberian voters only rewarded politicians for a school reform in the places where it was most effective.

We also contribute to the literature on yardstick competition in electoral politics (Salmon, 1987; Bodenstein & Ursprung, 2005; Besley & Case, 1995). While this literature typically considers voters comparing among competing jurisdictions, our paper suggests that geographic differences in service provision can be electorally important even *within* a jurisdiction. We also build on the literature showing that changes to public information have electoral ramifications (Stromberg, 2004; Snyder & Strömberg, 2010; Besley & Burgess, 2002).

Finally, we build on the robust social science literature examining the political and economic effects of highway construction. While our paper is the first to examine the effects of highway construction on electoral outcomes in the US, Voigtllnder and Voth (2018) shows that highway construction can drive[4] political support in an autocracy. Similarly, Harding (2015) finds that improvements to road quality increase incumbents' reelection rates in Ghana. Other work examines other outcomes of highway construction. Baum-Snow (2007) demonstrates that the IHS caused suburbanization and urban depopulation. Clayton Nall builds on this finding to show that highway construction contributed to the political polarization and class stratification of American geography in the 20th century (Nall, 2015, 2018).

The rest of the paper is organized as follows. Section 3.2 provides background information on the IHS. Section 3.3 describes the methodology and data used. Section 3.4 presents the main results on the effect of the IHS on political and other outcomes at the county level. Section 3.5 outlines our empirical findings on public good homogenization at the county pair level. Section 3.6 concludes.

3.2 Background: the Interstate Highway System

Each of the annual issues of the Highway Statistics Series from 1956 to 1996 provide excellent summaries of the IHS. Supplementing this series with the Federal-Aid Highway Acts, as well as with the cost-estimate reports of finishing the IHS, one can obtain detailed information on the funding and year-to-year changes in the IHS plans. In this section I present a summary on the evolution of the IHS.

The Federal-Aid Highway Act of 1944 gave birth to the IHS, back then called the National System of Interstate Highways. The Act called for the designation of a highway system of 40,000 miles to connect metropolitan areas, cities and industrial centers, as well as to connect the U.S. with Canada and Mexico at key border points. In 1947 the selection of the first 37,700 miles was announced. However, at the time there was no plan on how to fund the system, nor an estimate of

[4] if you will

how much it would cost, so its construction was uncertain. A map of the 1947 plan is presented in Figure 3.1.

In 1952, legislation approved some small funding towards what can be called a pilot stage in the program. The Act of 1952 devoted $25 million of federal funds for the fiscal year 1954 and a similar amount for the fiscal year 1955. States were required to match the federal funds with a 50% Federal - 50% State rule. Moreover, the funds were apportioned across states with a formula[5] that that assigned a weight of one-third to each of the following factors:

(1) Relative Population: the ratio which the population of each state bears to the total population of all the states (as shown by the latest available Federal census).

(2) Relative Area: The ratio which the area of each state bears to the total area of all the states.

(3) Relative Rural Delivery and Star Routes (RDSR) Mileage: the ratio which the mileage of rural delivery routes and star routes in each state bears to the total mileage of rural delivery and star routes in all the states at the close of the preceding fiscal year.

Two years later the Act of 1954, which expanded the pilot stage of the interstate program, was approved. It designated an appropriation of $175 million of federal funds for the fiscal year 1956 and a similar amount for the fiscal year 1957. For these years the apportionment formulas for the states were modified to give more weight to the state's population: (1) a weight of $2/3$ on relative population, (2) $1/6$ on relative area, and (3) $1/6$ on relative RDSR. Moreover, the matching funds rule changed to 60% Federal - 40% State.

Shortly after the Act of 1954 was passed, President Eisenhower started a campaign towards expanding the highway program with a speech given to the Governors' Conference.[6] After the speech, President Eisenhower asked General Clay to head a committee to propose a plan for constructing the interstate. At that time there was a consensus that there was a need for the IHS;

[5]Formula set forth by Section 21 of the Federal Highway Act of 1921.

[6]Since the President's mother was seriously ill the speech was delivered by Vice President Nixon, who read from the President's notes.

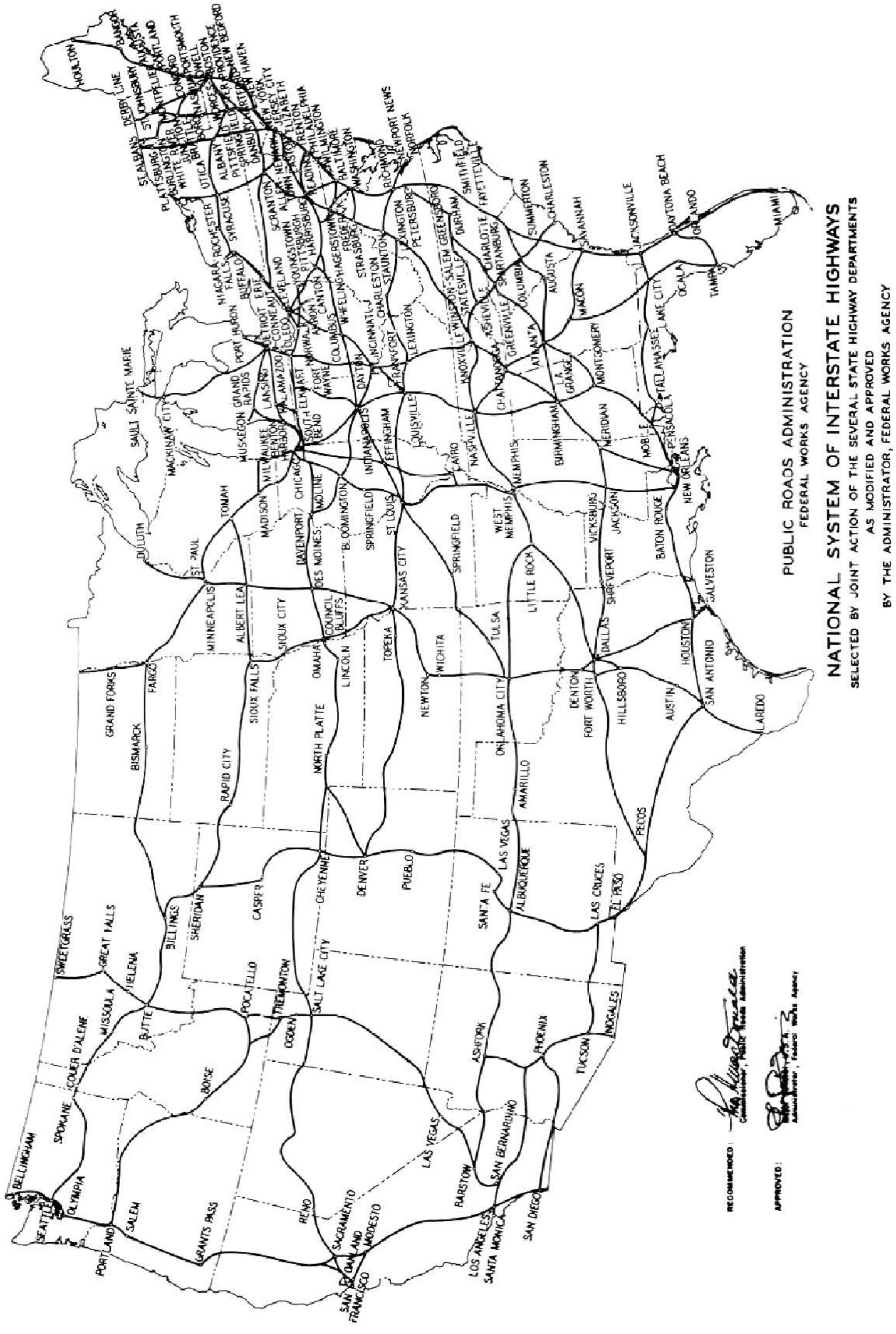


Figure 3.1: The Projected System of Interstate Highways in 1947

however, there was no agreement on how to pay for it.[7] Using information on a report that was currently being developed by the Bureau of Public Roads, the Clay committee estimated the program would cost $27.2 billion (January 1955). They suggested for the Federal Government to cover $25 billion and to finance it with a 30-year bond. The financial plan set forth by the Clay committee had very little support and was rejected by Congress.

After legislation failed in 1955, it was predicted that in 1956 (a presidential election year) the Democratic Congress would not approve such an important plan sought by a Republican president. However, Eisenhower continued to urge approval and worked with Congress to reach compromises. New legislation in 1956 proposed to finance the interstate with the creation of a Highway Trust Fund (HTF), which would collect a tax of 3 cents per gallon on gasoline and diesel, along with other excise taxes on highway users.[8] The idea was for the HTF to be modeled after the Social Security Trust Fund; revenue would go into the general treasury, but credited directly to the Fund. The HTF was a successful compromise which lead to the approval of the Federal-Aid Highway Act of 1956.[9]

The Act of 1956 is sometimes referred to as the IHS Act as it set forth a plan for completing the IHS. First, it created the HTF to finance highway federal-aid; at the time this included the IHS and the ABC program.[10] Second, it envisioned that the IHS would be completed in the following 13 years. Third, it provided more substantial federal-aid funds than its predecessors, totaling $25 billion to be spent during the 13 year period considered. Fourth, it changed the matching funds rule to 90% Federal - 10% State, which provided more incentives for states to invest in the IHS.[11] This matching rule prevailed until the final federal-aid appropriations took place in 1996. The state matching funds rule, together with the $25 billion appropriation, meant total funds equaled 6.2% of

[7] See https://www.fhwa.dot.gov/infrastructure/originalintent.cfm

[8] The HTF was also to be funded with taxes on tire rubber, tube rubber, new trucks, buses, and trailers. As of 2020 the HTF still exists, however it now collects a fuel tax of 18.4 cents per gallon on gasoline and 24.4 cents per gallon on diesel.

[9] The 1956 Act passed congress with 89 in favor and only 1 against, and was signed by President Eisenhower on June 29, 1956.

[10] The ABC program is a Federal-aid program that provides funds for Primary and Secondary Highway Systems, as well as for extensions of these systems within urban areas.

[11] The federal government actually covered 90.4% of the funds as section 108(e) of the Act of 1956 specified that the federal government would cover a percentage of the remaining 10% in any state where the ratio between the area of Federal lands and nontaxable Indian lands to the total area of the state exceeded 5%. The additional percentage was equal to 10% times such ratio and was capped at 5%. This rule affected only 12 states.

GDP.

For 1957 to 1959 the apportionment formula was the same as the one provided by the Act of 1954. For the subsequent years, the 1956 Act provided a different formula, solely based on the relative costs of completing the IHS. That is, the formula was equal to the ratio of the estimated cost of completing the system in each state compared with the cost in all states.[12] To keep this formula up to date, the cost-estimate of completing the IHS was to be updated periodically by the Secretary of Commerce.[13] The logic behind this method was for all states to finish construction of the IHS around the same time.

Even though subsequent acts, amendments and resolutions shaped the future years of the IHS, its essence remained linked to the Act of 1956. The most important changes were triggered by the rising estimated cost of the system, which delayed the end of its construction until 1996 and required considerably more appropriations than what the original plan considered.

Figure 3.2 shows how appropriations and expenditures of federal funds evolved from the beginning of the program. While the final appropriation took place in fiscal year 1996, expenditure continued in the 2000s because funds had been obligated but not yet spent. The procedure by which spending took place is also illustrated in Figure 3.2: (1) First, an estimate of the cost of completing the interstate was released. (2) Then, an authorization took place in a Federal Highway Act. These authorizations outline the amounts that would be available at the national level for the following couple of fiscal years. (3) Funds were then apportioned across states using formulas provided by legislation. The share each state receives is called the apportionment factor (AF). For each fiscal year apportionment factors were usually announced between 1 and 2 years in advance; however, they could be predicted with accuracy many years in advance using the formulas set forth by legislation. (4) Once the fiscal year of the appropriation was reached, states obligated funds in interstate highway projects. (5) Finally, as highways were built, spending took place. Payments to

[12] The Federal-Aid Highway Act of 1963 slightly changed the formula starting in fiscal year 1967. The new formula considered the ratio of the federal share of the estimated cost of completing the system in each state compared to the federal share of the estimated cost of completing the system in all states.

[13] This responsibility was later transferred to the Secretary of Transportation.

contractors for work completed were initially made from state funds[14] and the federal share was paid as reimbursements.

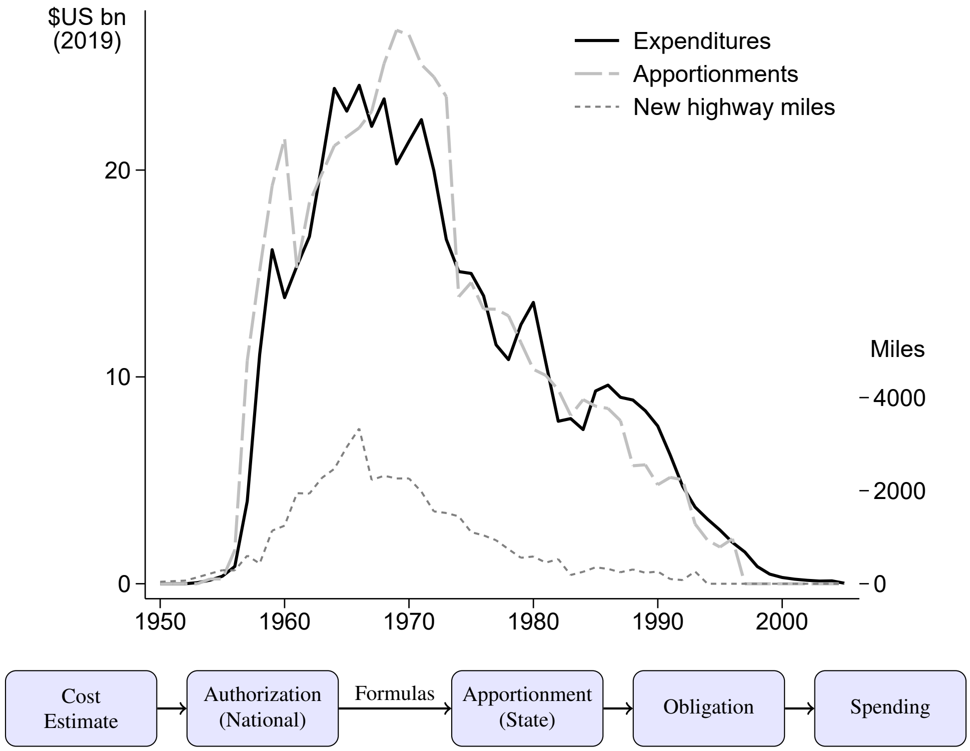


Figure 3.2: Federal Government Funds to Construct the IHS
(Billions of 2019 USD)

As years progressed a few more routes were added into the system, and others deleted. Figure 3.3 presents a digitized version of the 1947 map[15] together with the a digital map of the IHS as of May 2014.[16] Visual inspection of Figure 3.3 shows that the IHS followed the 1947 plan very closely. In fact, at the county-level, the correlation between the number of miles received by each county according to the 1947 plan, and the observed IHS (as of May 2014) is equal to 86%.[17]

States were required to spend apportioned funds within two years or forfeit them. This

[14]Sometimes from funds transferred to the state by cities, counties, or other local governments

[15]Digitization by the authors (Leff), using the USA Contiguous Equidistant Conic projection, which closely matched the layout of the 1947 plan.

[16]Interstate highways according to the National Highway Planning Network, version 14.05.

[17]This calculation uses the county boundary definitions from the 2015 census, and the 48 contiguous U.S. states. Based on 3,107 observations.

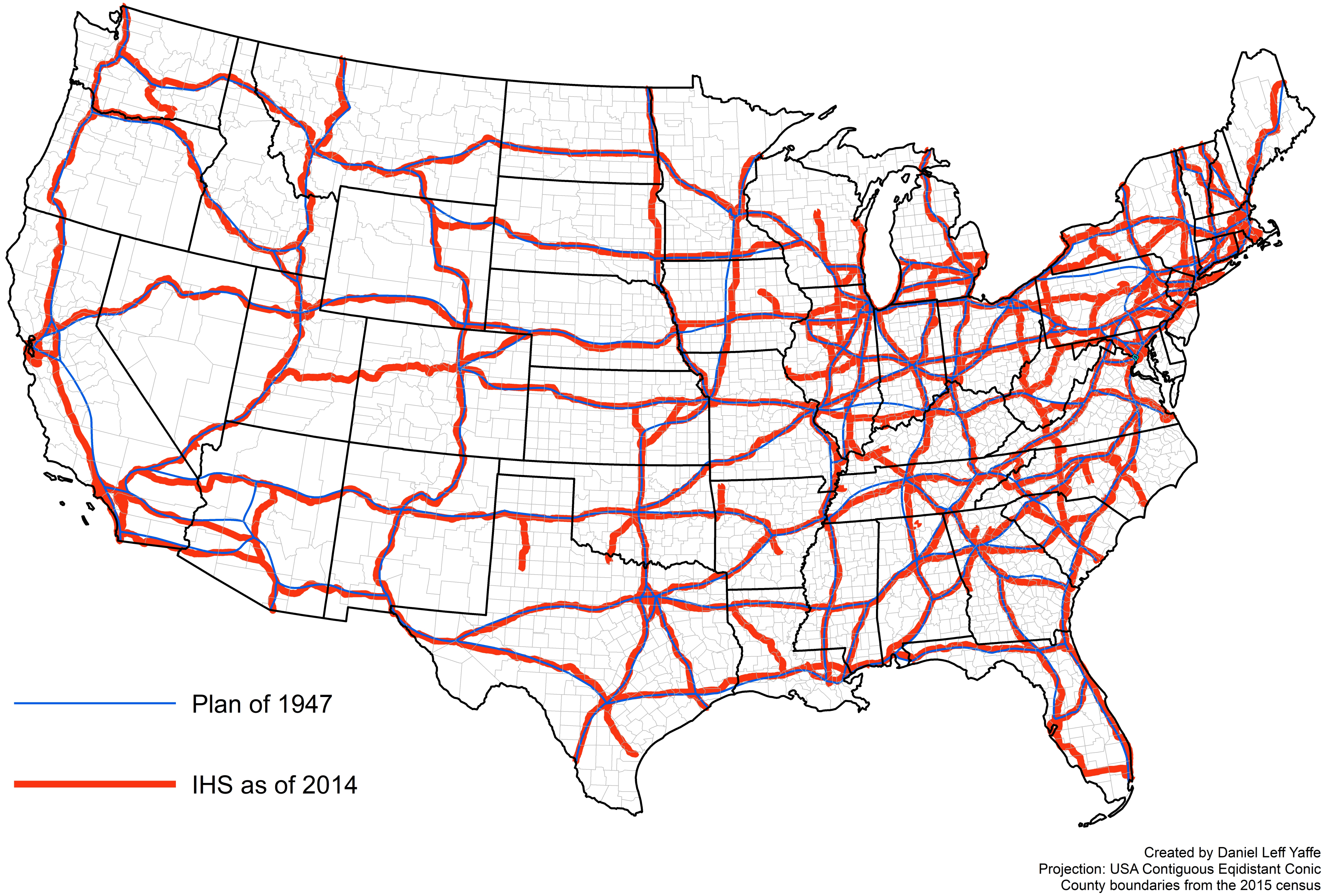


Figure 3.3: The 1947 Plan vs. the 2014 System

constrained somewhat governors' ability to manipulate the timing of construction for electoral gains. Still, it seems that they used what wiggle room they had to full effect. Figure 3.4 in the next section shows that both apportionments and construction increased as election years approached.

3.3 Data and Methodology

In this section we outline our empirical strategy for measuring the causal effect of IHS construction on gubernatorial incumbents' electoral fortunes.

Our primary explanatory variable is the number of miles of interstate highway built in a county $\times$ year (we index counties by i; years by t). We obtain this data from Baum-Snow (2007), who created the data by combining the PR-511 data set with a digital map of the interstate system.[18] Most counties don't change over the period we study, but some do. To address this, we collapse all counties into 3058 time-consistent county boundaries – this is what we mean in this paper when we refer to "counties."[19]

Our primary outcome variable of interest is the county-level vote share of the gubernatorial candidate from the incumbent governor's party.[20] We also examine vote share for the presidential candidate from the incumbent president's party. These data come from two databases of the Inter-university Consortium for Political and Social Research: (1) General Election Data for the United States, 1950-1990, and (2) Candidate Name and Constituency Totals, 1788-1990. We limit our analysis to the period 1950-1972.

Table 3.1 shows summary statistics for these variables at the county, county $\times$ year, and county $\times$ election term level.

[18] The PR-511 data set was created by the government, by requiring each state to report the completion month of each interstate highway within its borders.

[19] A Cartographic Boundary Shapefile at the county-level for the year 2000 was downloaded from the United States Census Bureau. This file included a total of 3108 counties for the 48 contiguous states. Using Census information on *Substantial Changes to Counties and County Equivalent Entities*, we aggregated counties and obtained 3058 counties with time-consistent boundaries from the year 1940 to the year 2000.

[20] Many states impose term limits on governors, creating limits on the electoral rewards an individual politician can reap. For this reason we focus on rewards accruing to the party.

Table 3.1: Summary statistics

	Mean	Median	SD	N
County				
Total miles ever built in county i	10.52	0.00	21.50	2925
Dummy: any miles ever built in county i	0.37	0.00	0.48	2925
County* $\times$ *year				
Num. new miles built in year t	0.44	0.00	2.55	73392
Num. new miles built in year t | county ever gets IHS	1.19	0.00	4.11	26904
Dummy: any new miles built in year t	0.05	0.00	0.22	73392
County* $\times$ *election term				
New miles built in term ending in year t	1.23	0.00	5.18	24337
Dummy: any new miles built in term ending in year t	0.10	0.00	0.29	24337
Incumbent party gub. vote share	60.18	56.34	18.65	23813

States with at least one 4-year election term in the period 1950-1972: AL, AZ, CA, CO, CT, DE, FL, GA, ID, IL, IN, KY, LA, ME, MD, MA, MI, MN, MS, MO, MT, NB, NV, NJ, NM, NY, NC, ND, OH, OK, OR, PA, SC, TN, UT, VA, WA, WV, WI, WY. Some of these states switched from 2-year to 4-year gubernatorial terms during the study period;

3.3.1 Empirical strategy

Here we lay out our strategy for recovering the causal effect of IHS construction on electoral outcomes.

Let Y_{it} denote the share of votes received by the gubernatorial incumbent party in county i, during the gubernatorial election of year t. Let X_{it} denote the number of interstate highway miles opened in county i during year t. If highway construction were randomly assigned across time and space, we might estimate the effect of opening an extra mile with the following specification:

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \tag{3.1}$$

where β is the parameter of interest and ε_{it} is the error term. Note that since gubernatorial elections are scheduled to occur every 2 or 4 years (depending on the state), the database to estimate equation (3.1) is by construction an unbalanced panel.

However, a naïve OLS regression of equation (3.1) is likely to deliver a biased estimate of β: political parties may assign X_{it} where it will be the most electorally beneficial. Descriptive statistics suggest this is indeed the case. Figure 3.4 shows that over the period we consider, construction (right axis) increased as election years approached. (Apportionments [left axis] also increased over the electoral cycle.)

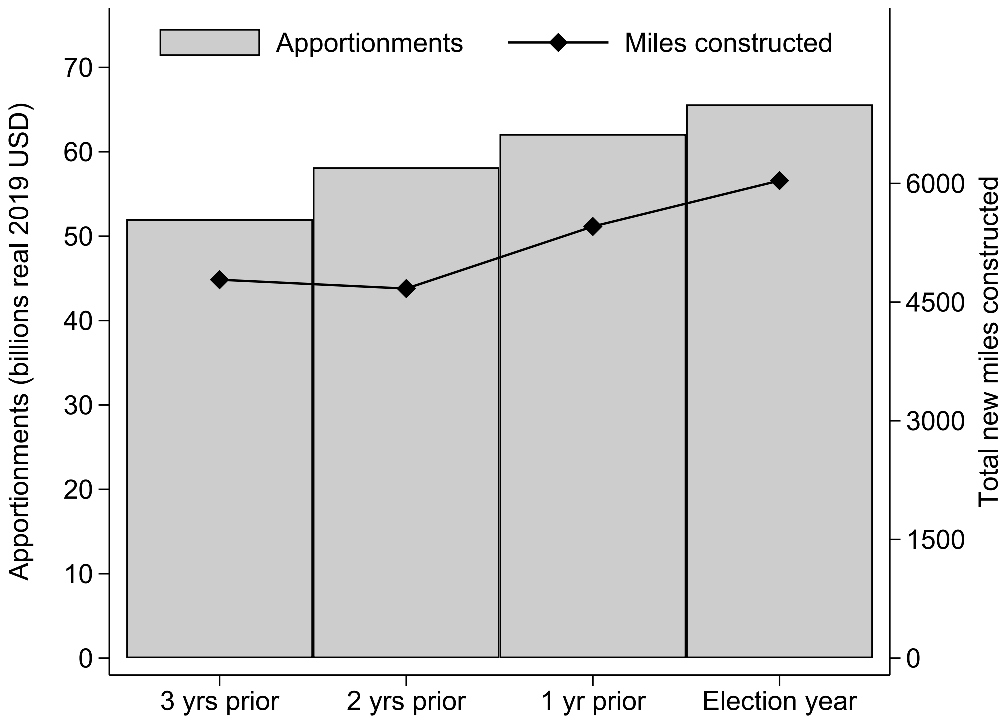


Figure 3.4: Timing of apportionments and construction in the election cycle
1954-1972. Apportionments and miles constructed both summed across years. Includes only full election terms of 4-year length. States with at least one 4-year election term in the period 1950-1972: AL, AZ, CA, CO, CT, DE, FL, GA, ID, IL, IN, KY, LA, ME, MD, MA, MI, MN, MS, MO, MT, NB, NV, NJ, NM, NY, NC, ND, OH, OK, OR, PA, SC, TN, UT, VA, WA, WV, WI, WY. Some of these states switched from 2-year to 4-year gubernatorial terms during the study period; this plot only includes 4-year terms from such states.

The same phenomenon is visible in Table 3.2, which shows that counties built about 15% more miles in election years during this period (an "effect size" of .07 miles on a base of .44). This is true even when controlling for county and year fixed effects. Apportionments also increase in election years. This may reflect the aligned interests of the congresspeople writing apportionment bills and the incumbent governors' in their home states.[21]

Table 3.2: Descriptive: more construction (and apportionments) in election years

	New miles constructed (County × year level)				Apportionments (State × year level)		
Gubernatorial election year	0.057**	0.066**	0.054	0.051*	4.147**	3.243	3.400
	(0.025)	(0.032)	(0.041)	(0.028)	(1.907)	(2.550)	(2.116)
Year FE		✓	✓		✓	✓	
FE	State	County	County	County	State	State	State
Term length	All	All	4yr	2yr	All	4yr	2yr
N	68281	68281	42889	25392	1066	686	380
DV Mean	0.442	0.442	0.492	0.358	28.306	30.729	23.932

New Miles regressions are at the county × year level and include 1950-1972. Apportionment regressions are at the state × year level and include 1954-1972. All regressions cluster standard errors at the state level.

Fixed effects

We address this bias in two steps: fixed effects and instrumental variables. First, we include county-level fixed effects to examine only within-county temporal variation in construction. This allows us to control for the fact that the number of miles assigned across counties is not randomly assigned; counties assigned more miles are different than those that are assigned less miles. We also include year fixed effects to control for the fact that federal appropriations at the state level tend to correlate with each other in large national trends (see Figure 3.2). We left with the following equation which we argue produces a less-biased estimate of β:

$$Y_{it} = \beta X_{it} + \mu_i + \gamma_t + \varepsilon_{it} \tag{3.2}$$

where β is the parameter of interest, μ_i are county fixed effects, γ_t are year fixed effects, and ε_{it} is the error term.

[21]Future research will test whether federal Representatives and Senators enjoyed electoral benefits of IHS construction.

However, fixed effects do not completely address the endogeneity problem, as a timing selection problem remains: in each year, state governments chose where to build the next portion of the interstate system. To address this problem, we create an instrument that captures the number of miles that would have been constructed in a given year had state governments allocated interstate highway construction uniformly across the federally assigned jurisdictions.

Instrumental variables

We construct a Bartik-style instrument which simulates how many miles would be constructed in each county in each year if state government allocated construction within each year uniformly across counties (Bartik, 1991; Goldsmith-Pinkham et al., 2020). We use the federal 1947 plan for the entire IHS system, which closely predicts which counties ended up seeing IHS construction (see Figure 3.3). Because this design was created in 1947 by the federal government for the purpose of facilitating trade and national defense, we think it is plausibly exogenous to the electoral strategies of politicians in individual states in the following decades.

Let $Plan47_i$ be the number of miles assigned to county i in the 1947 plan. This variable is estimated by digitizing the 1947 plan and measuring the number of miles inside each county using the USA Contiguous Equidistant Conic projection. Then, the instrument, denoted by Z_{it}, can be calculated by:

$$Z_{it} = \left(\frac{Plan47_i}{\sum_{i \in S(i)} Plan47_i} \right) \sum_{i \in S(i)} X_{it} \tag{3.3}$$

where $S(i)$ is a function that assigns each county to its respective state. For example, the 1947 plan assigned San Diego county 7.7% of all interstate miles in California. Then, San Diego's instrument for year t multiplies 7.7% times all the miles opened in California in year t. The 1947 plan should be a valid instrument as long as its creation was not influenced by the reelection strategies of incumbent parties between 1950 and 1972.

This leaves one potential source of endogeneity: the two-year window state governments have within which to spend apportioned funds (Figure 3.4). This allows politicians to manipulate

$\sum_{i \in S(i)} X_{it}$, the total IHS spending in state $S(i)$ in year t. In order to remove this source of endogeneity, we create an alternate instrument $\widetilde{Z}_{it}$ which uses state $\times$ year-level apportionments from Congress ($W_{S(i)t}$) in place of expenditures. These apportionments are of course strongly related to expenditures (see Figure 3.2), but are controlled by the federal government rather than state governments:[22]

$$\widetilde{Z}_{it} = \left(\frac{Plan47_i}{\sum_{i \in S(i)} Plan47_i} \right) W_{S(i)t} \tag{3.4}$$

We note a small source of measurement error arising from the timing of the variables: in the U.S., elections generally occur in November, while data on the number of opened miles covers the whole calendar year. In our data, 98.3% of the observations are for elections that happened in November, while the other 1.7% occurred before October. We leave these early-year elections in our data.

Another possible source of misspecification in equation (3.2) might arise due to counties having different areas. It is possible that a new highway mile has less impact on the share of votes in a county that has more area. In an alternate specification, we normalize the variables X_{it} and Z_{it} by area A_i, in the following way: $x_{it} = (X_{it}/A_i)\bar{A}$ and $z_{it} = (z_{it}/A_i)\bar{A}$, where $\bar{A} = \sum_i A_i/N$ and $N = 3,058$.[23] The inclusion of $\bar{A}$ in the formula simply scales the β coefficient to give it the same interpretation as before: the effect of opening an extra mile in county i on the share of votes received by the incumbent.

Finally, we consider alternate "treatment" definitions. Our main specification consider the main explanatory variable to be the number of miles built in a county in a year. However, it may be that the fact of having *any* interstate access is more important in voters' minds than the *number* of miles built. We introduce a specification which defines the explanatory variable as an indicator for whether any miles were built in the county in that year. Alternatively, voters may be attuned to construction not only in election years but during the entire election term of a sitting governor. We

[22]We thus instrument not only for the "share" on the left of our expression for Z_{it}, but also the "shift" on the right, as in Autor, Dorn, and Hanson (2013) equation (4).

[23]The area for each of the 3,058 counties was calculated using the USA Contiguous Albers Equal Area Conic projection.

create another alternate specification which lets the explanatory variable be the number of miles built in a county during a governor's entire election term. Finally, for completeness's sake, we create a specification whose explanatory variable is an indicator for whether *any* IHS construction happened in the county during the election term. These specifications' results appear in Table 3.4.

3.4 Results

Panel A of Table 3.3 presents the results of estimating equation (3.1) using different estimation methods and control variables. For every specification, standard errors are clustered at the state level.

Column (1) is our base OLS specification, with county and year fixed effects. In this specification, constructing an additional IHS mile in an election year is associated with a statistically significant increase in the incumbent governor's party's vote share of 0.126 pp. Is this large? Recall that the average county receives a bit more than one IHS mile per year *conditional* on receiving any. Recall too that the average gubernatorial candidate from the incumbent party received a vote share of 60% in this period.

Column (2) estimates the equation with the instrument Z_{it} (and the same battery of fixed effects). The new point estimate is equal to 0.646, although the precision is slightly lower ($p = .05$). The fact that the point estimate is higher with IV is consistent with the idea that political parties seem to target spending in counties where they have less of a voting base. Instead of spending where political support is already high, politicians seem to target spending where they have more votes to gain. It is important to note that the first-stage Kleibergen and Paap (2006) F-statistic of this specification is far above the standard weak instrument threshold of 10 (Staiger & Stock, 1997).

Columns (3) and (4) present results from two small extensions of the baseline model. In column (3) we add a lag of the new highway miles variable to control for the possibility that spending from the year prior to the election might also affect voters, using $Z_{i,t}$ and $Z_{i,t-1}$ as instruments for

Table 3.3: Effect of highway construction on incumbent governor's vote share

	OLS			IV: Z_{it}		IV: $\widetilde{Z}_{it}$	
Panel A: Not normalized by Area							
New miles this year	-0.256**	-0.026	0.125**	0.646*	0.883*	2.208**	2.521**
	(0.106)	(0.071)	(0.059)	(0.326)	(0.440)	(0.846)	(1.000)
L.New miles this year					-0.459		
					(0.458)		
New miles in nbor counties							-0.170*
							(0.098)
Year FE		✓	✓	✓	✓	✓	✓
County FE			✓	✓	✓	✓	✓
N	23813	23813	23813	23813	23813	19429	19429
DV Mean	60	60	60	60	60	59	59
KP 1S F				893	221	60	54
Panel B: Normalized by area							
New miles this year	-0.083*	-0.009	0.036	0.629*	0.877***	1.932***	2.571**
	(0.042)	(0.026)	(0.028)	(0.323)	(0.306)	(0.663)	(1.042)
L.New miles this year					-0.536		
					(0.624)		
New miles in nbor counties							-0.232*
							(0.120)
Year FE		✓	✓	✓	✓	✓	✓
County FE			✓	✓	✓	✓	✓
N	23813	23813	23813	23813	23813	19429	19429
DV Mean	60	60	60	60	60	59	59
KP 1S F				28	16	6	5

Observation is at the county $\times$ year level. All regressions include year and county fixed effects. Z_{it} instruments the IHS construction in $county_i \times year_t$ with the product of (a) the fraction of state $S(i)$'s total planned miles accounted for by county i according to the 1947 plan; and (b) the sum of highway miles constructed in state $S(i)$ in year t. The instrument $\widetilde{Z}_{it}$ is the product of (a), and (c) the amount of money apportioned by Congress for IHS construction in state $S(i)$ in year t, measured in millions of real 2019 US dollars. $\widetilde{Z}_{it}$ regressions include years 1954-1972; all other regressions include years 1950-1972. SE clustered by state in parentheses. KP 1S F reports 1st-stage F-statistics according to Kleibergen and Paap (2006). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$X_{i,t}$ and $X_{i,t-1}$. The estimate of this new parameter is statistically insignificant, but including it increases the point estimate of new miles constructed in the year of the election.

It is conceivable that the IHS brings benefits to adjacent counties even if it does not pass through them, and that voters in these counties might reward the governor. Column (4) tests whether opening a highway mile in a county that shares a border with county i affects votes for the incumbent in county i.[24] The results suggest that new highway miles in neighboring counties do not impact the share of votes received by the incumbent party.

Columns (5), (6), and (7) present results from regressions analogous to those in columns (2), (3), and (4), but using the $\widetilde{Z}_{it}$ instrument which incorporates apportionments rather than actual expenditure. The coefficients do not change signs, but jump markedly in magnitude. This is consistent with the evidence we have that politicians manipulate their spending toward election years.

Panel B presents the results when the explanatory variable of interest (new highway miles) and the instrument are normalized by the county's area as described before. The estimates remain similar in sign and magnitude, though there are some changes in precision. Note in particular that the first-stage F-statistics for Z_{it} and especially $\widetilde{Z}_{it}$ are quite low in some specifications, suggesting the instruments may be too weak to provide good inference.

3.4.1 Alternate treatment definitions tell the same story

It is possible that the effect of highway construction on voters is not simply linear in highway miles constructed during an election year. Table 3.4 shows the estimated coefficients from the fixed effects and IV specifications for alternative treatment definitions. Column (1) provides the basic year- and county-fixed-effect specification. Columns (2) and (3) include the same fixed effects, but let Z_{it} and $\widetilde{Z}_{it}$ (respectively) instrument for each alternative treatment variable presented in the panels A, B, and C.

Panel A's treatment variable is the number of highway miles constructed in the county during the governor's term – the four years leading up the election. Panel B defines treatment using an indicator variable for whether *any* highway miles were built in the year of the election. Panel C's treatment variable is an indicator for whether *any* new miles were constructed during those four years.

[24] We calculate a matrix indicating if two counties are contiguous using the Polygon Neighbors tool in ArcGIS.

Table 3.4: Effect of IHS construction on incumbent governor's party vote share: Alternate treatment definitions

	OLS:FE	IV: Z_{it}	IV: $\widetilde{Z}_{it}$
Panel A:			
New miles this term	0.106***	0.317*	0.631**
	(0.036)	(0.173)	(0.263)
N	23813	23813	19429
DV Mean	60	60	59
First Stage F		80	79
Panel B:			
Any new miles this year	1.351*	10.400*	35.230***
	(0.765)	(5.325)	(12.314)
N	23813	23813	19429
DV Mean	60	60	59
First Stage F		193	14
Panel C:			
Any new miles this term	1.325*	8.109*	19.606***
	(0.759)	(4.260)	(7.291)
N	23813	23813	19429
DV Mean	60	60	59
First Stage F		120	10

All specifications include county and year fixed effects. Standard errors clustered at state level in parentheses. 1950-1972. * p<0.10, ** p<0.05, *** p<0.01

These results do not alter the basic finding that IHS construction helps incumbent governors, but they shed light on the mechanisms that could be at play. In Panel A, coefficients for all three specifications are smaller than in Table 3.3, which examined the effect of construction during election years. This suggests that IHS construction helps incumbents most when it happens in election years (consistent with the insignificant coefficient on lagged IHS construction in Column 3 of Table 3.3). In Panels B and C, coefficients are much higher – this makes sense because since the explanatory variable is binary, it aggregates all amounts of construction. As with Panel A vs. Table 3.3, the coefficients are smaller in Panel C, suggesting that construction in election years is most impactful.

3.4.2 No effects on presidential vote share

Although we show that IHS construction caused higher vote share for incumbent governors, recall that 90% of the funding came from the federal government. Did federal politicians receive a similar boost to their vote share from IHS construction? Table 3.5 shows that in the case of presidential candidates, the answer is no. In nearly all specifications, the coefficients are not statistically significant, and some of the signs have switched, though these effects are in general not very precisely estimated. (The sample size is lower because there are fewer presidential election years than gubernatorial election years.) This result is consistent with more local politicians being able to more successfully claim credit. Further work will study whether federal legislators in the House and Senate, and/or local politicians, receive electoral benefits from highway construction.

3.4.3 Effects on county finances

The construction of the highway may affect other outcomes besides voting decisions. Highways are not only a public good in and of themselves; they also perform important functions in how society is organized. This in turn has implications for how governments fund themselves and what activities they engage in. For example, access to the highway might drive economic growth, leading to higher tax revenues and greater provision of other public goods. Or state governments who invest in building a highway in a county might treat that money fungibly and take away funding from other resources in the county, such as hospitals.

Table 3.6 shows the effect on some key elements of county finances (using the $\widetilde{Z}_{it}$ instrument based on state-level appropriations). In particular, we focus on county-level tax take, inter-governmental (IG) revenue from the state, direct expenditure, IG revenue for hospitals, IG revenue for highways, transit subsidies from the state, county's own expenditure on highways, and county's own total expenditure. All of these are measured in nominal USD per capita.

These preliminary results suggest that the highway did not have any immediate effect on total taxes, overall revenue from the state, or revenue for hospitals and highways. Direct expenditure

Table 3.5: Effect of highway construction on vote share for the presidential candidate from the presidential incumbent party

	OLS			IV: Z_{it}		IV: $\widetilde{Z}_{it}$	
	Panel A: Not normalized by Area						
New miles this year	-0.041	0.074**	0.052	0.349	0.579	-2.331	-2.809
	(0.047)	(0.031)	(0.041)	(0.308)	(0.418)	(1.606)	(1.763)
L.New miles this year					-0.384		
					(0.413)		
New miles in nbor counties							0.230**
							(0.108)
Year FE		✓	✓	✓	✓	✓	✓
County FE			✓	✓	✓	✓	✓
N	18257	18257	18257	18257	18257	15199	15199
DV Mean	48	48	48	48	48	49	49
KP 1S F				410	218	22	28
	Panel B: Normalized by area						
New miles this year	-0.037	0.021	0.029	0.013	0.148	-9.312	-9.094
	(0.025)	(0.027)	(0.025)	(0.308)	(0.344)	(29.993)	(20.936)
L.New miles this year					-0.265		
					(0.297)		
New miles in nbor counties							0.890
							(2.069)
Year FE		✓	✓	✓	✓	✓	✓
County FE			✓	✓	✓	✓	✓
N	18257	18257	18257	18257	18257	15199	15199
DV Mean	48	48	48	48	48	49	49
KP 1S F				53	37	0	0

Observation is at the county $\times$ year level. All regressions include year and county fixed effects. Z_{it} instruments the IHS construction in $county_i \times year_t$ with the product of (a) the fraction of state $S(i)$'s total planned miles accounted for by county i according to the 1947 plan; and (b) the sum of highway miles constructed in state $S(i)$ in year t. The instrument $\widetilde{Z}_{it}$ is the product of (a), and (c) the amount of money apportioned by Congress for IHS construction in state $S(i)$ in year t, measured in millions of real 2019 US dollars. $\widetilde{Z}_{it}$ regressions include years 1954-1972; all other regressions include years 1950-1972. SE clustered by state in parentheses. KP 1S F reports 1st-stage F-statistics according to Kleibergen and Paap (2006). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

seems to have dropped significantly, however. State transit subsidies seem to be crowded in, perhaps to complement increased transportation options. Municipalities' own expenditures on highways go down in response, but their total expenditure is unaffected.

Table 3.6: Effect of highway construction on county finances

	Total Taxes	Revenue from state	Direct expend-iture	State revenue hospitals	State revenue highways	State transit subsidy	Own expend highways	Own total expend
New miles	10.574	9.715	-10.418**	0.069	20.380	1.756**	-0.007**	-0.000
	(16.817)	(14.035)	(3.928)	(0.077)	(32.519)	(0.857)	(0.003)	(0.018)
N	58648	58648	58648	58648	58648	58648	58648	58648
DV Mean	122.859	108.366	28.124	0.038	322.065	6.490	0.027	0.210
First Stage F	31.487	31.487	31.487	31.487	31.487	31.487	31.487	31.487

Notes: Observation is at the county $\times$ year level. Outcomes are in USD per capita. All regressions include year and county fixed effects, and are instrumental variables regressions using $\widetilde{Z}_{it}$ Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

These results are preliminary, and we urge readers to take them with a grain of salt. There is so much data available on county-level finances that a researcher could easily run dozens of regressions and report only the results from those with significant (likely spurious) coefficients. We are still in the process of creating a strategy for examining this data in a credible way.

3.5 Public Good Provision and Connectivity

Highways play a role not only as a public good for a particular location *per se*, but also as connectors between different locations. This connection could play a role in how voters react to public spending writ large, above and beyond their reaction to the road itself as a public good.

In this section we document a new fact regarding the effects of connectivity on the allocation of government spending. We find that when neighboring counties are connected to one another by the interstate, public good provision becomes more homogeneous between these counties. To study the effect of connectivity on the homogeneity of spending, for any two pair of contiguous counties i and j, consider the following specification:

$$| Y_{it} - Y_{jt} | = \beta X_{ijt} + \mu_{ij} + \gamma_t + \varepsilon_{ijt} \tag{3.5}$$

where Y is a per capita measure of public good provision, X is a dummy variable for whether counties i and j are connected by an interstate highway at time t, μ_{ij} are fixed effects at the county-pair level, γ_t are time dummies, and ε_{ijt} is an error term.

Table 3.7 shows the results on three main outcomes: Intergovernment (IG) revenue from the state, direct expenditure, and IG revenue for hospitals. Getting connected by a highway decreases the per capita absolute difference of: a) IG revenue from the state government by 1.6 USD; b) direct expenditure by counties by 5.6 USD; and (c) Hospital by 2.08 USD.

Table 3.7: **Effect of IHS connection on Differences in Per Capita Public Good Spending**

	(1)	(2)	(3)
	IG Revenue from State	Direct Expenditure	Hospitals
X_{ijt}	-1.60**	-5.61**	-2.08**
	(0.68)	(2.32)	(0.88)
Time Dummies	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Obs.	57,368	57,368	57,368

Notes: * p<0.10, ** p<0.05, *** p<0.01. Driscoll-Kraay standard errors in parentheses, robust to heteroscedasticity, autocorrelation, and cross-sectional dependence. Lagged of dependent variable included as control.

These results are new and require more robustness testing. For now, we build a model in the following section to rationalize this new fact.

3.5.1 Connectivity: Model

We propose a reward-punishment retrospective probabilistic voting model to explain the new empirical findings. The model consists of one incumbent who wants to stay in office due to ego rents (Rogoff, 1990; Shi & Svensson, 2006) and who is responsible for allocating its jurisdiction resources between different location. After the incumbent spends all public resources, elections are held and voters, spread across different locations, decide to vote for the incumbent's reelection or vote for the opposite party according to the information they have about the incumbent public spending, the utility they get from the local public goods and their own idiosyncratic tastes (ideology realization that is drawn from a known ideology distribution).

In the model, there are information frictions. Agents do not fully observe what the incumbent does in all locations. They observe what the incumbent spends in their own location but only observe what the incumbent does in other locations if those locations are connected. In our setup, connections between different locations increase the information voters have on how the incumbent allocates resources. We assume that agents vote according to the direct utility they get from local public good provision but also according to the gains and losses coming from the relative allocation of public good provision between locations. Therefore, we develop a model that incorporates reference-dependence preferences in the voting function (in which the reference point is other locations public good supply), incorporating to the electoral incentives literature the idea developed by Kahneman, Tversky, Kahneman, and Tversky (1979) later applied by Kőszegi and Rabin (2006) on reference-dependent preferences and endogenous reference point determination.

Intuitively, there are three main reasons on why agents would care about the relative allocation of public goods. First, agents do not fully observe how the incumbent behaves and observing what other locations get increases knowledge when there is incomplete information. Voters may use the total amount of public goods to measure how productive the incumbent is given the total amount of taxes that are being paid. Second, if a location receives a low amount of public goods but voters do not observe what the exact budget is, they might think that total spending might be low due to an exogenous shock but if afterwards they find out that other locations (with similar socioeconomic characteristics) got a large amount of public goods, they know they are being discriminated which may impact on voting behavior. Finally, we may assume there is some kind of "keeping up with the joneses" type of mechanism in which voters might feel jealous if other locations receive more public goods even if their marginal utility of public good consumption is higher due to, for example, worse socio-economic conditions. In our main analysis we assume that all locations have the same socio-economic conditions and size so that any difference in public good allocation is driven by electoral motives.

In the equilibrium, as connections increase public information, they generate a change in incumbent's behavior. This happens as allocating more resources to some specific locations with large number of pivotal voters might hurt the voting share in other locations. In particular, when locations are not connected the incumbent may have incentives to exploit the heterogeneity in voter taste distribution, by spending more resources where she might capture more votes but if a connection is built between those locations, the incumbent has to take into account the negative spillover effect in other locations voting shares.

3.5.2 Set up of the model

The model consists of one incumbent that wants to stay in office and a mass 1 of voters that reward or punish the incumbent during elections by taking into account her behavior while in office. Voters are spread across different I locations, which are indexed by i and each one of the locations have a size θ_i which can be also interpreted as the share of agents in that location i.

Timing of the model. First the incumbent allocates g_i^k (public good expenditure of type k in location i), then agents of group i observe g_i^k and potentially observe $g_{j\neq i}^k$ and, finally, elections are held. The types of public goods represents different categories as hospitals, schools, or direct transfers such us welfare subsidies. For simplicity we assume there are two types of public goods, visible goods v and non-visible goods n. The visible goods are the ones that can be easily seen as hospitals, roads and schools. On the other hand, the non visible public goods are the ones that only are perceived by the people that receives them, for example direct transfers or welfare subsidies.

Elections and Voter's Behavior. Note that in retrospective voting models the equilibrium must be solved with backward induction. First, the decision of households have to be defined, so that the incumbent allocates resources optimally taking agents' best responses into consideration. Ex ante (from the incumbent point of view) the agent of group i will vote for the reelection if the expected utility generated by the transfers is higher than the ideology shock:

$$U(g_i, g_{j\neq i}, \beta_i) > \phi_i$$

Where g_i represents the vector of all types of public goods received by i, and β_i is a reduced form parameter that represents how big or small the marginal direct utility from an additional unit of public goods is. Moreover, as in the literature is commonly used, there is an idiosyncratic shock or ideology shock that will determine how much agents like, ex ante, the incumbent. We assume that ϕ, the ideology shock, is uniformly distributed on $\left[-\frac{1}{2\Phi_i}, \frac{1}{2\Phi_i}\right]$. The distributions are common knowledge but only voters from location i observe their own ideology draw. Note that the lower is the idiosyncratic shock the higher is the dislike towards the incumbent which implies that agents in that location need to receive more public goods to vote for the incumbent. Moreover, the mean ideology is the same across locations but the standard deviation is larger as Φ decreases. A higher standard deviation in the ideology distribution in practice means that there is a lower share of pivotal voters which makes capturing votes in that particular location more difficult for the incumbent. The utility function is given by:

$$U(g_i, g_{j\neq i}, \beta_i) = \sum_{k\varepsilon\{v,n\}} \left[H(g_i^k, \beta_i) + \alpha_i \sum_j \left[p_{ij}\lambda_{ij}(g_i^k - g_j^k) + (1 - p_{ij})(g_i^k - g_i^k) \right] W_{ij} \right]$$

where:

$$\lambda_{ij} = \begin{cases} \lambda^G \text{ if } g_i^k > g_j^k \\ \lambda^L \text{ if } g_i^k < g_j^k \end{cases}$$

$$p_{ij} = \begin{cases} 1 - \text{if i and j are connected and } k = v \\ 0 - \text{if i and j are not connected or } k = n \end{cases}$$

The first term $H(g_i, \beta_i)$ represents the direct utility generated by the public good supply. If the government invest more in schools, hospitals or in direct transfers, agents will have higher indirect utility. Recall that β_i is a reduced form parameter that represents how big or small the marginal direct utility from an additional unit of public goods is. For example, if the location is very rich, agents may use more private services which can reduce the return on the public good investment. We assumed that the marginal benefit from the public goods is decreasing, $dH(g_i^v, \beta_i)/dg_i^v > 0$, and that the higher is consumption, income or other socio-demographic characteristics the lower β_i is and, $dH(g_i^v, \beta_i)/d\beta > 0$.

The second term inside the summation comes from the difference between what they get and what they observe other people are receiving (their reference point). This is based on the idea that people care about the distribution of public goods and not only their own level. Agents compare how much transfers they receive relative to other locations and, for example, if their location receives less public goods (hospital, schools or parks) than other locations, this would generate a lower probability of voting for the incumbent in the elections. Nevertheless, agents observe what the locality j received if locality i and j are connected. We consider the case in which $p_{ij} = 1$ if there is a connection and $p_{ij} = 0$ if there is no connection. Therefore, if agents see what other locations are getting ($p_{ij} = 1$), they use the location j per capita spending as the reference point, if they don't observe what location j is receiving ($p_{ij} = 0$), they assume that the government is equally distributing resources among locations and, therefore, there is no change in the utility in terms of the relative distribution. Finally, we assume that if locations are not connected, agents from i cannot see any public good in location j, and if locations are connected, agents from i are only able to see location j visible goods but not the non-visible ones.

As Kőszegi and Rabin (2006) we assume a linear model where λ_{ij} represents how agents weight the losses ($\lambda_{ij} = \lambda^L$ if $g_i < g_j$) and gains ($\lambda_{ij} = \lambda^G$ if $g_i > g_j$) from the public good distribution. We assume that $\lambda^L > \lambda^G$ which means that the increase in utility from receiving relatively more public goods than other locations is smaller than the decrease in utility from receiving relatively fewer goods. The only difference with KR is that the reference point change depending on what the agents observe. Finally, W_{ij} is the weight for each pairwise comparison that agents use. We assume that $W_{ij} = 0$ for all non connected locations and $W_{ij} = 1/N$ for all the N connected ones.

$$\sum_v \left[H(g_i^v, \beta_i) + \alpha_i \sum_j p_{ij} \lambda^{L,G} (g_i^v - g_j^v) W_{ij} \right] + \sum_n H(g_i^n, \beta_i) > \phi_i$$

Incumbent's problem. Ex ante, the ideology draw is unknown by the incumbent but the distribution in each location is not. Therefore, the incumbent takes into account the probability of winning, as there is uncertainty due to the incomplete information setup and the idiosyncratic shock assumption. Agent i will vote for the incumbent with probability:[25]

$$Pr[U(g_i, g_{j\neq i}, \beta_i) > \phi_i] = \int_{-\frac{1}{2\Phi}}^{U(g_i, g_{j\neq i}, \beta_i)} f_i(\phi_i) d\phi_i = \Phi_i U(g_i, g_{j\neq i}, \beta_i) + \frac{1}{2}$$

[25]Recall that $\phi \sim U\left[-\frac{1}{2\Phi_i}, \frac{1}{2\Phi_i}\right]$ and that the pdf is $\frac{1}{\frac{1}{2\Phi_i} + \frac{1}{2\Phi_i}}$ for $\phi \in \left[-\frac{1}{2\Phi_i}, \frac{1}{2\Phi_i}\right]$ and 0 otherwise. Therefore: $\int_{x_1}^{\delta U(.,1)+U(.,2)} f_i(\phi_i) d\phi_i = \int_{-\frac{1}{2\Phi}}^{\delta U(.,1)+U(.,2)} \frac{1}{\frac{1}{2\Phi} + \frac{1}{2\Phi}} d\phi_i = \int_{-\frac{1}{2\Phi}}^{\delta U(.,1)+U(.,2)} \Phi d\phi_i$

In order to increase the probability of holding power, the incumbent maximizes the expected share of votes subject to the resource constraint. Thus, the incumbent's problem will be:

$$\max_{g_i^k \forall i,k} \sum_{i \in I} \theta_i \Phi_i U(g_i^k, g_{j \neq i}^k, \beta_i) + \frac{1}{2}$$

$$st: \sum_{k \in \{n,v\}} \sum_{i \in I} \theta_i g_i^v = T - X$$

Where g_i^k is per capita spending in local public goods in group i and public good type k, θ_i is the share population of group i, T is taxes raised and X is an exogenous and unobserved expenditure, so that it would be impossible for agents of group i to calculate g_j in case they do not directly observe it. Furthermore, the first order conditions with respect to g_i^v and g_i^n are respectively:

$$\theta_i \Phi_i \left[H'(g_i, \beta_i) + \alpha_i \sum_{j \neq i} p_{ij} \lambda^{L,G} W_{ij} \right] - \sum_{j \neq i} \theta_j \Phi_j \alpha_j p_{ji} \lambda^{L,G} W_{ji} = \Lambda \theta_i$$

$$\theta_i \Phi_i H'(g_i^n, \beta_i) = \Lambda \theta_i$$

Where Λ is the marginal increase in votes of one more dollar spend in public goods. If two locations are connected, an increase in the visible goods in one location is going to decrease the voting share in the other location. Nevertheless, an increase in government spending in goods that only are seen by the recipients are only going to increase the recipient utility without having a negative effect in other locations. The government will have incentives to invest more in invisible public goods to exploit the ideology distribution heterogeneity in her favor. Nonetheless, as the marginal utility from each type of public good is decreasing the incumbent will always provide all types of goods.

3.5.3 Connections and Government Expenditure Dynamics

In this section, we start by considering a simple case with 2 locations: A and B. Recall that without connections voters cannot observe what public spending is in other locations and therefore a connection in this setup increases public information. We assume first that both locations have the same socioeconomic characteristics ($\beta_a = \beta_b$) and population size ($\theta_a = \theta_b$), so that the incumbent only needs to decide how to spend its resources according to each location ideology distribution and public knowledge. Figure 3.5 represents the case with two locations for which a connection is built in a second period so we can compare the two static equilibria.

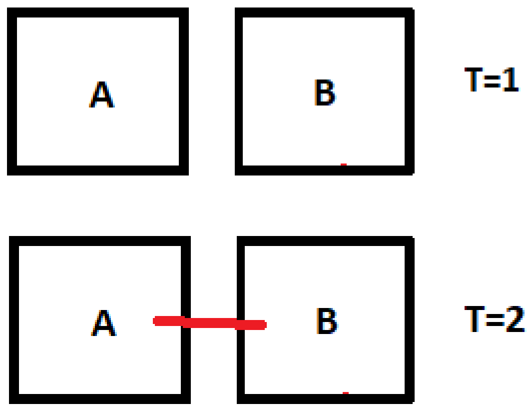


Figure 3.5: Locations and connections

Note that at T=1, there is no connection which means that $p_{ab} = p_{ba} = 0$. Without connections voters of A or B cannot see what agents in the other location are receiving (for both visible and non-visible goods). Therefore, there is no difference in terms of marginal utility in terms of the type of goods in the same location but there are differences in marginal utility across locations driven by the ideology distribution heterogeneity. Both conditions are determined by:

$$H'(g_i^v, \beta) = H'(g_i^n, \beta), \forall i$$

$$\Phi_a H'(g_a^k, \beta) = \Phi_b H'(g_b^k, \beta), \forall k$$

Note that if $\Phi_a < \Phi_b$, then $g_a < g_b$. This means that if the ideology distribution has lower variance, there are more pivotal agents at location B and therefore the marginal benefit in terms of votes is higher in that location which generates incentives to an office seeking candidate to supply more public goods in that location to try to capture more votes. This condition holds for every type of local public good k. Moreover, as the ideology distribution do not change within locations, the incumbent will equalize the marginal utilities for each type of goods.

When these locations are connected, there are new optimal conditions, as providing more visible goods to B will generate a larger discontent in A. Therefore, the incumbent has incentives to homogenize public good spending and to change the expenditure distribution across types of goods (not all goods are visible even when locations are connected). For the visible public goods the optimal condition changes as follows: [26]

$$before : \Phi_a H'(g_a^v, \beta) = \Phi_b H'(g_b^v, \beta)$$

$$after : \Phi_a H'(g_a^v, \beta) = \Phi_b H'(g_b^v, \beta) - 2\alpha \left[\Phi_a \lambda^L - \Phi_b \lambda^G \right]$$

Note that the new result depends on what the utility from gaining is relative to the disutility from losing. If λ^L is large enough relative to λ^G then there would be homogenization when connectivity increases. In the extreme case in which there is only loss from the relative distribution of public goods but not gains ($\lambda^G = 0, \lambda^L > 0$), then there would be homogenization of public good supply as the difference $|g_a - g_b|$ will decrease. Moreover, the strength of this mechanism is defined by α. As α decreases, the equilibrium level of public good expenditures gets closer to the pre-connection equilibrium. Note that if there is heterogeneity in terms of the importance of the reference dependence term inside the voting function, $\alpha_i \neq \alpha_j$, there can be deviations from the optimal public

[26]The optimal condition for the visible local public goods without specifying parameter values for lambda:

$$\Phi_a \left[H'(g_a^v, \beta) + \alpha_a \lambda^{L,G} \right] - \Phi_b \alpha_b \lambda^{L,G} = \Phi_b \left[H'(g_b^v, \beta) + \alpha_b \lambda^{L,G} \right] - \Phi_a \alpha_a \lambda^{L,G}$$

We know that the one that was benefiting from the non-connectivity was B which was the location with more pivotal agents. Therefore, the incumbent has incentives to spend more resources in B than in A, and so in equilibrium we will get that $g_b > g_a$, which means that location A has a loss from the public good allocation, and therefore we use λ^L, and B has a gain from the difference in goods allocation which means they have λ^G as parameter in their voting function. Therefore we get:

$$\Phi_a \left[H'(g_a^v, \beta) + \alpha_a \lambda^L \right] - \Phi_b \alpha_b \lambda^G = \Phi_b \left[H'(g_b^v, \beta) + \alpha_b \lambda^G \right] - \Phi_a \alpha_a \lambda^L$$

good allocation as well.

Interestingly, the dynamics on the non-visible public goods are richer because more connections mean that the voting share gain for the incumbent is higher than before using these types of goods. The optimal condition for this type of goods across locations is defined as:

$$before \text{ and } after: \Phi_a H'(g_a^n, \beta_a) = \Phi_b H'(g_b^n, \beta_b)$$

As this type of public goods cannot be seen the incumbent will equalize the marginal increase in votes across locations which means that she takes into consideration the marginal increase in direct utility and the ideology distribution as before and there is no change in the relative allocation, although there is a change in levels. This change in levels can be analyzed in the condition for the difference between public goods within location:

$$before: \Phi_a H'(g_a^n, \beta_a) = \Phi_a H'(g_a^v, \beta_a)$$

$$after: \Phi_a H'(g_a^n, \beta_a) - \alpha \left[\Phi_a \lambda^L - \Phi_b \lambda^G\right] = \Phi_a H'(g_a^v, \beta_a)$$

Assuming as before that the difference between λ^L and λ^G is large enough, there is a decrease in g^n relative to g^v in A and there is a decrease in g^v relative to g^n in B. The location that receives more public goods ex ante, receives more of non-visible goods so that agents in the other location do not perceive they receive less public goods. Moreover, note that if in period 1 B receives more non-visible goods and in period 2 there are more n public goods supplied, to equalize marginal utilities for n, g_b^n has to increase more than g_a^n (given that $H'(.) < 0$) which means that the difference will increase, while the difference between the visible goods will decrease. In appendix we provide exercises with more than two locations and different connection possibilities which generate the same patterns.

3.6 Conclusion

While it is generally believed that incumbent parties can influence the behavior of voters with government spending, there are only a handful of papers that attempt to measure the magnitude of this causal effect with modern identification methods. For the most part, research on the subject has been limited due to the endogeneity between these two variables. As politicians have incentives to target spending where it will benefit them the most in the upcoming election, a simple OLS regression is likely to deliver a biased estimate of the causal effect of interest, and this bias may act in either direction depending on whether candidates target core or swing voters.

In this paper, we digitize the 1947 plan of the Interstate Highway System of the United States and use it in an instrumental variables (IV) framework to estimate the causal effect of opening one highway mile, during an election year, on votes received by the incumbent party. The results suggest that receiving one extra highway mile causes the share of votes to the gubernatorial incumbent party to increase by 0.6 percentage points. Additionally, we find no effect from opening a highway mile in a contiguous county, or one year before the election. We also find no effect of highway construction on vote shares for the incumbent president's party's presidential candidate.

We then use the IHS to think about connection between counties. The interstate reduces the cost of moving between counties, and we study the effect of this increased connectivity. We first test the yardstick competition theory and analyze what happens with the public goods supplied by the county administration. We find that the difference in public goods between adjacent counties is diminished after they are connected by the highway, and we interpret this as the effect of higher competition between local governments. Moreover, we test whether this is caused by a homogenization of output per capita with measures of population and property taxes and we find no effect, suggesting that the local public good homogenization is not due to county homogenization of socioeconomic variables.

We also find that highway connections coincide with homogenization of local expenditure from the state government (both in direct state-supplied public goods and in transfers from the state to counties). Existing theories are not well-placed to explain this phenomenon, as the decision on these public good supplies are driven by the central government and therefore not subject to local incumbent competition. Therefore, we build a novel model to analyze this pattern using a probabilistic voting model incorporating reference dependence in households' utility. In our model, households care not only about what they get of local public goods but about what other locations get. If agents realize that they are not getting equal resources then they punish the incumbent. In our model, constructing highways increases available information between counties, which prevents the incumbent from allocating resources to exploit the ideology distribution heterogeneity as much as before.

ACKNOWLEDGEMENTS

Chapter 3, in part is currently being prepared for submission for publication of the material. Leff Yaffe, Daniel; Nakab, Alejandro; Sandholtz, Wayne Aaron. I very gratefully acknowledge my coauthors Daniel Leff Yaffe and Alejandro Nakab, who have taught me much.

References

Abdulkadiroğlu, A., Angrist, J. D., Hull, P. D., & Pathak, P. A. (2016, July). Charters without lotteries: Testing takeovers in New Orleans and Boston. American Economic Review, 106(7), 1878-1920.

Acemoglu, D. (2008). Oligarchic versus democratic societies. Journal of the European Economic Association, 6(1), 1–44. doi: 10.1162/JEEA.2008.6.1.1

Acemoglu, D. (2010, aug). Theory, General Equilibrium, and Political Economy in Development Economics. Journal of Economic Perspectives, 24(3), 17–32. Retrieved from http://pubs.aeaweb.org/doi/10.1257/jep.24.3.17 doi: 10.1257/jep.24.3.17

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy Does Cause Growth. Journal of Political Economy, 127(1). doi: 10.2139/ssrn.2411791

Agor, W. H. (1989). Intuition & strategic planning: How organizations can make productive decisions. The Futurist, 23(6), 20.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the economics of organizations. Journal of Economic Perspectives, 19(1), 9-32.

Andrabi, T., Bau, N., Das, J., & Khwaja, A. I. (2010). Are bad public schools public "bads"? Test scores and civic values in public and private schools. (Mimeo)

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. American Economic Review, 107(6), 1535-63.

Andrabi, T., Das, J., Khwaja, A. I., & Zajonc, T. (2011). Do value-added estimates add value? Accounting for learning dynamics. American Economic Journal: Applied Economics, 3(3), 29–54.

Aslam, M., Rawal, S., & Saeed, S. (2017). Public-private partnerships in education in developing countries: A rigorous review of the evidence. Ark Education Partnerships Group.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. American Economic Review, 103(6), 2121–2168.

doi: 10.1257/aer.103.6.2121

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., & Heil, A. (2020). Why Do People Stay Poor?

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. The Quarterly Journal of Economics, 122(3), 1235–1264.

Banerjee, A., Duflo, E., & Glennerster, R. (2008). Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system. Journal of the European Economic Association, 6(2-3), 487–500.

Banerjee, A., Hanna, R., Kyle, J., Olken, B. A., & Sumarto, S. (2019). Private outsourcing and competition: Subsidized food distribution in Indonesia. Journal of Political Economy, 127(1), 101-137.

Barrera-Osorio, F., Blakeslee, D. S., Hoover, M., Linden, L., Raju, D., & Ryan, S. P. (2017, September). Delivering education to the underserved through a public-private partnership program in Pakistan (Working Paper No. 23870). National Bureau of Economic Research.

Bartik, T. J. (1991). Who Benefits from State and Local Economic Development Policies? Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. doi: 10.4135/9781483330006.n10

Baum-Snow, N. (2007). Did highways cause suburbanization? Quarterly Journal of Economics, 122(2), 775–805. doi: 10.1162/qjec.122.2.775

Baysah, A. M., Jr. (2016, Nov). Liberia: Police charge youth activist for sodomy. Retrieved from https://web.archive.org/web/20161103182507/https://allafrica.com/stories/201611020824.html

Bennedsen, M., Nielsen, K. M., Pérez-González, F., & Wolfenzon, D. (2007). Inside the family firm: The role of families in succession decisions and performance. The Quarterly Journal of Economics, 122(2), 647–691.

Bercoff, J. J., & Meloni, O. (2009, jan). Federal budget allocation in an emergent democracy: Evidence from Argentina. Economics of Governance, 10(1), 65–83. Retrieved from http://link.springer.com/10.1007/s10101-008-0052-9 doi: 10.1007/s10101-008-0052-9

Besley, T., & Burgess, R. (2002). THE POLITICAL ECONOMY OF GOVERNMENT RESPONSIVENESS: THEORY AND EVIDENCE FROM INDIA. The Quarterly Journal of Economics. Retrieved from http://econ.duke.edu/{~}psarcidi/lunchf08/besburgess.pdf

Besley, T., & Case, A. (1995). Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition. The American Economic Review, 85(1), 25–45.

Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. The American economic review, 95(3), 616–636.

Betts, J. R., & Tang, Y. E. (2014). A meta-analysis of the literature on the effect of charter schools on student achievement (Tech. Rep.). Society for Research on Educational Effectiveness.

Bloom, E., Bhushan, I., Clingingsmith, D., Hong, R., King, E., Kremer, M., … Schwartz, J. B. (2007). Contracting for health: evidence from Cambodia. (mimeo)

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? Evidence from India. The Quarterly Journal of Economics, 128(1), 1–51.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools? The Economic Journal, 125(584), 647–674.

Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2014). Does working from home work? Evidence from a Chinese experiment. The Quarterly Journal of Economics, 130(1), 165–218.

Bodenstein, M., & Ursprung, H. W. (2005). Political yardstick competition, economic integration, and constitutional choice in a federation: A numerical analysis of a contest success function model. Public Choice, 124(3-4), 329–352. doi: 10.1007/s11127-005-2051-5

Bold, T., Kimenyi, M., Mwabu, G., Ng, A., & Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. Journal of Public Economics, 168, 1–20. Retrieved from https://doi.org/10.1016/j.jpubeco.2018.08.007 doi: 10.1016/j.jpubeco.2018.08.007

Brault, M. (2011). School-aged children with disabilities in U.S. metropolitan statistical areas: 2010. American community survey briefs (Tech. Rep.). ACSBR/10-12. US Census Bureau.

Bridge International Academies. (2017). Bridge International Academies' written evidence to the International Development Committee Inquiry on DFID's work on education: Leaving no one behind? (Tech. Rep.). House of Commons, International Development Committee.

Bruns, B., & Luque, J. (2014). Great teachers: How to raise student learning in Latin America and the Caribbean. World Bank Publications.

Bursztyn, L. (2016, oct). Poverty and the Political Economy of Public Education Spending: Evidence From Brazil. Journal of the European Economic Association, 14(5), 1101–1128. Retrieved from http://doi.wiley.com/10.1111/jeea.12174 doi: 10.1111/jeea.12174

Butler, D. M., & Nickerson, D. W. (2011, aug). Can Learning Constituency Opinion Affect How Legislators Vote? Results from a Field Experiment. Quarterly Journal of Political Science, 6(1), 55–83. Retrieved from http://www.nowpublishers.com/article/Details/QJPS-11019 doi: 10.1561/100.00011019

Cabral, S., Lazzarini, S. G., & de Azevedo, P. F. (2013). Private entrepreneurs in public services: A longitudinal examination of outsourcing and statization of prisons. Strategic Entrepreneurship Journal, 7(1), 6–25.

Chabrier, J., Cohodes, S., & Oreopoulos, P. (2016). What can we learn from charter school lotteries? The Journal of Economic Perspectives, 30(3), 57–84.

Crawfurd, L. (2017). School management and public-private partnerships in uganda. Journal of African Economies, 26(5), 539-560.

Cremata, E., Davis, D., Dickey, K., Lawyer, K., Negassi, Y., Raymond, M., & Woodworth, J. L. (2013). National charter school study (Tech. Rep.). Center for Research on Education Outcomes, Stanford University.

Cruz, C., Keefer, P., Labonne, J., & Trebbi, F. (2018). Making Policies Matter: Voter Responses to Campaign Promises.

Cruz, C., & Schneider, C. J. (2017, apr). Foreign Aid and Undeserved Credit Claiming. American Journal of Political Science, 61(2), 396–408. Retrieved from `http://doi.wiley.com/10.1111/ajps.12285` doi: 10.1111/ajps.12285

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School inputs, household substitution, and test scores. American Economic Journal: Applied Economics, 5(2), 29–57.

Das, J., & Zajonc, T. (2010). India shining and bharat drowning: Comparing two indian states to the worldwide distribution in mathematics achievement. Journal of Development Economics, 92(2), 175 - 187.

De La O, A. L. (2013). Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico. American Journal of Political Science, 57(1), 1–14. doi: 10.1111/j.1540-5907.2012.00617.x

De Figueiredo, M. F. P., Hidalgo, F. D., & Kasahara, Y. (2011). When Do Voters Punish Corrupt Politicians? Experimental Evidence from Brazil *. Retrieved from `https://law.utexas.edu/wp-content/uploads/sites/25/figueiredo{_}when{_}do{_}voters{_}punish.pdf`

de Ree, J., Muralidharan, K., Pradhan, M., & Rogers, H. (2018). Double for nothing? Experimental evidence on an unconditional teacher salary increase in indonesia. The Quarterly Journal of Economics, 133(2), 993-1039.

Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2013). Comparative cost-effectiveness analysis to inform policy in developing countries: a general framework with applications for education. Education Policy in Developing Countries, 285–338.

Dias, M., & Ferraz, C. (2019). Voting for Quality ? The Impact of School Quality Information on Electoral Outcomes.

DIVA-GIS. (2016). Liberia administrative areas. Retrieved 06/01/2016, from `http://biogeo.ucdavis.edu/data/diva/adm/LBR_adm.zip`

Dixit, A., & Londregan, J. (1996). The determinants of success of special interests in redistributive politics. Journal of Politics, 58(4), 1132–1155. doi: 10.2307/2960152

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. American Economic Review, 108(4-5), 899–934. doi: 10.1257/aer.20101199

Donaldson, D., & Hornbeck, R. (2016). RAILROADS AND AMERICAN ECONOMIC GROWTH: A "MARKET ACCESS" APPROACH. The Quarterly Journal of Economics, 131(2), 799–858. doi: 10.1093/qje/qjw002.Advance

Duflo, E., Dupas, P., & Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. Journal of Public Economics, 123, 92–110.

Duggan, M. (2004). Does contracting out increase the efficiency of government programs? Evidence from Medicaid HMOs. Journal of Public Economics, 88(12), 2549 - 2572.

Dunning, T., Grossman, G., Humphreys, M., Hyde, S. D., Mcintosh, C., Nellis, G., … Lierl, M. (2019). Voter information campaigns and political accountability : Cumulative findings from a preregistered meta-analysis of coordinated trials. Science Advances, 1–11.

Evans, D., & Yuan, F. (2017). The economic returns to interventions that increase learning. (mimeo)

Ferejohn, J. (1986). Incumbent performance and electoral control. Public Choice, 50, 5–25. Retrieved from `https://link.springer.com/content/pdf/10.1007/BF00124924.pdf`

Fernald, J. G. (1999). Roads to prosperity? Assessing the link between public capital and productivity. American Economic Review, 89(3), 619–638. doi: 10.1257/aer.89.3.619

Ferraz, C., & Finan, F. (2008, may). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. Quarterly Journal of Economics, 123(2), 703–745. Retrieved from `https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.2008.123.2.703` doi: 10.1162/qjec.2008.123.2.703

Fryer, R. G. J. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. The Quarterly Journal of Economics, 129(3), 1355–1407.

Galiani, S., Gertler, P., & Schargrodsky, E. (2005). Water for life: The impact of the privatization of water services on child mortality. Journal of political economy, 113(1), 83–120.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis. Chapman & Hall/CRC Boca Raton, FL, USA.

Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher incentives. American Economic Journal: Applied Economics, 2(3), 205–227.

Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. American Economic Journal: Applied Economics, 1(1), 112-35.

Golden, M., & Min, B. (2013, may). Distributive Politics Around the World. Annual Review of Political Science, 16(1), 73–99. Retrieved from `http://www.annualreviews.org/doi/10.1146/annurev-polisci-052209-121553` doi: 10.1146/annurev-polisci-052209-121553

Goldsmith-Pinkham, P., Sorkin, I., & Swift, J. (2020). Bartik Instruments: What, When, Why, and HowGoldsmith-Pinkham. American Economic Review, 110(January), 2586–2624.

Grossman, G. M., & Helpman, E. (1996). Electoral Competition and Special Interest Politics. The Review of Economic Studies, 63(2), 265. doi: 10.2307/2297852

Guiteras, R. P., & Mobarak, A. M. (2015). Does development aid undermine political accountability? Leader and Constituent Responses to a Large-Scale Intervention. NBER Working Paper, NBER Working Paper 21434. Retrieved from `http://faculty.som.yale.edu/MushfiqMobarak/papers/Sanitation{_}politics.pdf` doi: 10.3386/w21434

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Disruption versus tiebout improvement: The costs and benefits of switching schools. Journal of public Economics, 88(9), 1721–1746.

Hanushek, E. A., & Woessmann, L. (2016, sep). School resources and student achievement: A review of cross-country economic research. In Methodology of educational measurement and assessment (pp. 149–171). Springer International Publishing.

Harding, R. (2015). Attribution and accountability: Voting for roads in ghana. World Politics, 67(4), 656–689.

Hart, O., Shleifer, A., & Vishny, R. W. (1997). The proper scope of government: theory and an application to prisons. The Quarterly Journal of Economics, 112(4), 1127–1161.

Hjort, J., Moreira, D., Rao, G., & Santini, J. F. (2019). How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities. NBER Working Paper Series.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. Journal of Law, Economics, & Organization, 7, 24–52.

Hsieh, C.-T., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. Journal of public Economics, 90(8), 1477–1503.

Jones, M. P., Meloni, O., & Tommasi, M. (2012, jul). Voters as Fiscal Liberals: Incentives and Accountability in Federal Systems. Economics and Politics, 24(2), 135–156. Retrieved from http://doi.wiley.com/10.1111/j.1468-0343.2012.00395.x doi: 10.1111/j.1468-0343.2012.00395.x

Kahneman, D., Tversky, A., Kahneman, B. Y. D., & Tversky, A. (1979). Prospect Theory : An Analysis of Decision under Risk. Source: Econometrica, 47(2), 263–292. Retrieved from http://www.jstor.org/stable/1914185{%}5Cnhttp://www.jstor.org/{%}5Cnhttp://www.jstor.org/action/showPublisher?publisherCode=econosoc.{%}5Cnhttp://www.jstor.org

Katsimi, M., & Sarantides, V. (2012, apr). Do elections affect the composition of fiscal policy in developed, established democracies? Public Choice, 151(1-2), 325–362. Retrieved from http://link.springer.com/10.1007/s11127-010-9749-8 doi: 10.1007/s11127-010-9749-8

Key, V. O. (1966). The responsible electorate. Belknap Press of Harvard University Press.

Kőszegi, B., & Rabin, M. (2006). A Model of Reference-Dependent Preferences. Quarterly Journal of Economics, 121(4), 1133–1165.

King, S., Korda, M., Nordstrum, L., & Edwards, S. (2015). Liberia teacher training program: Endline assessment of the impact of early grade reading and mathematics interventions (Tech. Rep.). RTI International.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics, 133(1), 97–126. doi: 10.1016/j.jeconom.2005.02.011

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. Science, 340(6130), 297–300.

Kwauk, C., & Robinson, J. P. (2016). Bridge International Academies: Delivering quality education at a low cost in Kenya, Nigeria, and Uganda (Tech. Rep.). The Brookings Institution. Retrieved 09/08/2017, from http://www.bridgeinternationalacademies.com/wp-content/uploads/2016/09/Brookings-Millions-Learning-case-study.pdf

Larreguy, H., Montiel Olea, C. E., & Querubin, P. (2017). Political Brokers: Partisans or Agents? Evidence from the Mexican Teachers' Union. American Journal of Political Science, 61(4), 877–891. doi: 10.1111/ajps.12322

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. The Review of Economic Studies, 76(3), 1071–1102.

Leff-Yaffe, D. (2020). The Interstate Multiplier.

Lemos, R., & Scur, D. (2016). Developing management: An expanded evaluation tool for developing countries. (mimeo)

Levitt, S. D., & Snyder, J. M. (1997, feb). The Impact of Federal Spending on House Election Outcomes. Journal of Political Economy, 105(1), 30–53. Retrieved from http://www.journals.uchicago.edu/doi/10.1086/262064 doi: 10.1086/262064

Liaqat, A., Callen, M., Cheema, A., Khan, A., Naseer, F., Shapiro, J. N., ... San Diego, U. (2018). Political Connections and Vote Choice Evidence from Pakistan.

Liberia Institute of Statistics and Geo-Information Services. (2014). Liberia demographic and health survey 2013. Liberia Institute of Statistics and Geo-Information Services.

Liberia Institute of Statistics and Geo-Information Services. (2016). Liberia - household income and expenditure survey 2014-2015. Liberia Institute of Statistics and Geo-Information Services.

Litschig, S., & Morrison, K. M. (2013, oct). The Impact of Intergovernmental Transfers on Education Outcomes and Poverty Reduction. American Economic Journal: Applied Economics, 5(4), 206–240. Retrieved from http://pubs.aeaweb.org/doi/10.1257/app.5.4.206 doi: 10.1257/app.5.4.206

Loevinsohn, B., & Harding, A. (2005). Buying results? Contracting for health service delivery in developing countries. The Lancet, 366(9486), 676–681.

Lucas, A. M., & Mbiti, I. M. (2012). Access, sorting, and achievement: the short-run effects of free primary education in Kenya. American Economic Journal: Applied Economics, 4(4), 226–253.

MacRae, C. D. (1977). A Political Model of the Business Cycle. Journal of Political Economy, 85(2), 239–263.

Majumdar, S., & Mukand, S. W. (2004, aug). Policy Gambles. American Economic Review, 94(4), 1207–1222. Retrieved from http://pubs.aeaweb.org/doi/10.1257/0002828042002624 doi: 10.1257/0002828042002624

Manacorda, M., Miguel, E., & Vigorito, A. (2011, jul). Government Transfers and Political Support. American Economic Journal: Applied Economics, 3(3), 1–28. Retrieved from http://pubs.aeaweb.org/doi/10.1257/app.3.3.1 doi: 10.1257/app.3.3.1

Mani, A., & Mukand, S. (2007). Democracy, visibility and public good provision. Journal of

Development Economics, 83(2), 506–529. doi: 10.1016/j.jdeveco.2005.06.008

Marx, B. (2018). Elections as Incentives : Project Completion and Visibility in African Politics. (October).

May, S. (2017). Oral evidence: DFID's work on education: Leaving no one behind?, HC 639 (Tech. Rep.). House of Commons, International Development Committee.

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019, 04). Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania*. The Quarterly Journal of Economics, 134(3), 1627-1673. doi: 10.1093/qje/qjz010

Meager, R. (2016). Aggregating distributional treatment effects: A bayesian hierarchical analysis of the microcredit literature (Working paper). London School of Economics and Political Science.

Millennium Challenge Corporation. (2013). Liberia constraint analysis. Retrieved 11/03/2019, from http://www.liberianembassyus.org/uploads/PDFFiles/LIBERIA%20CONSTRAINTS%20ANALYSIS_FINAL%20VERSION.pdf

Ministry of Education - Republic of Liberia. (2015-2016). Education Management Information System (EMIS). http://moe-liberia.org/emis-data/.

Ministry of Education - Republic of Liberia. (2016a). Liberia education statistics report 2015-2106.

Ministry of Education - Republic of Liberia. (2016b). Memorandum of understanding between Ministry of Education, Government of Liberia and Bridge International Academies. Retrieved 6/08/2017, from www.theperspective.org/2016/ppp_mou.pdf

Ministry of Education - Republic of Liberia. (2017a). Getting to best education sector plan, 2017-2021.

Ministry of Education - Republic of Liberia. (2017b). PSL school allocation: Decision points. Retrieved 28/07/2017, from http://moe.gov.lr/wp-content/uploads/2017/06/Allocation-final.pdf

Muralidharan, K., Singh, A., & Ganimian, A. J. (2016). Disrupting education? Experimental evidence on technology-aided instruction in India (Tech. Rep.). National Bureau of Economic Research.

Muralidharan, K., & Sundararaman, V. (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in India. The Quarterly Journal of Economics, 130(3), 1011.

Nadeau, A., & Blais, R. (1992). The Electoral Budget Cycle. Public Choice, 74(4), 389–403.

Nall, C. (2015). The political consequences of spatial policies: How interstate highways facilitated geographic polarization. Journal of Politics, 77(2), 394–406. doi: 10.1086/679597

Nall, C. (2018). The road to inequality: how the federal highway program polarized america and undermined cities. Cambridge University Press.

Nordhaus, W. D. (1975). The Political Business Cycle. The Review of Economic Studies, 42(2), 169–190.

Patrinos, H. A., Barrera-Osorio, F., & Guáqueta, J. (2009). The role and impact of public-private partnerships in education. World Bank Publications.

Peltzman, S. (1992, may). Voters as Fiscal Conservatives. The Quarterly Journal of Economics, 107(2), 327–361. Retrieved from http://qje.oxfordjournals.org/content/107/2/327.abstract doi: 10.2307/2118475

Persson, T., & Tabellini, G. E. (2002). Political economics: explaining economic policy.

Pierskalla, J. H., & Sacks, A. (2019). Personnel Politics : Elections , Clientelistic Competition and Teacher Hiring in Indonesia. British Journal of Political Science. doi: 10.1017/S0007123418000601

Piper, B., & Korda, M. (2011). Egra plus: Liberia. program evaluation report (Tech. Rep.). RTI International.

Potrafke, N. (2010). The growth of public health expenditures in OECD countries: Do government ideology and electoral motives matter? Journal of Health Economics, 29(6), 797–810. doi: 10.1016/j.jhealeco.2010.07.008

Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Gaduh, A., Alisjahbana, A., & Artha, R. P. (2014, April). Improving educational quality through enhancing community participation: Results from a randomized field experiment in indonesia. American Economic Journal: Applied Economics, 6(2), 105-26. Retrieved from http://www.aeaweb.org/articles?id=10.1257/app.6.2.105 doi: 10.1257/app.6.2.105

Pritchett, L., & Sandefur, J. (2014, jan). Context Matters for Size: Why External Validity Claims and Development Practice do not Mix. Journal of Globalization and Development, 4(2). Retrieved from https://www.degruyter.com/view/j/jgd.2013.4.issue-2/jgd-2014-0004/jgd-2014-0004.xml doi: 10.1515/jgd-2014-0004

Rogoff, K. (1990). Equilibrium Political Budget Cycles. American Economic Review, 80(1), 21–36. Retrieved from https://ideas.repec.org/a/aea/aecrev/v80y1990i1p21-36.html

Rogoff, K., & Sibert, A. (1988). Elections and Macroeconomic Policy Cycles. The Review of Economic Studies, 55(1), 1. doi: 10.2307/2297526

Romero, M., Sandefur, J., & Sandholtz, W. (2017). Partnership schools for liberia (psl) program evaluation. AEA RCT Registry. doi: https://doi.org/10.1257/rct.1501-7.0

Romero, M., Sandefur, J., & Sandholtz, W. (2018). Partnership Schools for Liberia. Harvard Dataverse. Retrieved from https://doi.org/10.7910/DVN/5OPIYU doi: 10.7910/DVN/5OPIYU

Romero, M., Sandefur, J., & Sandholtz, W. A. (2019). Outsourcing Education: Experimental Evidence from Liberia. American Economic Review. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.20181478{&}{&}from=f doi: 10.1257/AER.20181478

Rubin, D. B. (1981). Estimation in parallel randomized experiments. Journal of educational and behavioral statistics, 6(4), 377–401.

Sabarwal, S., Evans, D. K., & Marshak, A. (2014). The permanent input hypothesis : the case of textbooks and (no) student learning in Sierra Leone (Policy Research Working Paper Series No. 7021). The World Bank.

Salmon, P. (1987). Decentralization as an incentive scheme (Tech. Rep.).

Samuels, D. J. (2002). Pork Barreling Is Not Credit Claiming or Advertising : Campaign Finance and the Sources of the Personal Vote in Brazil. Journal of Politics, 64(3), 843–863.

Sandholtz, W. A. (2019). Do voters reward service delivery? Experimental evidence from Liberia. Retrieved from https://www.socialscienceregistry.org/trials/2506.

Schermerhorn, J., Osborn, R., Uhl-Bien, M., & Hunt, J. (2011). Organizational behavior. Wiley.

Shi, M., & Svensson, J. (2006). Political budget cycles: Do they differ across countries and why? Journal of Public Economics, 90(8-9), 1367–1389. doi: 10.1016/j.jpubeco.2005.09.009

Singh, A. (2015). Private school effects in urban and rural india: Panel estimates at primary and secondary school ages. Journal of Development Economics, 113, 16–32.

Singh, A. (2016). Learning more with every year: School year productivity and international learning divergence. (Mimeo)

Snyder, J. M., & Strömberg, D. (2010). Press coverage and political accountability. Journal of Political Economy, 118(2), 355–408. doi: 10.1086/652903

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. Econometrica, 65(3), 557. doi: 10.2307/2171753

Stallings, J. A., Knight, S. L., & Markham, D. (2014). Using the stallings observation system to investigate time on task in four countries (Tech. Rep.). World Bank.

Stokes, S. C. (2005, aug). Perverse accountability: A formal model of machine politics with evidence from Argentina. American Political Science Review, 99(3), 315–325. Retrieved from http://www.journals.cambridge.org/abstract{_}S0003055405051683 doi: 10.1017/S0003055405051683

Stromberg, D. (2004, feb). Radio's Impact on Public Spending. The Quarterly Journal of Economics, 119(1), 189–221. Retrieved from https://academic.oup.com/qje/article-lookup/doi/10.1162/003355304772839560 doi: 10.1162/003355304772839560

Tuttle, C. C., Gleason, P., & Clark, M. (2012). Using lotteries to evaluate schools of choice: Evidence from a national study of charter schools. Economics of Education Review, 31(2), 237–253.

UNESCO. (2016). Global monitoring report 2016 (Tech. Rep.). United Nations.

UNICEF. (2013). The state of the world's children: Children with disabilities (Tech. Rep.). United Nations.

USAID. (2017). Request for proposals - SOL-669-17-000004, Read Liberia. Retrieved 6/08/2017, from https://www.fbo.gov/index?s=opportunity&mode=form&id=e53cb285301f7014f415ce91b14049a3&tab=core&tabmode=list&=

Useem, B., & Goldstone, J. A. (2002). Forging social order and its breakdown: Riot and reform in U.S. prisons. American Sociological Review, 67(4), 499-525.

van der Linden, W. J. (2017). Handbook of item response theory. CRC Press.

Vicente, P. C., & Wantchekon, L. (2009, jun). Clientelism and vote buying: lessons from field experiments in African elections. Oxford Review of Economic Policy, 25(2), 292–305. Retrieved from http://oxrep.oxfordjournals.org/cgi/doi/10.1093/oxrep/grp018 doi: 10.1093/oxrep/grp018

Voigtllnder, N., & Voth, H.-J. (2018). Highway to Hitler. SSRN Electronic Journal. doi: 10.2139/ssrn.2684404

Wantchekon, L. (2003, apr). Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin. World Politics, 55(03), 399–422. Retrieved from http://www.journals.cambridge.org/abstract{_}S0043887100003798 doi: 10.1353/wp.2003.0018

Werner, G. K. (2017). Liberia has to work with international private school companies if we want to protect our children's future. Quartz Africa. Retrieved 20/07/2017, from https://qz.com/876708/why-liberia-is-working-with-bridge-international-brac-and-rising-academies-by-education-minister-george-werner/

Woodworth, J. L., Raymond, M., Han, C., Negassi, Y., Richardson, W. P., & Snow, W. (2017). Charter management organizations (Tech. Rep.). Center for Research on Education Outcomes, Stanford University.

World Bank. (2007). A decade of action in transport an evaluation of world bank assistance to the transport sector, 1995-2005. Retrieved from http://documents.worldbank.org/curated/en/808261468139200365/pdf/396600Decade0of0transport0IEGWB01PUBLIC1.pdf doi: 10.1596/978-0-8213-7003-2

World Bank. (2013). Net ODA Received (% Of GDP). Retrieved 01/04/2019, from https://datacatalog.worldbank.org/net-oda-received-gdp

World Bank. (2014). Life expectancy. (data retrieved from World Development Indicators, http://data.worldbank.org/indicator/SE.PRM.NENR?locations=LR)

World Bank. (2015a). Conducting classroom observations: analyzing classrooms dynamics and instructional time, using the stallings' classroom snapshot'observation system. user guide (Tech. Rep.). World Bank Group.

World Bank. (2015b). World bank group support to public-private partnerships: Lessons from experience in client countries, FY02-12. World Bank.

World Bank. (2016). Deposit interest rate (%). (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/FR.INR.DPST?locations=LR)

World Bank. (2017). GDP per capita (current US$). (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/NY.GDP.PCAP.CD)

Young, F. (2018, Oct). Unprotected. Retrieved from https://features.propublica.org/liberia/unprotected-more-than-me-katie-meyler-liberia-sexual-exploitation/

Zhang, H. (2014). The mirage of elite schools: evidence from lottery-based school admissions in China. (Mimeo)

Zimmermann, L. (2018). The dynamic electoral returns of a large anti-poverty program.