Can Outsourcing Improve Liberia's Schools?

Preliminary Results from Year One of a Three-Year Randomized Evaluation of Partnership Schools for Liberia

Mauricio Romero, Justin Sandefur, and Wayne Aaron Sandholtz

Abstract

After one year, public schools managed by private contractors in Liberia raised student learning by 60 percent, compared to standard public schools. But costs were high, performance varied across contractors, and contracts authorized the largest contractor to push excess pupils and under-performing teachers onto other government schools.

Keywords: Public-Private Partnership; Randomized Controlled Trial; School Management

JEL Codes: I25, I28, C93, L32, L33

Online appendix: https://www.cgdev.org/sites/default/files/partnership-schools-for-liberia-online-appendix.pdf

Center for Global Development
www.cgdev.org

Working Paper 462
September 2017

Can Outsourcing Improve Liberia's Schools? Preliminary Results from Year One of a Three-Year Randomized Evaluation of Partnership Schools for Liberia

Mauricio Romero
University of California, San Diego

Justin Sandefur
Center for Global Development

Wayne Aaron Sandholtz
University of California, San Diego

We are grateful to the Minister of Education, George K. Werner, Deputy Minister Romelle Horton, Binta Massaquoi, Nisha Makan, and the PSL team, as well as Susannah Hares, Robin Horn, and Joe Collins from Ark EPG for their commitment throughout this project to ensuring a rigorous and transparent evaluation. Thanks to Arja Dayal, Dackermue Dolo, and their team at Innovations for Poverty Action who led the data collection. Avi Ahuja and Dev Patel provided excellent research assistance. The design and analysis benefited from comments and suggestions from Prashant Bharadwaj, Jeffrey Clemens, Joe Collins, Mitch Downey, Dean Karlan, Gordon McCord, Craig McIntosh, Owen Ozier, Lant Pritchett, Santiago Saavedra, and seminar participants at the Center for Global Development and UC San Diego. We're grateful to Michael Kremer, Karthik Muralidharan, and Pauline Rose who constituted the evaluation's academic oversight committee. Committee members had the opportunity to make both private and public comments on the report; all gave private comments, and none sent public comments. A lack of public comment by any individual does not imply endorsement of our conclusions, and all remaining errors are ours alone. The trial registry and pre-analysis plan for the evaluation are available at: https://www.socialscienceregistry.org/trials/1501. IRB approval was received from IPA (protocol # 14227) and the University of Liberia (protocol # 17-04-39) prior to any data collection. UCSD IRB approval (protocol # 161605S) was received after the baseline but before any other activities were undertaken. The evaluation was supported by the UBS Optimus Foundation and Aestus Trust. Romero acknowledges financial support from the Central Bank of Colombia through the Lauchlin Currie scholarship. The views expressed here are ours, and not those of the Ministry of Education of Liberia or our funders

The Center for Global Development is grateful for contributions from the Research on Improving Systems of Education (RISE) program and the UBS Optimus Foundation in support of this work.

Mauricio Romero, Justin Sandefur, and Wayne Aaron Sandholtz. 2017. "Can Outsourcing Improve Liberia's Schools? Preliminary Results from Year One of a Three-Year Randomized Evaluation of Partnership Schools for Liberia." CGD Working Paper 462. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/partnership-schools-for-liberia

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

Executive summary

After one year, public schools managed by private contractors in Liberia raised student learning by 60%, compared to standard public schools. But costs were high, performance varied across contractors, and contracts authorized the largest contractor to push excess pupils and under-performing teachers onto other government schools.

"Partnership Schools" are free, public schools managed by private contractors

- **Liberia's education system lags behind most of the world in both access and quality.** Net primary enrollment was only 38% in 2015, and in 2013, among adult women who finished elementary school, only 25% could read a complete sentence.

- **Under the new Partnership Schools for Liberia (PSL) program, the Liberian government delegated management of 93 public schools to eight contractors.** Teachers in PSL schools remained on government payroll; schools remained free to students and the property of the government; and contractors were prohibited from screening students based on ability or other characteristics.

- **In addition to new management, PSL brought extra resources.** While the government runs ordinary public schools on a budget of approximately $50 (USD) per pupil, PSL contractors received an additional $50 per pupil, as the total of $100 was deemed a realistic medium-term goal for public expenditure on primary education nationwide. While teachers are in short supply in Liberia's public schools, the Ministry of Education made special staffing arrangements for PSL.

- **The evaluation randomly assigned existing government schools to become PSL schools**. Because assignment to the PSL and comparison groups was random, differences between the two groups can be attributed to the program. Schools were randomized *after* contractors agreed on a school list, and students in the sample were selected from the enrollment logs of the school year *before* contractors arrived. Therefore the results are not biased by contractors selecting schools or rejecting students.

On average, partnership schools improved teaching and learning

- **Students in partnership schools scored 0.18 standard deviations higher in English and 0.18 standard deviations higher in mathematics** than students in regular public schools. While starting from a very low level by international standards, this is the equivalent of 0.56 additional years of schooling for English and 0.66 additional years of schooling for math.

- **The program increased teachers' quality of instruction.** Teachers in PSL schools were 20 percentage points more likely to be in school during a random spot check (from a base of 40% in control schools) and 16 percentage points more likely to be engaged in instruction during class time (from a base of 32% in control schools). This holds even after controlling for changes in the composition of teachers.

- **Students in partnership schools spent twice as much time learning each week**, when taking into account reduced absenteeism, increased time-on-task, and longer school days in PSL schools.

Costs were high, in terms of government staffing and private subsidies

- **Budget estimates for some contractors' in year 1 exceeded the program's long-term target.** Rather than $50 per pupil, contractors' *ex ante* budgets ranged from $57 for Youth Movement for Collective Action to $1,050 for Bridge International Academies (later revised to $663). Learning gains varied widely across contractors, and higher costs do not necessarily correlate with higher learning gains.

- **The government assigned PSL schools 37% more teachers than non-PSL schools, including first pick of better-trained, new graduates.** In the short term, without a significant increase in the supply of trained teachers, the staffing advantages given to PSL appear unsustainable at a larger scale.

Contracts authorized the largest contractor to push excess pupils and under-performing teachers onto other government schools

- **Overall, enrollment levels did not change. But there are signs that some children were turned away from their school when PSL arrived.** PSL contracts made provisions for contractors to cap class sizes. Classes that with enrollment below PSL's class-size caps before the program arrived saw increases in enrollment. But about 30% of students were in classes above PSL's caps, and in those cases enrollment fell by 20 students per grade (p-value .032). It appears most of these students were absorbed into other schools. This issue was mostly restricted to Bridge International Academies.

- **The same contractor also dismissed half of incumbent public teachers in its schools.** In theory, these teachers are still paid by the government and may be working in other public schools or collecting pay without working. Although weeding out poorly performing teachers is important, a reshuffling of teachers is unlikely to raise average performance in the system as a whole.

The program has not been tested in average Liberian schools. In the first year, the program was implemented (and evaluated) within a list of eligible schools agreed by contractors; these schools had higher staffing levels and better infrastructure and were located closer to roads than average Liberian schools.

Clear, uniform procurement rules might better align contractors' incentives with the public interest. Six of the eight contractors were contracted through an open, competitive bidding process. One contractor (Stella Maris) did not complete contracting, did little work, and produced low learning gains. Another (Bridge International Academies) was selected outside the competitive process, produced strong learning gains, but removed the majority of teachers and displaced some students. Revised contracts and competitive selection of contractors based on performance might mitigate these issues.

There is solid evidence of positive effects for Liberian children during the first year of PSL. But the program has yet to demonstrate it can work in average Liberian schools, with sustainable budgets and staffing levels, and without negative side-effects on other schools. Before making decisions about dramatically expanding the program, the remaining two years of the three-year pilot and evaluation could be used to test additional refinements and build up public sector capacity to hold contractors accountable.

Contents

List of Tables

List of Figures

1 Introduction

In September 2016, Liberia's Ministry of Education delegated management of ninety-three government primary schools to eight different private entities, ranging from local non-profit organizations to for-profit multinational companies. This initiative, known as "Partnership Schools for Liberia" (PSL), is patterned loosely on charter schools in the United States or academies in the United Kingdom. Schools are free to parents; selective admissions are prohibited; and teachers continue to draw salaries directly from the government. Private contractors receive a per-pupil subsidy to provide teacher training, school inputs, and to take over general school management. While the project has generated considerable controversy and international media attention, Minister of Education George K. Werner has encouraged critics to "judge us on the data—data on whether PSL schools deliver better learning outcomes for children" (Werner, 2017).

In addition to new management, the program also brought extra resources. While the government runs ordinary public schools on a budget of approximately \$50 (USD) per pupil, PSL schools received an additional \$50 on top of this, as the total of \$100 was deemed a realistic medium-term goal for public expenditure on primary education nationwide.[1] While teachers are in short supply in Liberia's public schools, the Ministry of Education made special staffing arrangements for PSL.

This report summarizes the findings from year one of a three-year randomized impact evaluation of PSL covering ninety-three schools, and measures the program's impacts, at the end of its first academic year, on enrollment and learning as well as a host of other student, parent, teacher, and school outcomes. Data were collected by independent survey teams managed by Innovations for Poverty Action (IPA) in September and October of 2016 for the baseline and May to June of 2017 for the midline results shown in this report. Treatment schools were randomly assigned to the PSL program from a list of one hundred and eighty-five eligible schools agreed upon by the Ministry and private contractors. Within both treatment and control schools, the survey teams sampled twenty pupils per school from the 2015/16 enrollment log, i.e., the year prior to the announcement and roll-out of the program. This sampling allows us to perform intention-to-treat analysis, ensuring that movement of pupils into or out of PSL schools (in response to the program) does not drive differences in test scores.

1.1 On average, partnership schools improved teaching and learning

Overall, we find that the PSL program increases learning outcomes. The effect on test scores of being randomly assigned to the PSL program after one academic year of treatment is $.18\sigma$ for English (p-value < 0.001) and $.18\sigma$ for math (p-value < 0.001).[2] To put these effect sizes in context, the average increase in test scores for each additional year of schooling in the control group is $.31\sigma$ in English and $.28\sigma$ in math. Thus, the treatment effect is equivalent to roughly 0.56 additional years of schooling for English ($.18\sigma/.31\sigma$) and 0.66 additional years of schooling for math ($.18\sigma/.28\sigma$). There is evidence that these gains do not reflect teaching to the test, as they are apparent in new questions administered only at the end of

[1]Because they were not subject to the same contracts, neither Bridge International Academies nor Stella Maris received the extra \$50 per pupil.

[2]Consistent with standard practice, we report many effect sizes in terms of standard deviations, notated with σ. Standard deviations are a measure of the dispersion of data points within a given dataset. An increase in test scores of 0.2σ is equivalent to pushing the average treated student from the 50th to the 57th percentile in the distribution of test scores.

the school year, and in conceptual questions with an entirely new format.[3] Similarly, we find no evidence that contractors engaged in student selection (though, as noted above, it would not affect our results on test scores); the probability of remaining in a treatment school is unrelated to age, gender, household wealth, or disability.

In unannounced spot checks, we find that teachers in PSL schools were 20 percentage points (p-value < 0.001) more likely to be in school (from a base of 40% in control schools). They were also 16 percentage points (p-value < 0.001) more likely to be engaging in either active or passive instruction during class time, and 27 percentage points (p-value < 0.001) less likely to be off-task (from a base of 32% and 59% respectively in control schools). Student attendance increased by 13 percentage points (from a base of 35%). Combining the effects of reduced student absenteeism, increased teacher time-on-task, and a longer school days in treatment schools (3.9 more hours a week of instructional time, p-value < 0.001), students in PSL schools spent roughly twice as long learning each week.[4]

Note that despite the positive treatment effect of the program, students in treatment schools are still far behind their international peers (see Appendix H), and teachers' time on task in PSL schools is still well below rates measured in middle-income countries (Bruns & Luque, 2014).

1.2 Costs were high, in terms of government staffing and private subsidies

Contractors vary considerably in terms of their total costs and cost structures. Rather than \$50 per student, per pupil ex-ante budgets ranged from a low of approximately \$57 per pupil for Youth Movement for Collective Action[5] to a maximum of \$1,050 per pupil in the case of Bridge International Academies (later revised down to \$663).[6] Note that these budgets include one-off start-up costs, recurring fixed costs, and variable costs per pupil. The long-run per pupil cost of a larger program might be considerably reduced. A more useful lens might be to identify the price point that induces private contractors to participate in PSL. At present, contractors have expressed interest in the program with an offer of \$50 subsidy per pupil, over and above the Ministry's \$50 expenditure per pupil in all schools (although several contractors continue to lobby for a price increase and in-kind support, and to fundraise from outside donors).

Using the optimistic long-term cost target of \$50, learning gains of $.19\sigma$ on average, and even 0.27σ for the best-performing contractors, appear expensive relative to the most cost-effective interventions in the academic literature (Kremer, Brannen, & Glennerster, 2013). In fairness, many education interventions have zero effect, or simply fail to measure either impacts or costs to allow for cost-effectiveness calculations (Evans & Popova, 2016). Furthermore, Liberia is a challenging environment, and both impact and cost parameters from other contexts may fail to replicate in this context.

The cost to the Liberian government of PSL came mostly through teacher salaries. PSL changed both the quantity and quality of teachers in public schools. Treatment schools have 2.6 more teachers (p-value

[3] We cannot rule out that contractors narrowed the curriculum and focused on English and mathematics (or conversely, that they generated learning gains in other subjects that we did not test).

[4] The scheduled instructional time increased from 16.5 to 20.4 hours per week; time-on-task during class time went up from 32% to 48%; and student attendance went from 35% to 48%. Hence, the effective teaching time in PSL schools was close to 10 (20.4×0.48) hours per week (compared to 5.3 in traditional public schools). Combining the effective teaching time with student attendance, the average student in PSL schools got 4.8 (10*0.48) hours per week of instructional time (compared to 1.9 in traditional public schools).

[5] Youth Movement for Collective Action began the evaluation as "Liberian Youth Network," or LIYONET. The group has now changed its name.

[6] Several caveats apply to the cost figures here, which are our own estimates based on contractors' self-reported budget data, and combine start-up costs, fixed costs, and variable costs. See Section 2.1.4 for more detail.

< 0.001) than schools in the control group (an increase of 37% from a base of 7). The additional teachers outweighed an enrollment increase of 19 students per school (p-value .24), to yield a net reduction in the pupil-teacher ratio of 6.9 (p-value < 0.001). While pupil-teacher ratios have not shown a robust relationship with learning outcomes in previous experiments in developing countries (Banerjee, Cole, Duflo, & Linden, 2007; Duflo, Dupas, & Kremer, 2015),[7] teacher quality appears to matter a great deal (Bruns & Luque, 2014; Buhl-Wiggers, Kerwin, Smith, & Thornton, 2017; Araujo, Carneiro, Cruz-Aguayo, & Schady, 2016). PSL contractors successfully lobbied the Ministry of Education to assign sought-after new graduates from teacher training institutes to PSL schools: average teacher age in PSL schools fell by 7.1 years (p-value < 0.001), and a measure of teachers' cognitive skills rose significantly ($.14\sigma$, p-value .018).[8] In the short term, without a significant increase in the supply of trained teachers, the staffing advantages given to PSL appear unsustainable at a larger scale.

1.3 Learning gains varied by contractor

The experimental evaluation was designed to study the impact of the PSL program at large, asking "What will a government like Liberia's achieve if it contracts out management of public schools to the private sector?" While this question dominates our analysis, there is also clear demand from the policy community to understand the relative performance of specific private contractors. This information serves two core policy functions: (a) providing the Ministry and donors with data to hold contractors accountable for results; and (b) providing the Ministry, donors, and contractors themselves with information about what worked well, and what did not.

We confront two fundamental obstacles in calculating contractor-specific impacts. First, contractors work in different counties and in schools with very different baseline conditions. While assignment to PSL overall is random, assignment to a specific contractor is not. We adjust for these baseline differences in a simple regression framework. Second, because randomization occurred at the school level and some contractors run only four or five treatment schools, the experiment is under-powered to estimate their effects. Nevertheless, ffrom a policy perspective, the randomized control trial (RCT) contains the only independent data on contractors' performance, and it is not optimal from a decision-making perspective to simply ignore this data. We take a Bayesian approach to this problem, estimating a hierarchical model along the lines proposed by Rubin (1981) and Gelman, Carlin, Stern, and Rubin (2014) to determine the best possible estimate of contractor-specific effects given small sample sizes. The net effect of the Bayesian adjustments is that the final estimate is an average of the overall effect and the contractor's individual effect, weighted according to the number of schools the contractor operates.

A key finding from this analysis is simply the existence of heterogeneity, or variance in contractors' effects. The variance of effects is larger than can be explained by chance and exists in learning in learning impacts, as well as in behaviors that might impose negative externalities on the broader education system. Heterogeneity is unsurprising but important: Merely contracting out school management by the Ministry of Education is not sufficient to generate consistent results, as the identity of the contractor appears to

[7]Note that while Angrist and Lavy (1999) had found a causal effect between class sizes and test scores using quasi-experimental evidence, a recent revisit of the issue using the same estimation strategy, in the same setting, found no effect of class sizes on achievement (Angrist, Lavy, Leder-Luis, & Shany, 2017).

[8]Once the Education Management Information System (EMIS) data for the 2016/2017 school year is released, we will reexamine this issue to study whether teachers who were fired were allocated to other public schools.

matter quite a lot.

Results on learning can be roughly grouped into three categories. In the first group, the Youth Movement for Collective Action (YMCA), Rising Academies, Bridge International Academies, and Street Child generated an increase in learning of 0.27σ across all subjects. In the second group, BRAC and More than Me produced an increase in learning of 0.15σ. In the third group, consisting of Omega and Stella Maris, estimated learning gains are on the order of 0.01σ, and indistinguishable from zero in both cases.

1.4 Contracts authorized the largest contractor to push excess pupils and underperforming teachers onto other government schools

We find that 74 percent of teachers in Bridge International Academies schools at baseline had been released,[9] contradicting the program's intent that contractors train and manage existing government teachers. Although weeding out bad teachers is important, a reshuffling of teachers is unlikely to raise average performance in the system as a whole.

A similar pattern is observed with student enrollment. Class-size caps were authorized by contracts, but were generally not enforced in control schools or by contractors other than Bridge International Academies. In schools and grades where baseline enrollment was above the theoretical cap for PSL (i.e., already oversubscribed schools, holding 30% of students at baseline), the program reduced enrollment by 20 pupils per class (p-value .032). This effect was driven by Bridge International Academies. Most of these students were absorbed by nearby traditional public schools.[10] In schools and grades where baseline enrollment was below the cap, enrollment increased by 4.8 pupils per class (p-value < 0.001), highlighting the importance of contracting details.

1.5 Policy challenges

In the first year, the program was implemented (and evaluated) within a list of eligible schools agreed by contractors; these schools had higher staffing levels, better infrastructure, and were closer to roads than average Liberian schools. It remains to be seen whether similar impacts can be replicated in more disadvantaged schools without raising costs.

Clear, uniform procurement rules might better align contractors' incentives with the public interest. Six of the eight contractors were contracted through an open, competitive bidding process. One contractor (Stella Maris) did not complete contracting, did little work, and produced low learning gains. Another (Bridge International Academies) was selected outside the competitive process, produced strong learning gains, but removed the majority of teachers and displaced some students. Revised contracts and competitive selection of contractors based on performance might mitigate these issues.

This is the first year of a three-year evaluation, but our preliminary results provide solid evidence of positive effects for Liberian children during the first year. Impacts may increase as contractors establish their operations. However, the program has yet to demonstrate it can work in average Liberian schools,

[9] The total number of teachers in Bridge schools at baseline was 236 according to EMIS data. Of these, 177 were on the government payroll.

[10] Note that our survey, which tracked a sub-sample of these students, suggests the majority of those who were excluded re-enrolled in other schools, rather than dropping out of school altogether. Once the EMIS data for the 2016/2017 school year is released, we will revisit this issue to study changes in total enrollment.

with sustainable budgets and staffing levels and without negative side-effects on other schools. The remaining two years of the three-year pilot and evaluation could be used to test further refinements before any significant expansion of the program. In addition, future survey rounds will seek to further unpack the principal factors underlying PSL's impacts. We review additional policy issues in Section 7.

2 Evaluation design

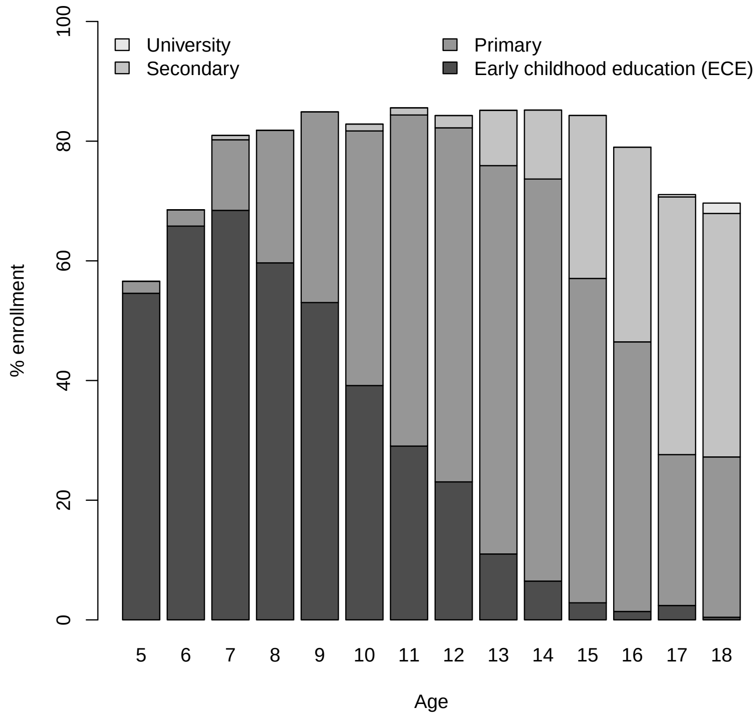
2.1 The program

2.1.1 Context

The PSL program breaks new ground in Liberia by delegating management of government employees to private contractors, but it is worth noting that a strong role for private actors—such as NGOs and USAID contractors—in providing school meals, teacher support services, and other assorted programs in government schools is the norm, not an innovation. Over the past decade, Liberia's basic education budget has been roughly $40 million per year (about 2-3% of GDP), while external donors contribute about $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 of 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 was tendering a $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 of this financing sytem 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 in preschool and primary attend private schools (Ministry of Education - Republic of Liberia, 2016a).

A second broad feature of Liberia's education system that is important for the PSL program is its performance: Not only are learning levels low, but simple 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 (The World Bank, 2014). Low *net* enrollment is partially explained by an extraordinary backlog of over-age children (see Figure 1), particularly in early childhood education, where the median student is eight years old (Liberia Institute of Statistics and Geo-Information Services, 2016). Learning levels are low: Only 25% of adult women who finish elementary school can read a complete sentence (Liberia Institute of Statistics and Geo-Information Services, 2014) (there is no information for men).

Figure 1: Enrollment by age



Note: Authors' calculations based on 2014 Household Income and Expenditures Survey.

2.1.2 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 contractors to run ninety-three existing public primary and pre-primary schools.[11] Contractors receive funding on a per-pupil basis and in exchange are responsible for the daily management of the schools.

Eight contractors were allocated rights to manage public schools by the government under the PSL program. After an open and competitive bidding process led by the Ministry of Education with the support of the Ark Education Partnerships Group (henceforth Ark, a UK charity), the Liberian government selected seven organizations, of which six passed financial due diligence. Stella Maris did not complete this step and, although included in our sample, was never paid. The government made a separate agreement with Bridge International Academies (not based on a competitive tender), but considers Bridge part of the PSL program. The organizations are as follows, ordered by the number of schools they manage that are part of the RCT: 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

[11]There are nine grades per school: three early childhood education grades (Nursery, K1, and K2) and six primary grades (grade 1 - grade 6).

Collective Action[12] (4 schools), and Stella Maris (4 schools).[13]

PSL schools remain public schools that should be free of charge and non-selective (i.e., contractors are not allowed to charge fees or to discriminate in admissions, for example on learning levels). Traditional public schools are not fully free. Although public primary education is nominally free starting in Grade 1,[14] 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 contractors receive *additional* funding of USD 50 per student (with a maximum of USD 3,250 or 65 students per grade).[15] Contractors have complete autonomy over the use of these funds (e.g., they can be used for teacher training, school inputs, or management personnel).[16] On top of that, contractors may raise more funds on their own.

Contractors 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 also 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. First, PSL schools are managed by private organizations. Second, PSL schools were theoretically guaranteed one teacher per grade in each school, plus extra funding. Third, private contractors 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.

[12]Youth Movement for Collective Action began the evaluation as "Liberian Youth Network," or LIYONET. The group has since changed its name.

[13]Bridge International Academies is managing two additional demonstration schools that were not randomized and are thus not part of our sample. Omega Schools opted not to operate two of their assigned schools, which we treat as non-compliance. Rising Academies opted not to operate one of their assigned schools (which we treat as non-compliance), and was given one non-randomly assigned school in exchange (which is outside our sample). Therefore, the set of schools in our analysis is not identical to the set of schools actually managed by PSL contractors.

[14]Officially, public schools are free, but in reality most charge informal fees. See Section 3.4 for statistics on these fees.

[15]Neither Bridge International Academies nor Stella Maris received the extra $50 per pupil. Stella Maris did not complete financial due diligence, and Bridge International Academies had a separate agreement with the Ministry of Education.

[16]Contractors may spend some of their 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 in practice. Only 8% of teachers in PSL schools were paid by contractors at the end of the school year. Information interviews with contractors indicate that in most cases, the contractors are paying these salaries while awaiting placement of the teachers on the government payroll, and they expect to be reimbursed by the government once that occurs.

Table 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	Contractor
Who sets curriculum?	Government	Government + contractor supplement
Funding		
Primary user fees (annual USD)	Zero	Zero
ECE user fees (annual USD)	$38	Zero
Extra funding per pupil (annual USD)	NA	$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 $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 contractors were authorized to cap class sizes at 65.
[c] Bridge International Academies has first pick, before other contractors, of the new teacher-training graduates.

2.1.3 What do contractors do?

A core feature of PSL is that contractors 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 contractors' own description of their model — where the incentives to exaggerate may be strong, and activities may be defined in non-comparable ways across contractors — we administered a survey module to teachers in all treatment schools, asking if they had heard of the contractor, and if so, what activities the contractor had engaged in. We summarize teachers' responses in Figure 2, which shows considerable variation in the specific activities and the total activity level of contractors.

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

Figure 2: What did contractors do?

		Contractor							
		Stella M	YMCA	Omega	BRAC	Bridge	Rising	St. Child	MtM
Contractor Support	Operator 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 contractor(%)	46	96	100	95	100	100	100	100
	Has anyone from (contractor) been to this school?(%)	42	88	100	94	100	100	99	100
Contractor 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
	Organize 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 Contractor 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

The figure reports simple proportions (not treatment effects) of teachers surveyed in PSL schools who reported whether or not the contractor responsible for their school had engaged in each of the activities listed. The sample size, n, of teachers interviewed with respect to each contractor is: Stella Maris, 26; Omega, 141; YMCA, 26; BRAC, 170; Bridge, 157; Street Child, 80; Rising Academy, 31; More than Me, 46. Recall that the standard error for a proportion, p, is $\sqrt{(p(1-p))/n}$. *This sample only includes compliant treatment schools.*

2.1.4 Cost data and assumptions

The government designed the PSL program based on the estimate that it spends roughly $50 per child on teacher salaries in all public schools, and it planned to continue to do so in PSL schools (Werner, 2017).[17] On top of this, contractors would be offered a $50 per-pupil payment to cover their costs.[18] This cost figure was chosen because $100 was deemed a realistic medium-term goal for public expenditure on primary education nationwide (Werner, 2017). To locate this in a global context, $50 is about what was spent per primary pupil by governments in Guinea in 2014, Afghanistan in 2015, Ghana in 2001, or India in 1999.

[17] As shown in Section 3, 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 $70.

[18] As noted above, neither Bridge International Academies nor Stella Maris received the extra $50 per pupil.

$100 is comparable to Tanzania in 2014, Pakistan in 2015, Ghana in 2014, or India in 2010. (The World Bank, 2015b, 2015a)

In the first year, contractors spent far more than this amount.[19] *Ex ante* per-pupil contractor budgets (on top of the Ministry's costs) ranged from a low of approximately $57 for Youth Movement for Collective Action to a high of $1,050 for Bridge International Academies (see Figure 3a). *Ex post* per-pupil contractor expenditure (on top of the Ministry's costs) ranged from a low of approximately $48 for Street Child to a high of $663 for Bridge International Academies (see Figure 3b). 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 contractors would be irrelevant for policy evaluation in a public-private partnership with this structure. If the contractors are willing to make an agreement in which the government pays $50 per pupil, contractors' losses are inconsequential to the government. In practice, philanthropic donors have stepped in to fund some contractors' high costs under PSL. Thus we present analyses in this report using both the Ministry's $50 long-term cost target and contractors' actual budgets.[20]

Contractors' budgets for the first year of the program are likely a naïve measure of program cost, as these budgets combine start-up costs, fixed costs, and variable costs.[21] It is possible to distinguish start-up costs from the other costs as shown in Figure 3, and these make up a small share of the first-year totals for most contractors. But it is not possible to distinguish fixed from variable costs in the current budget data. In informal interviews, some contractors (e.g., Street Child) profess to operate a mostly 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 precipitously if scaled; however, we have no direct evidence of this, and our best estimate is that Bridge's international operating cost, at scale, is between $191 and $220 per pupil annually.[22]

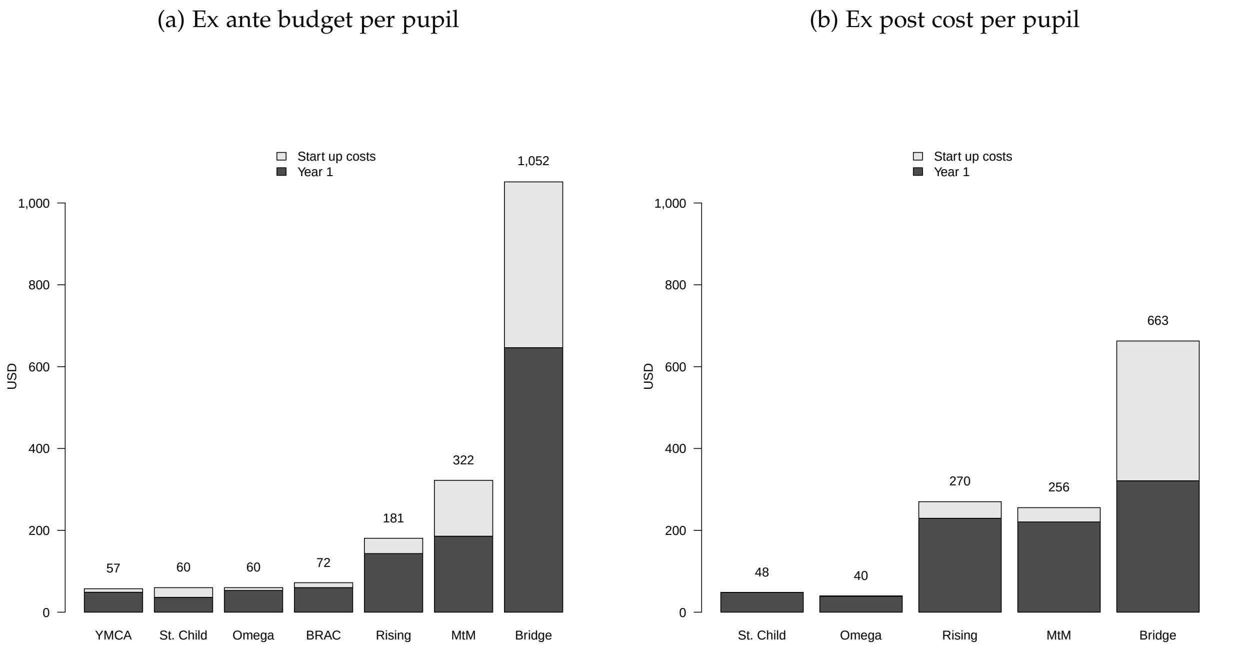
[19]Several caveats apply to the cost figures here, which are our own estimates based on contractors' self-reported budget data, and combine start-up costs, fixed costs, and variable costs. At the time of writing, the most comparable cost data we have access to are contractors' *ex ante* budgets, rather than actual expenditures. Five contractors also submitted (self-reported) data to the evaluation team on actual expenditures. Based on their reports and our calculations, More Than Me had a total expenditure per child in the first year of $255.55 (recurring: $220.55, start up cost: $34.95); Bridge International Academies had a total expenditure per child in the first year of $662.74 (recurring: $321.15, start up cost: $341.59, though the company itself prefers to omit the costs of customizing their software for Liberia, developing testing materials, internal monitoring and evaluation, etc., bringing the start-up cost figure down to $50); Street Child had a total expenditure per child in the first year of $48.48; Rising Academies had a total expenditure per child in the first year of $270 (recurring: $229.5, start up cost: $40.5); and Omega Schools had a total expenditure per child in the first year of $39.75 (recurring: $39.10, start up cost: $0.65). Some contractors (e.g., Street Child) profess to operate a mostly variable cost model, while others (e.g., Bridge) report that their costs are almost entirely fixed, and unit costs would fall precipitously if scaled.

[20]It is important to note that while some contractors relied almost exclusively on the $50 per child subsidy from the PSL pool fund, others have raised additional money from donors. Notably, Bridge International Academies relied entirely on direct grants from donors and opted not to participate in the competitive bidding process for the $50 per pupil subsidy which closed in June 2016. However, Bridge did subsequently submit an application for this funding in January 2017, which was not approved, but allows us access to their budget data. Bridge instead followed a bilateral memorandum of understanding (MOU) signed with the government of Liberia (Ministry of Education - Republic of Liberia, 2016b). In practice, they operated as part of the larger PSL program. A noteworthy difference is that Bridge was authorized to cap class sizes somewhere between 45 and 55 students per class, while other contractors were authorized to cap them at 65.

[21]Another possibility is that contractors are spending more during the first years of the program to prove effectiveness, but will lower expenditure once they are locked in a long-term contract.

[22]In written testimony to the UK House of Commons, Bridge stated that its fees were between $78 and $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, Bridge co-founder Shannon May stated that the company had supplemented its fee revenue with more than $12 million in the previous year (May, 2017). This is equivalent to an additional $120 per pupil, and implies Bridge spends between $191 and $220 per pupil at its current

Figure 3: Budget and costs as reported by contractors



Note: Numbers in 3a are based on contractors' ex-ante budgets, as 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 3b are based on self-reported data on ex post expenditures (inclusive of both fixed and variable costs) submitted to the evaluation team by five contractors in various formats. Numbers do not include the cost of teaching staff borne by the Ministry of Education.

2.1.5 Challenges

Before going into the results, we want to highlight some of the challenges contractors faced when setting up to manage their schools. The first is simply the amount of time they had. The final school allocation (after filtering based on contractor's location preferences and randomization) was given to contractors on July 18th. The first day of the academic year was September 5th. That is, contractors had less than two months to visit their schools, engage the community and teachers, conduct teacher training, and set up their management systems. Additionally, three of the contractors did not have a local presence in Liberia before the program.

The second hurdle is that setting up a functioning operation in Liberia is onerous. As a frame of reference, according to the World Bank (2017)'s Doing Business report, there are only 16 countries in the world where it is harder to set up a business: Enforcing contracts can take over three years; importing goods is burdensome (only five countries rank lower); and it takes more than a year to get a business connected to the electric grid (and even then, electricity flow is unreliable). Traveling to schools is a non-trivial task. Less than 6% of roads in the country are paved, and during the rainy season (which lasts 7–9 months) most of the roads are impassable.[23] Only three recently paved roads are in good condition

global scale.

[23]Note, however, that schools in the RCT–both treatment and control–are closer to paved roads than most schools in the country, as shown in Table A.1.

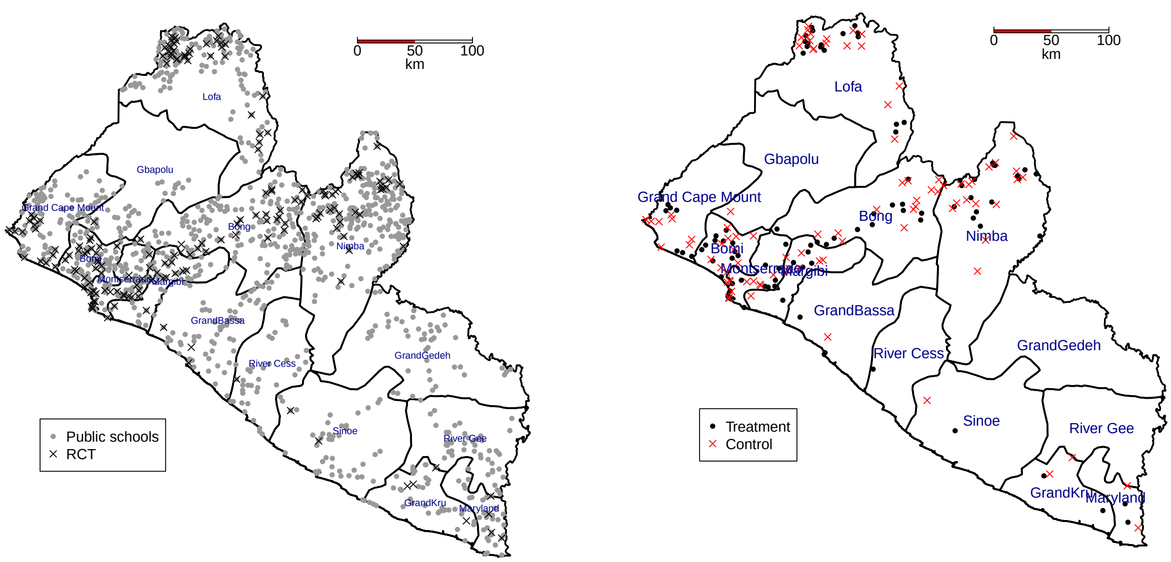
throughout the year: The road between Monrovia and Ganta, the road between Monrovia and Buchanan Port, and the road between Monrovia and Bo (Logistics Capacity Assessment - Wiki, 2016).

2.2 Experimental design

2.2.1 Sampling and random assignment

Liberia has 2,619 public primary schools. Private contractors 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.[24] 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. Figure 4a shows all public schools in Liberia and those within our sample, while Table A.1 in Appendix A shows the difference between schools in the experiment and other public schools. On average, schools in the experiment are closer to the capital (Monrovia) and have more students, greater resources, and better infrastructure.[25]

Figure 4: Public primary schools in Liberia



(a) Public schools in Liberia and those within the RCT.

(b) Geographical distribution of treatment and control schools, original treatment assignment.

We paired schools in the experiment sample within each district according to a principal component analysis (PCA) index of school resources.[26] This pairing implicitly stratified treatment by school resources

[24] As indicated by Education Management Information System (EMIS) data.

[25] 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.

[26] 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;

within each private contractor, but not across contractors. We gave a list of "counterparts" to each contractor based on their location preferences, so that each list had twice the number of schools they were to operate. Two contractors, Omega Schools and Bridge International Academies, required schools with 2G connectivity. Additionally, each contractor submitted to the government a list of the counties they were willing to work in. Note that Bridge International Academies had first pick of schools. Once each contractor approved this list, we randomized the treatment assignment within each pair.[27]

Private contractors did not manage all the schools originally assigned to treatment. After contractors 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 EMIS 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. Contractors did not operate in these schools and we treat them as non-compliant, presenting results in an intention-to-treat framework. Replacement schools were provided to these contractors, 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). Contractors 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.[28] Omega Academies opted not to operate two of their assigned schools, which we treat as non-compliance. Rising Academies opted not to operate one of their assigned schools (which we treat as non-compliance), and was given one non-randomly assigned school in exchange (which is outside our sample). Bridge International Academies is managing two extra demonstration schools that were not randomized and are thus not part of our sample. Therefore, the set of schools in our analysis is not identical to the set of schools actually managed by PSL contractors. Figure 4b shows the original treatment assignment. Appendix L contains a complete list of the schools related to the PSL program, and Table L.1 summarizes the overlap between schools in our main sample and the set of schools actually managed by PSL contractors.

Treatment assignment may change the student composition across schools. Thus, to prevent differences in the composition of students from driving differences in test scores, we sampled 20 students (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 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.

2.2.2 Timeline of research and intervention activities

We conducted the baseline survey in September/October 2016 and the follow-up survey in May/June 2017. A second follow-up survey will take place in March/April 2019 conditional on continuation of the project and preservation of the control group. See Figure A.1 in Appendix A for a timeline of intervention

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.

[27] There is one threesome due to logistical constraints in the assignment of schools across counties, which resulted in one extra treatment school.

[28] We analyzed results for this "final" treatment and control school list, and they are almost identical to the results for the "original" list—perhaps unsurprisingly, given that they only differ by four pairs of schools. Results for this final list of treatment and control schools are available upon request.

and research activities. Note that we collected the baseline data 2 to 8 weeks after the beginning of treatment. Thus, we focus on slow-moving characteristics and administrative data collected before the program began when checking balance between treatment and control schools to verify whether treatment was truly randomly assigned (see Section 2.2.5). As discussed below (in Section 3.1), there is evidence that this baseline was already "contaminated" by very short-run treatment effects on fast-moving outcomes such as teacher attendance and even test scores, with implications for our estimation strategy (e.g., we cannot control for several baseline characteristics).

2.2.3 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. For the math portion 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 F has details on the construction of the test.

We estimate an item response theory (IRT) model for each round of data collection.[29] IRT models are the standard in the assessments literature for generating comparative test scores. 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.[30] There are two important and 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.

2.2.4 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, 2015). Furthermore, we conducted school-level surveys to collect information about school facilities, the teacher roster, input availability (e.g., textbooks) and expenditures.

[29] Note that the overlap between baseline and follow-up is small, and therefore we do not estimate the same IRT model across rounds.

[30] 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 (2015b, 2016); Muralidharan, Singh, and Ganimian (2016); Mbiti et al. (2017). Das and Zajonc (2010) provide a nice introduction to IRT models, while van der Linden (2017) provides a full treatment of IRT models.

We asked principals how they use their time, and 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.[31] Additionally, we asked principals to complete two commonly used human resource instruments to measure individuals' "intuitive score" (Agor, 1989) and "time management profile" (Schermerhorn, Osborn, Uhl-Bien, & Hunt, 2011).

At follow-up, we surveyed a random subset of households from 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.

2.2.5 Balance and attrition

As mentioned above, baseline 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 at baseline (see Table 2). We intentionally leave test scores out of this table table as short-run treatment effects are already noticeable at baseline—we postpone discussing these effects to Section 3.1. 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); 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 administrative data (EMIS), the number of students, infrastructure, and resources available to students are not statistically different across treatment and control schools (see Table A.2 in Appendix A).

We took great care to avoid differential 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. Panel C shows that attrition from our original sample is balanced between treatment and control (and is below 4% overall).[32]

[31] While management practices are difficult to measure, previous work has constructed detailed instruments to measure them in schools (e.g., see Bloom, Lemos, Sadun, and Van Reenen (2015); Crawfurd (in press); Lemos and Scur (2016)). Due to budget constraints, we checked easily observable differences in school management.

[32] Appendix E has more details on the tracking and attrition that took place in each round of data collection.

Table 2: Baseline balance: Observable, time-invariant school and student characteristics

	(1) Control	(2) Treatment	(3) Difference	(4) Difference (F.E)
Panel A: School characteristics (N = 185)				
Facilities (PCA)	-0.001	-0.075	-0.074	-0.067
	(0.169)	(0.156)	(0.230)	(0.231)
% holds some classes outside	14.130	13.978	-0.152	0.000
	(3.652)	(3.615)	(5.138)	(5.094)
% rural	80.435	79.570	-0.865	-0.361
	(4.159)	(4.204)	(5.913)	(4.705)
Travel time to nearest bank (mins)	68.043	75.129	7.086	7.079
	(6.308)	(7.165)	(9.547)	(8.774)
Panel B: Student characteristics (N = 3,499)				
Age in years	12.289	12.405	0.116	0.056
	(0.070)	(0.068)	(0.170)	(0.112)
% male	56.413	54.942	-1.470	-1.767
	(1.184)	(1.192)	(2.013)	(1.257)
Wealth index	0.025	-0.008	-0.034	0.008
	(0.037)	(0.037)	(0.140)	(0.059)
% in top wealth quartile	0.218	0.198	-0.020	-0.017
	(0.010)	(0.010)	(0.026)	(0.014)
% in bottom wealth quartile	0.285	0.267	-0.018	-0.012
	(0.011)	(0.011)	(0.039)	(0.019)
ECE before grade 1	0.822	0.835	0.013	0.013
	(0.009)	(0.009)	(0.024)	(0.017)
Panel C: Attrition (N = 3,499)				
% interviewed	96.01	95.98	-0.03	-0.23
	(0.47)	(0.47)	(0.63)	(0.44)

Baseline data was collected 2 to 8 weeks after the beginning of treatment; hence, the focus here is on time-invariant characteristics (note that some of these characteristics may vary in response to the program in the long run, but are time-invariant given the the duration of our study). 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 (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 motorbike, a fan, and a phone. Panel C shows the attrition rate (proportion of students interviewed at baseline who we were unable to interview at follow-up). The standard errors are clustered at the school level. The sample is the original treatment and control allocation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3 Main 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.[33]

3.1 Test scores

Following our pre-analysis plan, we report treatment-effect estimates based on three specifications. The first specification amounts to a simple comparison of post-treatment outcomes for treatment and control individuals, where 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_i is an individual error term.

$$Y_{isg} = \alpha_g + \beta_1 treat_s + \varepsilon_i \quad (1)$$

$$Y_{isg} = \alpha_g + \beta_2 treat_s + \gamma_2 X_i + \delta_2 Z_s + \varepsilon_i \quad (2)$$

$$Y_{isg} = \alpha_g + \beta_3 treat_s + \gamma_3 X_i + \delta_3 Z_s + \zeta_3 Y_{isg,-1} + \varepsilon_i \quad (3)$$

The second specification adds controls for baseline characteristics measured at the individual level (X_i) and school level (Z_s).[34] Finally, in equation (3) we use an ANCOVA specification (i.e., controlling for baseline individual outcomes).

Adding controls, as in equation (2), should increase the precision of our results. However, controlling for baseline outcomes, as in equation (3), may also risk attenuation bias in the treatment effect estimates if the baseline outcomes are imbalanced. This is, in fact, what we observe in our baseline data. Students in treatment schools score higher at baseline than those in control schools by .076σ in math (p-value=.077) and .091σ in English (p-value=.049).

There is some evidence that this imbalance is not simply due to "chance bias" in randomization, but rather a treatment effect that materialized in the weeks between the beginning of the school year and the baseline survey. First, 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).[35] Second, time-invariant student characteristics are balanced across treatment and control (see Table 2). Third, the effects on English and math appear to materialize in the later weeks of the fieldwork, as shown in Figure A.2, consistent with a treatment effect rather than imbalance.[36] Thus we face a trade-off between precision and attenuation bias in choosing between the three specifications above. Our preferred specification is equation (2), although we report all three results.

[33] A randomized controlled trial registry entry and the pre-analysis plan, are available at: https://www.socialscienceregistry.org/trials/1501.

[34] These controls were specified in the pre-analysis plan and are listed in Table A.11.

[35] Note that there is evidence that schooling (Brinch & Galloway, 2012) and cognitive training (Jaeggi, Buschkuehl, Jonides, & Shah, 2011) can increase performance on abstract reasoning tests.

[36] As mentioned in Section 2, we collected the baseline data 2 to 8 weeks after the beginning of treatment. While most contractors started the school year on time, 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.

Table 3 shows results from student tests at baseline and at follow-up one year later. The first two columns show differences between control and treatment schools' test scores at baseline (September/October 2016), while the last four columns show the difference in May/June 2017. In our preferred specification (Column 5) the treatment effect of PSL after one year is .18σ for English (p-value < 0.001) and .18σ for math (p-value < 0.001). Table A.4 in Appendix A shows both the ITT and the treatment-on-the-treated (ToT) effect (i.e., treatment effect only for students that actually attended a PSL school in 2016/2017), while Table A.5 shows the ITT effect using different measures of student ability.

Table 3: ITT treatment effects on learning

	Baseline		One-year follow-up			
	Difference	Difference (F.E.)	Difference	Difference (F.E.)	Difference (F.E. + Controls)	Difference (ANCOVA)
	(1)	(2)	(3)	(4)	(5)	(6)
English	0.06	0.09**	0.17**	0.17***	0.18***	0.13***
	(0.08)	(0.05)	(0.08)	(0.04)	(0.03)	(0.02)
Math	0.08	0.08*	0.18***	0.19***	0.18***	0.14***
	(0.07)	(0.04)	(0.06)	(0.04)	(0.03)	(0.02)
Abstract	0.05	0.05	0.05	0.05	0.05	0.03
	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
Composite	0.08	0.09*	0.18***	0.19***	0.19***	0.14***
	(0.07)	(0.05)	(0.07)	(0.04)	(0.03)	(0.02)
New modules			0.18**	0.20***	0.19***	0.16***
			(0.07)	(0.04)	(0.04)	(0.03)
Conceptual			0.12**	0.14***	0.12***	0.10***
			(0.05)	(0.04)	(0.04)	(0.04)
Observations	3,499	3,499	3,498	3,498	3,498	3,498

Columns 1-2 use baseline data and show the difference between treatment and control (Column 1), and the difference taking into account the randomization design—i.e., including "pair" fixed effects—(Column 2). Columns 3-6 use May/June 2017 data and show the difference between treatment and control (Column 3) in test scores, the difference 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 difference using an ANCOVA style specification that controls for baseline test scores (Column 6).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To ease any concerns that differences in pre-treatment student ability drive the difference on test scores after one year, we estimate the treatment effect separately for students tested during the first and the second half of baseline field work (see Figure A.2). As discussed above, the imbalance in baseline test scores is only apparent for students tested later at baseline. Yet the difference in test scores at the one-year follow-up is almost identical regardless of which students are included in the sample. Mechanically, the treatment effects using an ANCOVA style specification become smaller for students tested later at baseline.

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 over new modules that were not in the baseline test is significant (.19σ, p-value $<$ 0.001), and statistically indistinguishable from the treatment effect over all the items (.19σ, p-value $<$ 0.001). Second, the treatment effect over the conceptual questions (which do not resemble the format of standard textbook exercises) is positive and significant (.12σ, p-value .0014). However, we cannot rule out that contractors narrowed the curriculum by focusing on English and mathematics or, conversely, that they generated learning gains in other subjects that we did not test.

Although reporting the impact of interventions in standard deviations is the norm in the education and experimental literature, we also report results as "equivalent years of schooling" (EYOS) following Evans and Yuan (2017). Results in this format are easier to communicate to policymakers and the general public, by juxtaposing treatment effects with the learning from business-as-usual schooling. In our data the average increase in test scores for each extra year of schooling in the control group is .31σ in English and .28σ in math. Thus, the treatment effect is roughly 0.56 EYOS for English and 0.66 EYOS for math. See Appendix G for a detailed explanation of the methodology to estimate EYOS, and a comparison of EYOS and standard deviation across countries. Additionally, Appendix H shows absolute learning levels in treatment and control schools for a subset of the questions that are comparable to other settings, to allow direct comparisons with learning levels in other countries. Note that despite the positive treatment effect of the program, students in treatment schools are still behind their international peers.

3.2 Enrollment, attendance and student selection

The previous section showed that education quality, measured in an ITT framework using test scores, 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. The brief answer is that 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. This does not appear to be driven by selection of "better" students, but simply contractors capping class sizes and eliminating double shifts.[37] As shown in Section 5.4, 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 4. There are a few noteworthy items. First, treatment schools are slightly larger before treatment: They have 34 (p-value .078) students more on average at baseline.[38] Second, PSL schools have on average 52 (p-value $<$ 0.001) more students than control schools in the 2016/2017 academic year, which results in a net increase (after controlling for baseline differences) of 19 (p-value .24) students per school.[39]

Since contractor compensation is based on the number of students enrolled rather than the number of students actively attending school, it is possible that increases in enrollment do not translate into increases in student attendance. An independent measure of student attendance conducted by our enumerators during a spot check shows that students are 13 (p-value $<$ 0.001) percentage points more likely to be in

[37] Three Bridge International Academies treatment schools (representing 24% of total enrollment in Bridge treatment schools) had double shifts in 2015/2016, but not in 2016/2017. One Omega Schools treatment school (representing 6.8% of total enrollment in Omega treatment schools) had double shifts in 2015/2016, but not in 2016/2017. Note that the MOU between Bridge and the Ministry of Education explicitly authorized eliminating double shifts (Ministry of Education - Republic of Liberia, 2016b).

[38] Note that Table A.2 uses EMIS data, while Table 4 uses data independently collected by IPA. While the difference in enrollment in the 2015/2016 academic year is only significant in the latter, the point estimates are remarkably similar across both tables.

[39] Once the EMIS data for the 2016/2017 school year are released, we will reexamine this issue to study whether increases in enrollment come from children previously out-of-school or from children previously enrolled in other schools.

school during class time (see Panel A, Table 4).

Turning to the question of student selection, the proportion of students with disabilities is not statistically different in PSL schools and control schools (Panel A, Table 4).[40] 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 more likely to be enrolled in school at all (see Panel B, Table 4). We complement the selection analysis using student-level data on wealth, gender, and age in Table A.6 in Appendix A. We find no evidence that any group of students is systematically excluded from PSL schools.

Table 4: ITT treatment effects on enrollment, attendance and selection

	(1) Control	(2) Treatment	(3) Difference	(4) Difference (F.E)
Panel A: School level data (N = 1,663)				
Enrollment 2015/2016	259.73	293.60	33.87*	33.65*
	(11.14)	(15.95)	(19.45)	(19.00)
Enrollment 2016/2017	258.05	310.80	52.74***	52.36***
	(13.16)	(12.01)	(17.82)	(15.54)
Difference	-1.68	17.19	18.87	18.71
	(10.11)	(12.88)	(16.38)	(15.79)
Attendance % (spot check)	35.10	48.31	13.21***	12.89***
	(2.84)	(2.55)	(3.82)	(3.03)
% of students with disabilities	0.39	0.58	0.20	0.20
	(0.07)	(0.12)	(0.14)	(0.14)
Panel B: Student level data (N = 3,493)				
% enrolled in the same school	83.32	80.51	-2.81	0.67
	(0.88)	(0.92)	(3.90)	(2.19)
% enrolled in school	93.97	94.11	0.13	1.22
	(0.58)	(0.56)	(1.37)	(0.89)
Days missed, previous week	0.85	0.85	-0.00	-0.05
	(0.03)	(0.03)	(0.10)	(0.07)

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 (Column 4). Our enumerators conducted the attendance 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. Standard errors are clustered at the school level. The sample is the original treatment and control allocation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Contractors 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 explore whether there is any heterogeneity in enrollment by how binding these class-caps are. We estimate whether the caps are binding for each student by examining whether the average enrollment in her grade cohort and the two

[40]We do, however, note that 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).

adjacent grade cohorts (i.e., one grade above and below) was larger prior to treatment than the theoretical class-size cap under PSL. We average over three cohorts because some contractors 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):

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

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 contractor 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 contractor o. We label a grade-school combination as "constrained" if $c_{igso} > 1$.

Column 1 in Table 5 shows that enrollment in constrained school-grades decreases, while enrollment in unconstrained school-grades increases. Thus, schools far below the cap are driving the total (positive) treatment effect on enrollment and schools near or above the cap partially offset it with declining enrollment. Our student data reveal this pattern as well: Columns 2 and 3 in Table 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 school overall). But in constrained classes students are less likely to be enrolled in the same school. While there is no effect on overall school enrollment, previous research has shown that switching schools is disruptive for children (Hanushek, Kain, & Rivkin, 2004). Figure A.3 in Appendix A shows the enrollment treatment effect across all grades (left panel), across grades with no constraints (middle panel), and across grades with constraints (right panel). Finally, note that test-scores did improve for students in constrained classes. This result has to be treated carefully as it includes the positive treatment effect on students who did not change schools (possibly compounded by smaller class sizes) with the effect on students removed from their schools.

Table 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	4.79***	4.10***	1.69**	0.15***
	(0.99)	(1.41)	(0.73)	(0.034)
Constrained=1 × Treatment	-19.9**	-12.9*	0.19	0.32**
	(9.23)	(7.61)	(3.89)	(0.13)
No. of obs.	1,663	3,631	3,491	3,496
Mean control (Unconstrained)	-0.23	82.23	93.45	0.13
Mean control (Constrained)	-6.93	83.23	94.19	-0.09
α_0 = Constrained - Unconstrained	-24.72	-17.04	-1.50	0.17
p-value ($H_0 : \alpha_0 = 0$)	0.01	0.02	0.71	0.18

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. Standard errors are clustered at the school level. The sample is the original treatment and control allocation. Note that there were 195 constrained classes at baseline (holding 30% of students), and 1,470 unconstrained classes at baseline (holding 70% of students).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.

3.3.1 Inputs and resources

Teachers, one of the most important inputs of education, change in several ways (see Panels A/B in Table 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,[41] 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 a simple memory, math, word association, and abstract thinking test).[42] While teachers in PSL schools earn higher wages, previous experimental literature has found that large unconditional increases in teacher salaries have no effect on student performance in the short run (de Ree, Muralidharan, Pradhan, & Rogers, 2015).

Our enumerators conducted a "materials" check during classroom observations (See Panels C - Table 6). Since we could not conduct classroom observations in schools that were out of session during our visit, Table A.7 in Appendix A presents Lee 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

[41]Once the EMIS data for the 2016/2017 school year are released, we will reexamine this issue to study whether teachers who were fired were allocated to other public schools. Note that while the majority of released teachers are on the government's payroll, some of the dismissed teachers are thus they have not necessarily been assigned to other public schools.

[42]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.

25 percentage points (p-value < 0.001) more likely to have a textbook and 8.7 percentage points (p-value .052) more likely to have writing materials (both a pen and a copybook).

Table 6: ITT treatment effects on inputs and resources

	(1) Control	(2) Treatment	(3) Difference	(4) Difference (F.E)
Panel A: School-level outcomes (N = 185)				
Number of teachers	7.02	9.62	2.60***	2.61***
	(0.33)	(0.29)	(0.44)	(0.37)
Pupil-teacher ratio (PTR)	40.15	33.27	-6.88***	-6.94***
	(1.85)	(1.16)	(2.18)	(1.97)
New teachers	1.77	4.81	3.03***	3.01***
	(0.21)	(0.27)	(0.34)	(0.35)
Teachers re-assigned	2.11	3.24	1.13**	1.10**
	(0.27)	(0.39)	(0.48)	(0.47)
Panel B: Teacher-level outcomes (N = 1,167)				
Age in years	46.37	39.09	-7.28***	-7.10***
	(0.53)	(0.45)	(1.02)	(0.68)
Experience in years	15.79	10.59	-5.20***	-5.26***
	(0.47)	(0.33)	(0.76)	(0.51)
% has worked at a private school	37.50	47.12	9.62**	10.20***
	(2.09)	(1.77)	(3.76)	(2.42)
Test score in standard deviations	-0.01	0.13	0.14*	0.14**
	(0.05)	(0.04)	(0.07)	(0.06)
% certified (or tertiary education)	58.05	60.11	2.06	4.20
	(1.94)	(1.64)	(4.87)	(2.99)
% on government payroll	73.73	68.35	-5.38	-4.16
	(1.75)	(1.60)	(4.27)	(2.97)
% paid by contractor	0.00	7.18	7.18***	7.54***
	(0.00)	(0.89)	(1.80)	(1.48)
% of months paid on time	46.31	38.93	-7.38*	-5.72*
	(2.49)	(1.77)	(4.03)	(3.18)
Salary (USD/month)–Conditional on salary>0	104.54	121.36	16.82**	13.90***
	(3.63)	(2.09)	(6.56)	(4.53)
Panel C: Classroom observation (N = 133)				
Number of seats	20.38	20.46	0.07	0.51
	(1.71)	(1.50)	(2.27)	(2.02)
% with students sitting on the floor	4.48	2.44	-2.04	-3.26
	(2.55)	(1.71)	(3.07)	(2.28)
% with chalk	77.61	96.34	18.73***	17.93***
	(5.13)	(2.09)	(5.54)	(5.91)
% of students with textbooks	17.16	36.32	19.15***	24.56***
	(4.23)	(4.74)	(6.35)	(6.40)
% of students with pens/pencils	78.46	88.41	9.95**	8.70*
	(3.74)	(2.20)	(4.34)	(4.42)

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 (Column 4). Standard errors are clustered at the school level. The sample is the original treatment and control allocation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.3.2 School management

Two important management changes are shown in Table 7: PSL schools are 8.7 percentage points more likely to be in session during a regular school day (p-value .057), and have a longer school day that translates into 3.9 more hours a week of instructional time (p-value < 0.001). In addition, 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, we find that management practices (as measured by a PCA index[43] normalized to a mean of zero and standard deviation of one in the control group) are $.4\sigma$ (p-value < 0.001) higher in PSL schools. This effect size can be viewed as a boost for the average treated school from the 50th to the 66th percentile in management practices.

Table 7: ITT treatment effects on school management

	(1) Control	(2) Treatment	(3) Difference	(4) Difference (F.E)
% school in session	83.70	92.47	8.78*	8.66*
	(3.87)	(2.75)	(4.75)	(4.52)
Instruction time (hrs/week)	16.50	20.40	3.90***	3.93***
	(0.49)	(0.60)	(0.77)	(0.73)
Intuitive score (out of 12)	4.03	4.08	0.04	0.02
	(0.14)	(0.14)	(0.20)	(0.19)
Time management score (out of 12)	5.69	5.60	-0.09	-0.10
	(0.14)	(0.13)	(0.19)	(0.19)
Principal's working time (hrs/week)	20.60	21.43	0.83	0.84
	(1.51)	(1.23)	(1.94)	(1.88)
% of time spent on management	53.64	74.06	20.42***	20.09***
	(2.97)	(2.85)	(4.12)	(3.75)
Index of good practices (PCA)	-0.00	0.41	0.41***	0.40***
	(0.10)	(0.07)	(0.12)	(0.12)
Observations	92	93	185	185

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 (Column 4). Intuitve 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 A.8. 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. The sample is the original treatment and control allocation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[43] 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 A.8 has details on every component of the good practices index.

3.3.3 Teacher behavior

An important component of school management is teacher accountability and its effects on teacher behavior. Note that the changes in the composition of teachers in PSL schools confound any effect 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 directly by the Liberian government. In theory, private contractors 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 contractors can hold teachers accountable through monitoring and support, rather than rewards and threats.

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 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 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 7.4 percentage points (p-value < 0.001) less likely to have missed school the previous week.

Classroom observations also show changes in teacher behavior and pedagogical practices. First, teachers in PSL schools are 16 percentage points (p-value < 0.001) more likely to be engaging 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 27 percentage points (p-value < 0.001) less likely to be off-task.[44] 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). Additionally, teachers are 6.6 percentage points (p-value .0098) less likely to hit students in PSL schools.

[44] 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 8: ITT treatment effects on teacher behavior

	(1) Control	(2) Treatment	(3) Difference	(4) Difference (F.E)
Panel A: Spot checks (N = 185)				
% on schools campus	40.38	60.32	19.94***	19.79***
	(2.63)	(2.40)	(3.56)	(3.48)
% in classroom	31.42	47.02	15.60***	15.37***
	(2.61)	(2.76)	(3.80)	(3.62)
Panel B: Student reports (N = 185)				
Teacher missed school previous week (%)	25.09	17.77	-7.32***	-7.43***
	(1.56)	(1.13)	(1.92)	(1.96)
Teacher never hits students (%)	48.26	54.75	6.49**	6.56***
	(1.77)	(1.94)	(2.63)	(2.51)
Teacher helps outside the classroom (%)	46.74	50.02	3.28	3.42
	(1.87)	(1.89)	(2.66)	(2.29)
Panel C: Classroom observations (N = 185)				
Instruction (active + passive) (% of class time)	32.28	48.82	16.53***	16.35***
	(3.79)	(3.37)	(5.07)	(4.72)
Classroom management (% class time)	8.26	19.03	10.77***	10.69***
	(1.46)	(2.17)	(2.62)	(2.78)
Teacher off-task (% class time)	59.46	32.15	-27.31***	-27.04***
	(4.41)	(3.98)	(5.94)	(5.74)
Student off-task (% class time)	45.34	49.99	4.65	4.31
	(4.06)	(3.51)	(5.37)	(4.62)

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 (Column 4). Our enumerators conducted the attendance spot check in the middle of a school day. 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 A.7 has the results without imputing values. Standard errors are clustered at the school level. The sample is the original treatment and control allocation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As previously mentioned, 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 analysis in Appendix A 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 A.7 in Appendix A shows an ITT treatment effect of 14 percentage points (p-value < 0.001) on teacher attendance. Importantly, zero is not part of the Lee bounds for this effect. This aligns with previous findings showing that management practices have significant effects on worker performance (Bloom, Liang, Roberts, & Ying, 2014; Bloom, Eifert, Mahajan, McKenzie, & Roberts, 2013; Bennedsen, Nielsen, Pérez-González, & Wolfenzon, 2007).

3.4 Other outcomes

Student data (Table 9, Panel C) and household data (Table 9, Panel A) show that the program increases both student and parental satisfaction. Students in PSL schools are happier (measured by whether they think going to school is fun or not), and parents with children in PSL schools (enrolled in 2015/2016) are 7.4 percentage points (p-value .022) more likely to be satisfied with the education their children are receiving. Table C.1 in Appendix C has detailed data on student, parental, and teacher support and satisfaction with PSL.

Contractors are not allowed to charge fees and PSL should be free at all levels, including early-childhood education (ECE) for which fees are normally permitted in government schools. We interviewed both parents and principals regarding fees. Note that in both treatment and control schools parents are more likely to report paying fees than schools are to report charging them. Similarly, 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 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 schools 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.

On top of reduced fees, contractors often provide textbooks and uniforms free of charge to students (see Section 2.1.3). Indeed, interviews with parents reveal that household expenditure on fees, textbooks, and uniforms drops (see Table A.9 for details). In total, household expenditures on children's education decreases by 6.6 USD (p-value .11) in PSL schools.

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, but 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 A.10 for details).

To complement the effect of the program on cognitive skills, we study student attitudes and opinions (see Table 9, Panel C). Some of the control group rates are noteworthy: Only 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. 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 (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 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 contractors per se but rather to the program as a whole.

Table 9: ITT treatment effects on household behavior, fees, and student attitudes

	(1) Control	(2) Treatment	(3) Difference	(4) Difference (F.E)
Panel A: Household behavior (N = 1,116)				
% satisfied with school	67.47	74.89	7.43**	7.45**
	(2.50)	(2.00)	(3.20)	(3.23)
% paying any fees	73.37	47.85	-25.52***	-25.68***
	(1.93)	(2.03)	(4.70)	(3.25)
Fees (USD/year)	8.03	5.69	-2.34**	-2.93***
	(0.42)	(0.41)	(0.96)	(0.60)
Expenditure (USD/year)	73.38	66.43	-6.96	-6.58
	(3.49)	(3.10)	(7.12)	(4.12)
Engagement index (PCA)	-0.09	-0.11	-0.02	-0.03
	(0.04)	(0.03)	(0.08)	(0.06)
Panel B: Fees (N = 184)				
% with > 0 ECE fees	30.77	11.83	-18.94***	-18.98***
	(4.87)	(3.37)	(5.92)	(5.42)
% with > 0 primary fees	29.67	12.90	-16.77***	-16.79***
	(4.82)	(3.50)	(5.95)	(5.71)
ECE Fee (USD/year)	1.42	0.57	-0.85**	-0.87***
	(0.29)	(0.20)	(0.35)	(0.33)
Primary Fee (USD/year)	1.22	0.54	-0.68**	-0.70**
	(0.25)	(0.18)	(0.31)	(0.31)
Panel C: Student attitudes (N = 3,498)				
School is fun	0.53	0.58	0.05**	0.05**
	(0.01)	(0.01)	(0.02)	(0.02)
I use what I'm learning outside of school	0.49	0.52	0.04	0.05***
	(0.01)	(0.01)	(0.02)	(0.02)
If I work hard, I will succeed.	0.55	0.60	0.05*	0.04***
	(0.01)	(0.01)	(0.03)	(0.02)
Elections are the best way to choose a president	0.88	0.90	0.02*	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
Boys are smarter than girls	0.69	0.69	-0.00	0.01
	(0.01)	(0.01)	(0.02)	(0.01)
Some tribes in Liberia are bad	0.79	0.76	-0.03	-0.03**
	(0.01)	(0.01)	(0.02)	(0.01)

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 (Column 4). Standard errors are clustered at the school level. The sample is the original treatment and control allocation. The index for parent engagement is the first component from a principal component analysis across several measures of parental engagement; see Table A.10 for details.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4 Understanding mechanisms

The question of mechanisms can be divided into two parts: What changed? And which changes mattered for learning outcomes? We answer the first question in the previous section. In this section we use observational data analysis to answer the latter question as best as possible.

There are three related goals in the analysis below: (i) to highlight which mechanisms correlate with learning gains; (ii) to uncover how much of the treatment effect is the result of an increase in resources (e.g., teachers and per-child expenditure); and (iii) to estimate whether PSL schools are more productive (i.e., whether they use resources more effectively to generate learning). To attain these goals we use mediation analysis, and follow the general framework laid out in Imai, Keele, and Yamamoto (2010) and Imai, Keele, and Tingley (2010).[45]

The mediation effect of a learning input (e.g., teacher attendance) is the change in learning gains that can be attributed to changes in this input caused by treatment. Formally, we can estimate the mediation effect via the following two equations:

$$M_{isg} = \alpha_g + \beta_4 treat_s + \gamma_4 X_i + \delta_4 Z_s + u_i \tag{4}$$

$$Y_{isg} = \alpha_g + \beta_5 treat_s + \gamma_5 X_i + \delta_5 Z_s + \theta_5 M_{isg} + \varepsilon_i \tag{5}$$

where Y_{isg} is the test score 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_i and u_i are individual error terms. X_i and Z_s are individual and school-level controls measured before treatment, while M_{isg} are the potential mediators for treatment (i.e., learning inputs measured post-treatment). Equation 4 is used to estimate the effect of treatment on the mediator (β_4), while equation 5 is used to estimate the effect of the mediator on learning (θ_5).

The mediation effect is $\beta_4 \times \theta_5$, i.e., the effect of the mediator on learning gains (θ_5) combined with changes in the mediator caused by treatment (β_4). The direct effect is β_5. The mediation effect and the direct effect are in the same units (the units of Y_{isg}), and are therefore comparable.

The crux of a mediation analysis is to get consistent estimators of θ_5 (and therefore of β_5). Imai, Keele, and Yamamoto (2010) show that under the following assumption:

Assumption 1 (Sequential ignorability)

$$Y_i(t', m), M_i(t) \perp\!\!\!\perp T_i | X_i = x \tag{6}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | X_i = x, T_i = t \tag{7}$$

where $Pr(T_i = t|X_i = x) > 0$ and $Pr(m_i(t) = m|T_i = t, X_i = x) > 0$ for all values of t, x and m.

the OLS estimators for θ_5 and β_5 are consistent. While randomization implies that equation 6 in Assumption 1 is met, we do not have experimental variation in any of the possible mediators and thus unobserved variables may confound the relationship between mediators and learning gains, violating equation 7 in Assumption 1 (Green, Ha, & Bullock, 2010; Bullock & Ha, 2011). To mitigate omitted variable bias we use the rich data we have on soft inputs (e.g., hours of instruction and teacher behavior) and hard inputs (e.g., textbooks and number of teachers) and include a wide set of variables in M_{is}. But two problems arise: a)

[45] This framework is closely related to the framework used by Heckman, Pinto, and Savelyev (2013); Heckman and Pinto (2015), and there is a direct mapping between the two.

As Bullock and Ha (2011) state, "it is normally impossible to measure all possible mediators. Indeed, it may be impossible to merely *think* of all possible mediators". Thus, despite being extensive, the list may be incomplete. b) It is unclear what the relevant mediators are, and adding an exhaustive list of them will reduce the degrees of freedom in the estimation and lead to multiple-inference problems. As a middle ground between these two issues, we use a Lasso procedure to select controls that are relevant from a statistical point of view, as opposed to having the researcher choose them *ad hoc*. The Lasso procedure is akin to OLS but penalizes according to the number of controls used (see James, Witten, Hastie, and Tibshirani (2014) for a recent discussion).

We use two sets of mediators. The first only includes raw inputs: teachers per student, textbooks per student, and teachers' characteristics (age, experience, and ability). Results from estimating equation 5 with these mediators are shown in Columns 2 and 3 of Table 10. The second includes raw inputs as well as changes in the use of these inputs (e.g., teacher behavior measurements, student attendance, and hours of instructional time per week). Results from estimating equation 5 with these mediators are shown in Columns 4 and 5 of Table 10. For reference, we include a regression with no mediators (Column 1) that replicates the results from Table 3.

Note that the treatment effect of PSL is positive even after controlling for more and better inputs (Columns 2 and 3). However, the drop in the point estimate, compared to Column 1, suggests that changes in inputs explain about half of the total treatment effect. The persistence of a "direct" treatment effect in these columns suggests that changes in the use of inputs are an important mechanism as well.

The results from Columns 3 and 4 provide ancillary evidence that changes in the use of inputs (i.e., management) are important pathways to impact. After controlling for how inputs are used (e.g., teacher attendance) the "direct" treatment effect is close to zero.

Table 10: Mediation analysis

		Inputs		Inputs & Management	
	(1)	(2)	(3)	(4)	(5)
Treatment	0.19***	0.09**	0.11**	0.01	0.01
	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)
Writing materials		-0.00	-0.00		-0.00
		(0.00)	(0.00)		(0.00)
PTR		-0.00	-0.00	-0.00	-0.00
		(0.00)	(0.00)	(0.00)	(0.00)
Teachers' age		-0.01***	-0.02***	-0.01***	-0.01***
		(0.00)	(0.00)	(0.00)	(0.00)
% exp. in private schools		-0.00	-0.00		-0.00
		(0.00)	(0.00)		(0.00)
Textbooks			-0.00		-0.00
			(0.00)		(0.00)
Teachers' experience			0.01**		0.01
			(0.01)		(0.01)
Teachers' test score			0.06		0.07
			(0.05)		(0.05)
Certified teachers			0.00		0.00
			(0.00)		(0.00)
Teacher attendance				0.00***	0.00***
				(0.00)	(0.00)
% of time spent on management					-0.04
					(0.08)
Index of good practices (PCA)					0.07***
					(0.02)
Student attendance					0.12
					(0.10)
Instruction (Classroom obs)					-0.00
					(0.00)
Hrs/week					0.01*
					(0.00)
No. of obs.	3,498	3,498	3,498	3,498	3,498
R2	0.53	0.53	0.53	0.53	0.54
Mediators	None	Lasso	All	Lasso	All

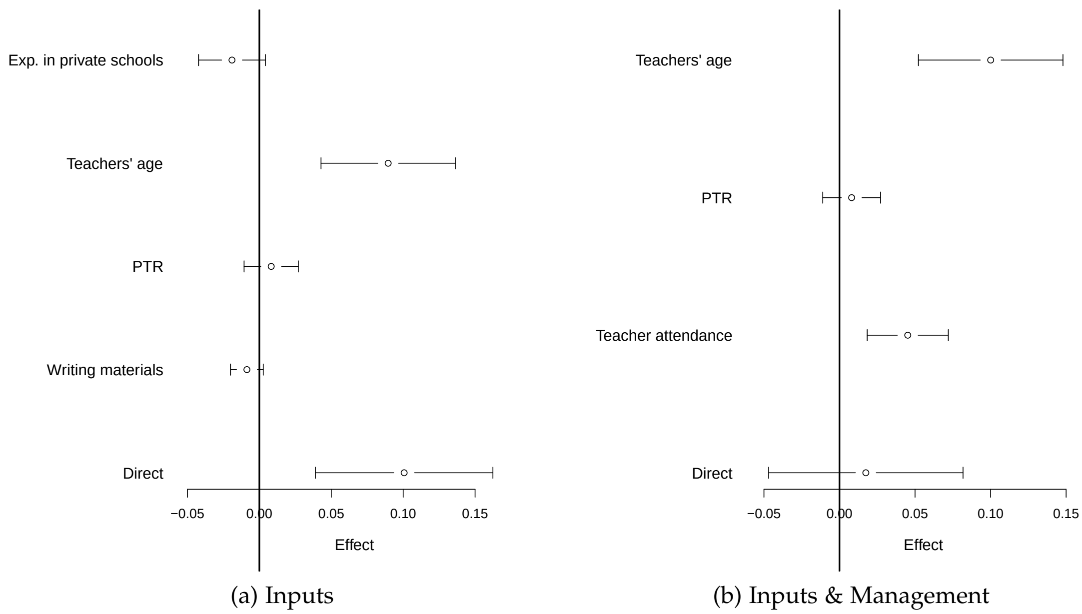
Column 1 replicates the results from Table 3 and columns 2 and 3 include only raw inputs. Columns 4 and 5 include raw inputs and the use of these inputs. Column 2 and column 4 only include mediators selected via Lasso, and columns 3 and 5 include all the mediators.
Standard errors are clustered at the school level. The sample is the original treatment and control allocation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note that in Section 3 we estimated equation (4) for several mediators. Combining those results with the results from Table 10, we show in Figure 5 the mediation effect ($\beta_4 \times \theta_5$) for the intermediate outcomes selected by Lasso, as well as the direct effect (β_5). The left panel uses only raw inputs as mediators, while the right panel also includes changes in the use of inputs.

Approximately half of the overall increase (37.4%–53.5%) in learning appears to have been due to changes in the composition of teachers (measured by teacher's age, a salient characteristic of new teaching graduates). Once we allow changes in the use of inputs to act as mediators, teacher attendance accounts for 24.1% of the total treatment effect. Although changes to teacher composition make it impossible to claim that teacher attendance increases purely due to management changes, our estimates from Section 3.3.3 suggest that contractors are able to increase teacher attendance even if the pool of teachers is held constant. Finally, note that 41.6% of the total treatment effect is a residual (the direct effect) when we only control for changes in inputs, but this drops to 9.09% when we control for changes in the use of inputs.

In short, roughly half of the overall increase in learning appears to have been due to changes in the composition of teachers. Teacher attendance (which may reflect underlying managerial practice) explains much of the residual not explained by the younger, better-trained teachers. These results suggest that extra resources (new and younger teachers) are an important pathway to impact in the PSL program, but changes in management practices play an equally important role.

Figure 5: Direct and mediation effects



Note: Direct (β_5) and mediation effects ($\beta_4 \times \theta_5$) for the mediators selected via Lasso. Note that the point estimates within the same panel are directly comparable to each other. Point estimates and 90% confidence intervals are plotted. Panel 5a shows treatment effects allowing only change in inputs as mediators. Panel 5b shows treatment effects allowing change in inputs and in the use of inputs as mediators.

5 Contractor comparisons

The main results in Section 3 address the impact of the PSL program from a policy-maker's perspective, answering the question, "What can the Liberian government achieve by contracting out management of public schools to a variety of private organizations?" We now turn from that general policy question to the narrower programmatic question of measuring the impact of specific contractors.

5.1 Methodology: Bayesian hierarchical model

There are two hurdles to estimating contractor-specific treatment effects. First, the assignment of contractors to schools was not random, which resulted in (non-random) differences in schools and locations across contractors (see Appendix I for more details). While the estimated treatment effects for each contractor are internally valid, they are not comparable to each other without further assumptions. Second, the sample sizes for most contractors 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 contractors' schools will score better than others for reasons unrelated to PSL), as well as interactions of those characteristics with a treatment dummy to make sure we capture heterogeneous treatment effects that go beyond any differences in location/schools (to account for the fact that raising scores through PSL relative to the control group will be easier in some contexts than others). The covariates over which we control are chosen via Lasso and are those that predict high learning gains in the control group.

Because randomization occurred at the school level and some contractors are managing only four or five treatment schools, the experiment is under-powered to estimate their effects.[46] Additionally, since the "same program" was implemented by different contractors, it would be naïve to treat contractors' estimators as completly independent from each other.[47] We take a Bayesian approach to this problem, estimating a hierarchical model (Rubin, 1981) (see Gelman et al. (2014) and Meager (2016) for a recent discussion). Intuitively, by allowing dependancy across contractors' treatment effects, the model "pools power" across contractors, and in the process pulls estimates for smaller contractors toward the overall average (a process known as "shrinkage"). The results of the Bayesian estimation are a weighted average of contractors' own performance and average performance across all contractors, where the proportions depend on the contractor's sample size.[48] We apply the Bayesian estimator after adjusting for baseline school differences and estimating the treatment effect of each contractor on the average school in our sample.[49]

5.2 Baseline differences

As discussed in Section 2.2.1 and shown in Table A.1, PSL schools are not a representative sample of public schools. Furthermore, there is heterogeneity in school characteristics across contractors. This is unsurprising since contractors stated different preferences for locations and some volunteered to manage schools in more remote and marginalized areas. We show how the average school for each contractor differs from the average public school in Liberia in Table 11 (Table I.1 in Appendix I shows simple summary statistics for the schools of each contractor). We reject the null that contractors' schools have similar characteristics

[46] There are not enough schools per contractor to get reliable standard errors by clustering at the school level. Therefore, when comparing contractors we collapse the data to the school level.

[47] Note that in a frequentist framework treatment estimates for contractors are considered independent when compared to each other.

[48] This model assumes that the true treatment effect for each contractor 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 final estimated effect is the weighted average of contractors' own performance and average performance across all contractors. The "weight" given to the contractor's own performance depends on the contractor'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.

[49] Coincidentally, the textbook illustration of a Bayesian hierarchical model is the estimate of treatment effects for an education intervention run in eight different schools with varied results (Rubin, 1981; Gelman et al., 2014).

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 contractors by over 100 students. Most contractors have schools with better infrastructure than the average public school in the country, except for Omega and Stella Maris. Finally, it should be noted that 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 contractors' schools.

Table 11: Baseline differences between treatment schools and average public schools, by contractor

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

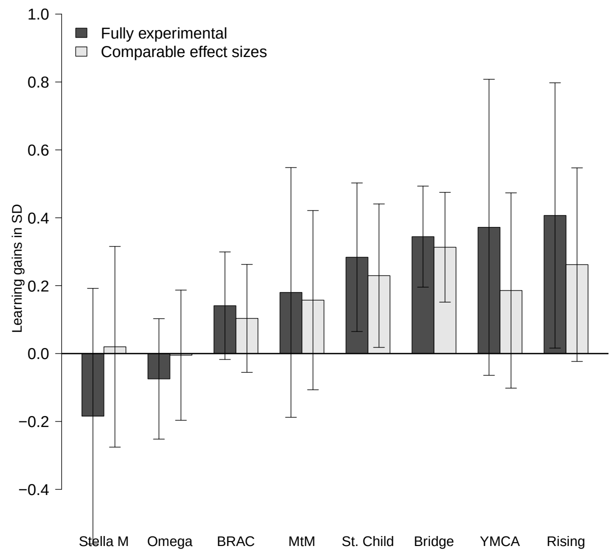
This table presents the difference between public schools and the schools operated by each contractor. The information for all schools is taken from the 2015/2016 EMIS data, and therefore is pre-treatment information. Column 9 shows the p-value for testing $H_0 : \beta_{BRAC} = \beta_{Bridge} = \beta_{YMCA} = \beta_{MtM} = \beta_{Omega} = \beta_{Rising} = \beta_{St.Child} = \beta_{StellaM}$. Standard errors are clustered at the school level. The sample is the original treatment and control allocation. Since some contractors had no schools with classes above the class caps, there is no data to estimate treatment effects over constrained classes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.3 Learning outcomes

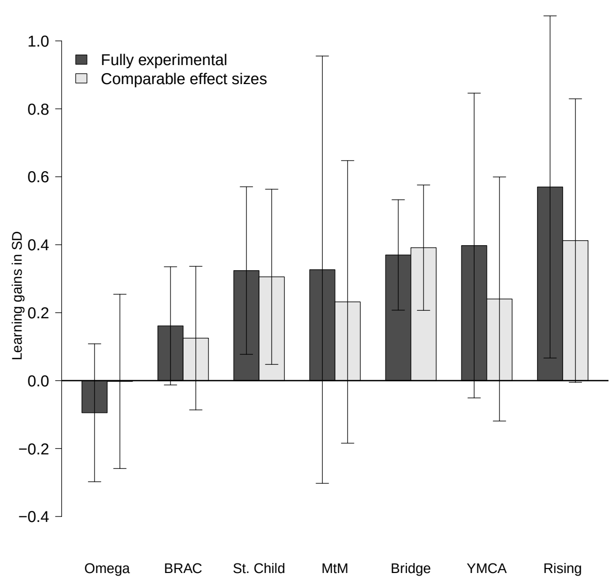
The raw treatment effects on test scores for each individual contractor shown in Figure 6 are internally valid, but not comparable. They are positive and significantly different from zero for three contractors: Rising Academies, Bridge International Academies, and Street Child. They are positive but statistically insignificant for Youth Movement for Collective Action, More Than Me, and BRAC. The estimates which we label as "comparable treatment effects" differ in two respects: They adjust for baseline differences and "shrink" the estimates for smaller contractors using the Bayesian hierarchical model. While the comparable effects are useful for comparisons, the raw experimental estimates remain cleaner for non-comparative statements (e.g., whether a contractor had an effect or not).

Intention-to-treat (ITT) treatment effects are shown in Figure 6a. That is, the effect over both compliers and non-compliers (i.e., over all students enrolled in a treatment school in 2015/2016, regardless of whether they attended an actual PSL school in 2016/2017). Treatment-on-the-treated (ToT) treatment effects are shown in Figure 6b. That is, the effect over compliers (i.e., students who actually attended a PSL school in 2016/2017). Non-compliance can happen either at the school level (if a contractor opted not to operate a school or the school did not meet the eligibility criteria), or at the student level (if the student no longer attends a treatment school). Comparable ITT treatment effects across contractors from the Bayesian hierarchical model are also shown in Panel A of Table 12. Table A.12 in Appendix A has the raw experimental treatment effects by contractor.

Figure 6: Treatment effects by contractor



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



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

Note: These figures show the raw, fully experimental treatment effects and the comparable treatment effects after adjusting for differences in school characteristics and applying a Bayesian hierarchical model. Figure 6a shows the intention-to-treat (ITT) effect, while Figure 6b shows the treatment-on-the-treated (ToT) effect. The ToT effects are larger than the ITT effects due to contractors replacing schools that did not meet the eligibility criteria, contractors refusing schools, or students leaving PSL schools. Note that Stella Maris had full non-compliance at the school level and therefore there is no ToT effect for this contractor.

There is considerable heterogeneity in the results. The data suggest contractors' learning impacts fall into three categories, based on a k-means clustering algorithm. In the first group, YMCA, Rising Academies, Street Child, and Bridge International Academies generated an increase in learning of 0.27σ across all subjects. In the second group, BRAC and More than Me generated an increase in learning of 0.15σ. In the third group, consisting of Omega and Stella Maris,[50] estimated learning gains are on the order of 0.01σ, and indistinguishable from zero in both cases. Below we explore whether "top-performing" contractors were more likely to behave in ways that might impose negative externalities on the broader education system, and hence whether better performance came at a cost to the education system as a whole.[51]

5.4 Non-learning outcomes and contracting flaws

Economists typically approach outsourcing in a principal-agent framework: A government (the principal) seeks to write a complete contract defining the responsibilities of the private contractor (the agent). This evaluation is part of that effort. In real-world settings, contracts are inevitably incomplete. It is impossible to pre-specify every single action and outcome that a private contractor must concern themselves with when managing a school. Economists have offered a number of responses to contractual incompleteness. One approach focuses on fostering competition among contractors via the procurement process and parental choice (Hart, Shleifer, & Vishny, 1997). Another, more recent approach puts greater focus on the identity of the contractors, on the premise that some agents are more "mission motivated" than others (Besley & Ghatak, 2005; Akerlof & Kranton, 2005). If contractors have intrinsic motivation and goals that align with the principal's objectives then they are unlikely to engage in pernicious behavior. This may be the case for non-profit contractors whose core mission is education. In the particular case of Liberia, this may also be true for for-profit contractors who are eager to show their effectiveness and attract investors and philanthropic donors. But, if contractors define their objectives more narrowly than the government, they may neglect to pursue certain government goals.

We examine three indicators illustrating how contractors and government goals may diverge under PSL: contractors' willingness to manage any school (as opposed to the best schools); contractors' willingness to work with existing teachers and improve their pedagogical practices and behavior (as opposed to having the worst performing teachers transferred to other public schools, imposing a negative externality on the broader school system); and contractors' commitment to improving access to quality education (rather than learning gains for a subset of pupils). In short, we're concerned with contractors rejecting "bad" schools, "bad" teachers, and excess pupils.

We already studied school selection in Section 5.2. To measure teacher selection, we study the number of teachers dismissed and the number of new teachers recruited (Table 12 - 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 contractor (Bridge International

[50] Non-compliance likely explains the lack of effect for these two contractors. Stella Maris never took control of its assigned schools, and Omega had not taken control of all its schools by the end of the school year. Our teacher interviews reflect these contractors' absence: in 3 out of four Stella Maris schools, all of the teachers reported that no one from Stella had been at the school in the previous week, and in 6 out of 19 Omega schools all of the teachers reported that no one from Omega had been at the school in the previous week.

[51] We had committed in pre-analysis to compare for-profit to non-profit contractors. This comparison yields no clear patterns.

Academies), while successful lobbying for additional teachers was common across several contractors. Although weeding out bad teachers is important, a reshuffling of teachers is unlikely to raise average performance in the system as a whole.

While enrollment increased across all contractors, the smallest treatment effect on this margin is for Bridge, which is consistent with that contractor being the only one enforcing class size caps (see Figure A.5 in Appendix A for additional details). As shown above, in classes where class-size caps were binding (10% of all classes holding 30% of students at baseline), enrollment fell by 20 students per grade. This issue was mostly restricted to Bridge International Academies (see Panel C, Table 12).

Table 12: Comparable ITT treatment effects by contractor

	(1) BRAC	(2) Bridge	(3) YMCA	(4) MtM	(5) Omega	(6) Rising	(7) St. Child	(8) Stella M	(9) p-value
Panel A: Student test scores									
English (standard deviations)	0.16*	0.25***	0.25	0.20	0.05	0.25	0.23*	0.07	0.089
	(0.09)	(0.09)	(0.16)	(0.15)	(0.11)	(0.16)	(0.12)	(0.17)	
Math (standard deviations)	0.07	0.33***	0.13	0.13	-0.01	0.24	0.22*	-0.00	0.025
	(0.10)	(0.10)	(0.18)	(0.16)	(0.11)	(0.18)	(0.13)	(0.19)	
Composite (standard deviations)	0.10	0.31***	0.19	0.16	-0.00	0.26	0.23*	0.02	0.033
	(0.10)	(0.10)	(0.17)	(0.16)	(0.12)	(0.17)	(0.13)	(0.18)	
Panel B: Changes to the pool of teachers									
% teachers dismissed	-8.84	50.36***	12.37	14.01	-5.22	1.43	-2.21	-7.02	<0.001
	(6.35)	(7.01)	(12.87)	(11.28)	(6.65)	(11.91)	(8.98)	(12.80)	
% new teachers	38.60***	70.87***	35.93*	48.36***	23.44**	20.68	37.05**	-8.55	0.0027
	(11.04)	(12.89)	(20.77)	(18.74)	(11.77)	(20.28)	(14.95)	(25.84)	
Age in years (teachers)	-5.53***	-9.10***	-3.46	-7.63***	-5.79***	-7.99***	-6.53***	-5.92**	0.12
	(1.70)	(2.17)	(3.56)	(2.55)	(1.72)	(2.74)	(2.08)	(2.70)	
Test score in standard deviations (teachers)	0.12	0.25*	0.06	0.23	0.18	0.17	0.23	0.17	0.47
	(0.13)	(0.14)	(0.23)	(0.18)	(0.13)	(0.18)	(0.16)	(0.18)	
Panel C: Enrollment and access									
Δ Enrollment	26.54	5.92	22.38	11.56	25.13	16.94	22.28	19.78	0.91
	(25.45)	(26.65)	(33.05)	(32.22)	(25.70)	(31.95)	(28.74)	(32.76)	
Δ Enrollment (constrained grades)	52.93	-45.78***	–	–	-22.11	52.21	18.40	–	0.0031
	(39.15)	(11.55)	(–)	(–)	(39.47)	(39.25)	(51.82)	(–)	
Student attendance (%)	18.04***	10.75*	15.86**	21.78**	11.07*	19.04**	18.21***	13.39	0.22
	(5.43)	(5.69)	(8.08)	(8.81)	(5.74)	(8.32)	(6.81)	(8.16)	
% of students still in any school	-1.23	1.26	-1.99	-3.33	-0.98	-2.51	-1.02	-1.96	0.80
	(3.38)	(3.52)	(5.21)	(5.61)	(3.56)	(5.22)	(4.20)	(5.05)	
% of students still in same school	0.52	2.48	0.14	0.40	0.72	0.76	0.29	0.30	0.82
	(1.79)	(1.94)	(2.85)	(2.59)	(1.88)	(2.58)	(2.26)	(2.64)	
Panel D: Satisfaction									
% satisfied with school (parents)	11.90*	11.60*	8.36	3.10	1.66	2.02	-0.84	9.82	0.19
	(6.34)	(6.47)	(9.38)	(8.72)	(6.28)	(9.29)	(8.49)	(9.41)	
% students that think school is fun	4.17	2.83	4.80	2.74	3.40	3.57	2.84	0.24	0.58
	(3.89)	(3.59)	(5.95)	(5.43)	(3.99)	(5.50)	(4.60)	(6.51)	
Observations	40	45	8	12	38	10	24	8	

This table presents the ITT treatment effect for each contractor, after adjusting for differences in baseline school characteristics, based on a Bayesian hierarchical model. Thus, this number should be interpreted as the difference between treatment and control schools, not as the mean in treatment schools. Column 9 shows the p-value for testing $H_0 : \beta_{BRAC} = \beta_{Bridge} = \beta_{YMCA} = \beta_{MtM} = \beta_{Omega} = \beta_{Rising} = \beta_{St.Child} = \beta_{StellaM}$. Some operators had no schools with class sizes above the caps. Standard errors are shown in parentheses. Estimation is conducted on collapsed, school-level data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

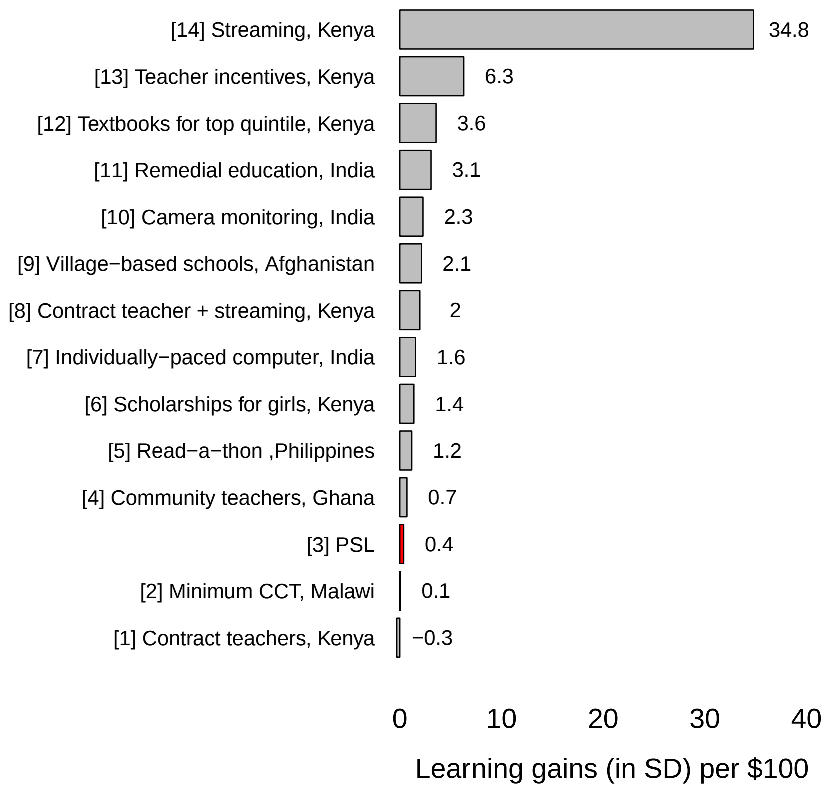
6 Cost-effectiveness analysis

From a policy perspective, the relevant question is not only whether the PSL program had a positive impact (especially given its bundled nature), but rather whether it is the best use of scarce funds. Cost-effectiveness analysis compares programs designed to achieve a common outcome with a common metric — in this case learning gains — by their cost per unit of impact. Inevitably, this type of analysis requires a host of assumptions, which must be tailored to a given user and policy question (see Dhaliwal, Duflo, Glennerster, and Tulloch (2013) for a review). Section 2.1.4 outlined various assumptions behind the cost estimates for each contractor.

Given the contested nature of these assumptions and the difficulty of modeling the long-term unit cost of PSL in a credible way, we opt to present only basic facts here. We encouraged operators to publish their *ex post* expenditure data in the same repository as our survey data, and some have agreed to do this. Readers who are willing to make stronger assumptions than we are about the cost structure of different contractors (e.g., separating out fixed and variable costs), are welcome to do cost-effectiveness calculations for each contractor using this data.

Instead, we perform a single cost-effectiveness calculation assuming a cost of $50 per pupil (the lowest possible cost associated with the program). Given that the ITT treatment effect is .19σ, test scores increased 0.38σ per $100 spent. Taking these estimates at face value suggests that so far in its first year PSL is not a cost-effective program for raising learning outcomes. We acknowledge that many education interventions have either zero effect or provide no cost data for cost-effectiveness calculations (Evans & Popova, 2016). However, a review by Kremer et al. (2013) of other interventions subject to experimental evaluation in developing countries highlights various interventions that yield higher per-dollar gains than PSL (see Figure 7).

Figure 7: Cost per child and treatment effects for several education interventions



Note: Figures show the learning gains per 100 (2011) USD. For more details on the calculations for [1], [2], [5]-[14] see https://www.povertyactionlab.org/policy-lessons/education/increasing-test-score-performance. Data for [4] is taken from Kiessel and Duflo (2014). The original studies of each intervention are as follows: [1],[8],[14] Duflo, Dupas, and Kremer (2011); Duflo et al. (2015); [2] Baird, McIntosh, and Özler (2011); [5] Abeberese, Kumler, and Linden (2014); [6] Kremer, Miguel, and Thornton (2009); [7] and [11] Banerjee et al. (2007); [9] Burde and Linden (2013); [10] Duflo, Hanna, and Ryan (2012); [12] Glewwe, Kremer, and Moulin (2009); [13] Glewwe, Ilias, and Kremer (2010).

However, it is unclear whether cost-effectiveness calculations from other contexts and interventions are relevant to the Liberian context and comparable to our results. First, test design is crucial to estimates of students' latent ability (and thus to treatment effects on this measure).[52] Since different interventions use different exams to measure students' ability, it is unclear that the numerator in these benefit-cost ratios is comparable.[53] The second problem is external validity. Even if treatment estimates were comparable across settings, treatment effects probably vary across contexts. This does not mean we cannot learn from different programs around the world, but implementing the same program in different settings is unlikely to yield identical results everywhere. Finally, the cost of implementing a program *effectively* (the denominator) is also likely to be variable across settings.

An important feature of our experiment is its real-world setting, which may increase the likelihood that gains observed in this pilot could be replicated at a larger scale. Previous research has shown that inter-

[52]For example, Table A.5 shows how PSL treatment estimates vary depending on the measure of students' ability we use.
[53]For more details, see Singh (2015a)'s discussion on using standard deviations to compare interventions.

ventions successfully implemented by motivated non-profit organizations (NGO) often fail when implemented at scale by governments (e.g., see Banerjee, Duflo, and Glennerster (2008); Bold, Kimenyi, Mwabu, Ng'ang'a, and Sandefur (2013); Dhaliwal and Hanna (2014); Kerwin and Thornton (2015); Cameron and Shah (2017)). The public-private partnership is designed to bypass the risk of implementation failure when taken up by the government, simply because the government is never the implementing agency. Note, however, that the program may still fail if the government withdraws support or removes all oversight.

7 Conclusions and policy discussion

To reiterate our main findings, random assignment of a school to the PSL program led to higher test scores by a margin of .18σ for English (p-value < 0.001) and .18σ for math (p-value < 0.001) on average across all schools and operators. Teachers in PSL schools were 20 percentage points more likely to be in school during a random spot check (from a base of 40% in control schools) and 16 percentage points more likely to be engaged in instruction during class time (from a base of 32% in control schools). Other indicators of welfare also improved: the program increased parental satisfaction with their school by 7.4 percentage points (p-value .022), and increased the share of children who say school is "fun" by 5.7 percentage points (p-value .022).

We explore three basic mechanisms for these effects: more material resources, more and better-qualified teachers, and better management. All three clearly improved. The evaluation is not designed to disentangle these mechanisms experimentally, but non-experimental analysis can offer some clues. Our estimates suggest additional books, supplies, and other inputs had little impact. Roughly half of the overall increase in learning appears to have been due to changes in the composition of teachers. Teacher attendance (which may reflect underlying managerial practice) explains much of the residual not explained by the younger, better-trained teachers.

Making the leap from empirical findings into policy recommendations involves considerable subjective judgment. We leave that task to the Ministry of Education. Instead, we focus here on translating our results into clear policy *challenges* — i.e., places where the empirical results do not conform to the goals of the project as we have understood them — without prescribing specific solutions.

1. The program risks becoming too expensive for the government to contemplate, even in optimistic fiscal scenarios. Our understanding is that the Ministry of Education sought to pilot a program at a cost of $50 per pupil (roughly a doubling of the primary education budget), under the assumption this was a feasible financing target for the government in the medium term. Even if we ignore the high expenditures budgeted by some contractors in the first year (on the assumption these incorporate once-off sunk costs as well as fixed costs that will not grow with scale), the costs of PSL accruing to the Ministry have crept up higher than advertised, and they are anticipated to rise even higher in the second year. Pupil-teacher ratios are lower in PSL schools, so the cost of teachers per pupil is higher for the program than the Ministry at large, and PSL schools have received additional infrastructure support from the Ministry. In year two, there are plans to raise the per-pupil subsidy above $50, and to include an additional meal program not covered by contractors (Ministry of Education - Republic of Liberia, 2017b).

2. The program concentrates this influx of new funding on a narrow set of schools that were already advantaged at baseline (i.e., schools were chosen because they had above-average resources and infrastructure: .97 more teachers and 1.4 more classrooms than the average public school). In the second year, the Ministry has sought to improve the equitable targeting of the program by concentrating new PSL schools in the Southeast, one of the most deprived parts of the country. Within the Southeast, however, the list of schools announced for PSL's expansion is again better-equipped and better-staffed than the region as a whole (1.3 more teachers and .93 more classrooms than the average public school in the region). The hypothesis that the PSL model can scale up and generate learning gains even in the most disadvantaged Liberian schools remains untested.

3. In places where schools were already oversubscribed, PSL has led to exclusion of students from the primary school where they would otherwise be enrolled. In the first year, PSL operators took over management of some schools whose enrollment levels per grade were already above the class-size caps agreed under PSL, and in some cases these schools were previously running double shifts to accommodate more students (something many PSL operators were unwilling to do). The result was easy to foresee. PSL led to reduced enrollment in these over-subscribed schools by 20 pupils per class (p-value .032). In a small number of schools where this phenomenon was most pronounced, several hundred students were removed, as documented in local press reports (Senah, 2016; Mukpo, 2017) and confirmed in our data. To date, there has been no public comment or statement on plans to avoid repeating this experience in future years.

4. The allocation of both higher quantities and higher quality teachers to PSL schools — without expanding the overall pool of teachers — has led to a situation in which some of the program's gains come at the direct expense of other schools. As noted above, PSL led to the assignment of 2.6 additional teachers per school and 1.1 additional teachers exiting per school. Large-scale dismissal of teachers was unique to one operator (Bridge International Academies), but successful lobbying for additional teachers was common across several operators. This contradicts a core selling point of PSL as we understood it, which was that the program would improve the management and training of government teachers, rather than replace them. Although weeding out bad teachers is important, a reshuffling of teachers is unlikely to raise average performance in the system as a whole.

5. Modest user fees persist. The Ministry of Education instructed operators that PSL schools should be free to parents, with no fees at any level, including early-childhood education (ECE) for which fees are normally permitted in government schools. In practice, parents in control schools report paying \$8 per pupil, and PSL reduces these amounts by \$2.9 (p-value < 0.001). It is not clear that PSL contractors play any direct role in the levying of fees; nevertheless, the existence of fees in PSL schools undermines one of the key marketing points of the program.

6. While the experiment cannot make strong claims about which operator behaviors led to particularly good or bad learning outcomes, we show that the two operators with the lowest treatment effects on learning (Omega and Stella Maris), were also largely absent from their schools for much or all of the year. In 3 out of 4 Stella Maris schools and 6 out of 19 Omega schools, all of the teachers reported that no one from their respective contractor had been at the school in the previous week. Several of

the other issues raised above—i.e., items (1), (2), (3), and (4)—are particularly acute in the case of Bridge International Academies. Given these results, it is notable that both Stella Maris and Bridge have been assigned new schools in the second year of the program.

One interpretation of our results is that contracting rules matter. In retrospect, there were some obvious gaps in the first-year contracts that the government might consider revising in light of the challenges above. For instance, contracts might forbid class-size caps or simply 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, to avoid creating more "ghost teachers" (on the payroll, but not functionally employed) as appears to have happened in the first year. Still, contracts are by nature incomplete, and the mechanisms to select contractors and weed out those who perform badly may matter more than refining the details of the contract.

Fixing the contracts and procurement process is not just a question of technical tweaks; it reflects a key governance challenge for the program. In year one, six of the eight contractors successfully passed through a competitive procurement process. One operator (Stella Maris) did not complete contracting, did little work, and produced low learning gains. Another (Bridge International Academies) was selected outside the competitive process, produced strong learning gains, but removed the majority of teachers and displaced some students. This underlines the importance of uniform contracting rules and competitive bidding in a public-private partnership such as this.

There is solid evidence of positive effects for Liberian children from partnership schools during the first year of this three-year evaluation, and there is potential to improve performance through successive iterations in the remaining two years. The program should aim to demonstrate it can work in average Liberian schools, with sustainable budgets and staffing levels, and without negative side-effects on other schools—while continuing to produce the significant improvements in teaching, learning, school attendance, and student and parent satisfaction documented here.

References

Abeberese, A. B., Kumler, T. J., & Linden, L. L. (2014). Improving reading skills by encouraging children to read in school: A randomized evaluation of the sa aklat sisikat reading program in the philippines. *Journal of Human Resources*, *49*(3), 611–633.

Agor, W. H. (1989). Intuition & strategic planning: How organizations can make. *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. doi: 10.1257/0895330053147930

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. (2008). A dime a day: The possibilities and limits of private schooling in Pakistan. *Comparative Education Review*, *52*(3), 329–355.

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. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.20140774 doi: 10.1257/aer.20140774

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.

Angrist, J. D., & Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, *114*(2), 533–575.

Angrist, J. D., Lavy, V., Leder-Luis, J., & Shany, A. (2017). *Maimonides rule redux* (Tech. Rep.). National Bureau of Economic Research.

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics*, *131*(3), 1415–1453.

Baird, S., McIntosh, C., & Özler, B. (2011). Cash or condition? evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, *126*(4), 1709–1753.

Banerjee, A. V., 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. Retrieved from http://qje.oxfordjournals.org/content/122/3/1235.abstract doi: 10.1162/qjec.122.3.1235

Banerjee, A. V., 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.

BBC Africa. (2016). *Liberia – the country that wants to privatise its primary schools.* Retrieved 1/06/2017, from http://www.bbc.com/news/world-africa-36074964

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.

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

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. doi: 10.1111/ecoj.12267

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.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2013). *Scaling up what works: experimental evidence on external validity in kenyan education.* (Mimeo)

Brault, M. (2011). *School-aged children with disabilities in us 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.

Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises iq scores. *Proceedings of the National Academy of Sciences*, *109*(2), 425-430. doi: 10.1073/pnas.1106077109

Bruns, B., & Luque, J. (2014). *Great teachers: How to raise student learning in latin america and the caribbean*. World Bank Publications.

Buhl-Wiggers, J., Kerwin, J., Smith, J., & Thornton, R. (2017). *The impact of teacher effectiveness on student learning in africa.* (mimeo)

Bullock, J. G., & Ha, S. E. (2011). Mediation analysis is harder than it looks. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), (p. 959). Cambridge University Press.

Burde, D., & Linden, L. L. (2013). Bringing education to afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics*, *5*(3), 27–40.

Cameron, L., & Shah, M. (2017). *Scaling up sanitation: Evidence from an rct in indonesia.* (mimeo)

Crawfurd, L. (in press). School management in Uganda.
(Journal of African Economies)

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. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304387809000273 doi: http://dx.doi.org/10.1016/j.jdeveco.2009.03.004

de Ree, J., Muralidharan, K., Pradhan, M., & Rogers, H. (2015). *Double for nothing? experimental evidence on the impact of an unconditional teacher salary increase on student performance in indonesia* (Working Paper No. 21806). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w21806 doi: 10.3386/w21806

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.

Dhaliwal, I., & Hanna, R. (2014). *Deal with the devil: The successes and limitations of bureaucratic reform in india* (Tech. Rep.). National Bureau of Economic Research.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, *101*(5), 1739–74. doi: 10.1257/aer.101.5.1739

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.

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, *102*(4), 1241–78. doi: 10.1257/aer.102.4.1241

Evans, D., & Popova, A. (2016). What really works to improve learning in developing countries? an analysis of divergent findings in systematic reviews. *The World Bank Research Observer*, *31*(2), 242–270.

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

Foreign Policy. (2016). *Liberia's education fire sale.* Retrieved 20/07/2016, from http://foreignpolicy.com/2016/06/30/liberias-education-fire-sale/

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. Retrieved from http://www.aeaweb.org/articles?id=10.1257/app.1.1.112 doi: 10.1257/app.1.1.112

Green, D. P., Ha, S. E., & Bullock, J. G. (2010). Enough already about "black box" experiments:

Studying mediation is more difficult than most scholars suppose. *The ANNALS of the American Academy of Political and Social Science*, *628*(1), 200-208. Retrieved from http://dx.doi.org/10.1177/0002716209351526 doi: 10.1177/0002716209351526

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.

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.

Heckman, J., & Pinto, R. (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews*, *34*(1-2), 6-31. Retrieved from http://dx.doi.org/10.1080/07474938.2014.944466 doi: 10.1080/07474938.2014.944466

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052-86. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.103.6.2052 doi: 10.1257/aer.103.6.2052

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, *15*(4), 309.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science*, *25*(1), 51–71.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, *108*(25), 10081-10086. doi: 10.1073/pnas.1103228108

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An introduction to statistical learning with applications in r.* (Springer)

Kerwin, J. T., & Thornton, R. (2015). *Making the grade: Understanding what works for teaching literacy in rural uganda.* (mimeo)

Kiessel, J., & Duflo, A. (2014). *Cost-effectiveness report: The teacher community assistant initiative (TCAI).* Retrieved 6/08/2017, from http://www.poverty-action.org/sites/default/files/publications/TCAI_Cost-Effectiveness_2014.3.26.pdf

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.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, *340*(6130), 297–300. Retrieved from http://science.sciencemag.org/content/340/6130/297 doi: 10.1126/science.1235350

Kremer, M., Miguel, E., & Thornton, R. (2009). Incentives to learn. *The Review of Economics and Statistics*, *91*(3), 437–456.

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

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects.

The Review of Economic Studies, *76*(3), 1071–1102.

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

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.

Logistics Capacity Assessment - Wiki. (2016). *Liberia road network.* Retrieved 09/07/2017, from http://dlca.logcluster.org/display/public/DLCA/2.3+Liberia+Road+Network;jsessionid=B5C4A37554B63CF040F07A185B5AD3FE

Mail & Guardian Africa. (2016a). *An Africa first! Liberia outsources entire education system to a private American firm. why all should pay attention.* Retrieved 20/07/2016, from http://mgafrica.com/article/2016-03-31-liberia-plans-to-outsource-its-entire-education-system-to-a-private-company-why-this-is-a-very-big-deal-and-africa-should-pay-attention

Mail & Guardian Africa. (2016b). *An update on bridge academies in Liberia, and why people need dreams - and yes, sweet lies - too.* Retrieved 20/07/2016, from http://mgafrica.com/article/2016-05-07-an-update-on-bridge-academies-in-liberia-and-why-people-need-dreams-and-yes-sweet-lies-too

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., Rajani, R., & Manda, C. (2017). *Inputs, incentives, and complementarities in primary education: Experimental evidence from tanzania.* (Mimeo)

Meager, R. (2016). *Aggregating distributional treatment effects: A bayesian hierarchical analysis of the microcredit literature* (Tech. Rep.).

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

Mukpo, A. (2017). *In Liberia, a town struggles to adjust to its new charter school.* Retrieved 28/07/2017, from https://gemreportunesco.wordpress.com/2017/04/12/in-liberia-a-town-struggles-to-adjust-to-its-new-charter-school

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.

OHCHR. (2016). *Un rights expert urges Liberia not to hand public education over to a private company.* Retrieved 1/06/2017, from http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=18506

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

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

Schermerhorn, J., Osborn, R., Uhl-Bien, M., & Hunt, J. (2011). *Organizational behavior*. Wiley. Retrieved from https://books.google.com/books?id=8eRtuZeIguIC

Senah, G. (2016). *At Kendeja Public School, more than 300 students left unenrolled.* Retrieved 28/07/2017, from http://www.bushchicken.com/at-kendeja-public-school-more-than-300-students-left-unenrolled

Singh, A. (2015a). *How standard is a standard deviation? a cautionary note on using sds to compare across impact evaluations in education.* Retrieved 31/07/2017, from http://blogs.worldbank.org/impactevaluations/how-standard-standard-deviation-cautionary-note-using-sds-compare-across-impact-evaluations

Singh, A. (2015b). 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)

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.

The New York Times. (2016). *Liberia, desperate to educate, turns to charter schools.* Retrieved 20/07/2016, from http://www.nytimes.com/2016/06/14/opinion/liberia-desperate-to-educate-turns-to-charter-schools.html

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

The World Bank. (2015a). *Gdp per capita (current us£).* (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/NY.GDP.PCAP.CD)

The World Bank. (2015b). *Government expenditure per student, primary (% of gdp per capita).* (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.XPD.PRIM.PC.ZS)

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&=

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

Vox World. (2016). *Liberia is outsourcing primary schools to a startup backed by mark zuckerberg.* Retrieved 20/07/2016, from http://www.vox.com/2016/4/8/11347796/liberia-outsourcing-schools

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/

World Bank. (2015). *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. (2017). *Doing business 2017: Equal opportunity for all* (Tech. Rep.).

World development report. (2004). The World Bank. doi: 10.1596/0-8213-5637-2