Outsourcing Service Delivery in a Fragile State: Experimental Evidence from Liberia

Mauricio Romero[*]   Justin Sandefur[†]   Wayne Aaron Sandholtz[‡]

April 1, 2019[§]

Abstract

In 2016, the Liberian government delegated management of 93 randomly-selected public schools to private providers. Providers received USD 50 per pupil, on top of USD 50 per pupil annual expenditure in control schools. After one academic year, students in outsourced schools scored 0.18σ higher in English and mathematics. Teachers in outsourced schools were 50% more likely to be in school and 43% more likely to be engaged in instruction during class time. We do not find evidence of heterogeneity in learning gains or enrollment across students. We do find significant heterogeneity across providers, both in learning gains and other unintended consequences of outsourcing.

Keywords: Public-Private Partnership; Randomized Controlled Trial; School Management
JEL Codes: I25, I28, C93, L32, L33

1   Introduction

Governments often enter into public-private partnerships as a means to raise capital or to leverage the efficiency of the private sector (World Bank, 2015b). But contracts are inevitably incomplete, and thus contracting out the provision of public services to private providers will have theoretically ambiguous impacts on service quality (Hart, Shleifer, & Vishny, 1997; Holmstrom & Milgrom, 1991). While private contractors may face stronger incentives for cost efficiency than civil servants, they may cut costs through actions that are contractually permissible but not in the public interest.

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

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

The ITT effect on student test scores after one year of the program is 0.18σ for English and 0.18σ

1

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

These gains in test scores reflect a combination of additional inputs and improved management. Some providers spent more than 600 USD per student, while others spent as little as USD 50 per student. While some providers incurred larger costs in the first year, assuming all providers will eventually reach the budget target of USD 50 per pupil implies that the program can increase test scores in treated students by 0.44σ per USD 100 spent. We estimate this yields a positive net present value for the program investment after considering the income gains associated with schooling, but may render it less cost-effective than business-as-usual education policies (under certain assumptions discussed in Section 5). The program also increased management quality, as proxied by teacher time on task. Teachers in PSL schools were 50% more likely to be in school during a spot check (a 20-percentage-point increase, from a base of 40%) and 43% more likely to be engaged in instruction during class time (a 15-percentage point increase, from a base of 35%). Teacher attendance and time on task improved for incumbent teachers, which we interpret as evidence of better management.

Since each provider was assigned schools in a matched-pair design, we are able to estimate (internally valid) treatment effects for each provider. While the assignment of treatment within matched pairs was random, the assignment of pairs to providers was not, resulting in non-random differences in schools and locations across providers. Therefore, the raw treatment effects for each individual provider are internally valid but they are not comparable without further assumptions (see Section 4 for more details). In the Appendix, we also present treatment effects adjusting for baseline differences and "shrinking" the estimates using a Bayesian hierarchical model — with qualitatively similar results. While the highest-performing providers generated increases in learning of over 0.36σ, the lowest-performing providers had

---

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

no impact on learning. The group of highest-performing providers includes both the highest spender and some of the lowest-cost organizations. These results suggests that higher spending by itself is not necessary for improving learning outcomes.[2]

In addition to heterogeneity in test scores across providers, we also find heterogeneity in provider actions that are within the boundaries of the contract but may be socially undesirable. While the contract did not allow cream-skimming, it did not prohibit providers from expelling students from oversubscribed schools and shifting underperforming teachers to other schools.[3] While most providers kept students in oversubscribed schools and retained existing teachers, one provider did not. This provider, Bridge International Academies, removed pupils after taking control of schools with large class sizes, and removed 74% of incumbent teachers from its schools.[4] Additionally, teachers or staff of two PSL providers (More than Me Academy and Youth Movement for Collective Action) have been accused of sexual abuse since the start of the program, and an investigate report published in late 2018 alleged that More than Me Academy employed a serial child rapist in its schools prior to the start of PSL (Baysah, 2016; Young & Flynn, 2018). These events underscore the challenge of ensuring that private providers act in the public interest in a world of incomplete contracts (as outlined by Hart et al. (1997)).

We make several contributions to both research and policy. Proponents of outsourcing in education argue that combining public finance with private management has the potential to overcome a trade-off between efficiency and equity (Patrinos, Barrera-Osorio, & Guáqueta, 2009). On the efficiency side, private schools tend to be better managed than their public counterparts (N. Bloom, Lemos, Sadun, & Van Reenen, 2015; Muralidharan & Sundararaman, 2015). On the equity side, fee-charging private schools may increase inequality and induce socio-economic stratification in education (Hsieh & Urquiola, 2006; Lucas & Mbiti, 2012; Zhang, 2014). Thus, in theory, publicly-financed but privately-managed schools may increase efficiency without compromising equity. Most of the empirical evidence to date on outsourcing education comes from the U.S., where charter schools appear to improve learning outcomes when held accountable by a strong commissioning body (Cremata et al., 2013; Woodworth et al., 2017). However, there is limited

---

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

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

[4]Several providers enrolled more students than they were paid for. The Ministry allowed all providers to replace up to 40% of underperforming teachers, yet our results show no discernible effect on teacher exit rates for other providers. Differences in behavior with uniform contracts suggest differences in mission alignment (à la Besley and Ghatak (2005) or Akerlof and Kranton (2005)) may be important when outsourcing public services.

evidence on whether private administration of public schools can improve learning outcomes in developing countries, where governments tend to have limited capacity to write complete contracts and enforce them. Two noteworthy studies which examine close analogs to PSL in the U.S. are Abdulkadiroğlu, Angrist, Hull, and Pathak (2016) who study charter takeovers (where traditional public schools are restarted as charter schools, similar to our setting) in Boston and New Orleans and Fryer Jr (2014) who studies the implementation of a bundle of best practices from high-performing charter schools into low-performing, traditional public schools in Houston, Texas. In line with our results, both studies find increases in test scores. We provide some of the first experimental estimates on contracting out management of existing public schools in a developing country.[5]

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

Finally, we contribute to the broader literature on outsourcing service delivery. Hart et al. (1997) argue that the bigger the adverse consequences of non-contractible quality shading, the stronger the case for governments to provide services directly. Empirically, in cases where quality is easy to measure and to enforce, such as water services (Galiani, Gertler, & Schargrodsky, 2005), outsourcing seems to work. Similarly, for primary health care, where quality is measurable (e.g., immunization and antenatal care coverage), outsourcing improves outcomes in general (Loevinsohn & Harding, 2005; E. Bloom et al., 2007). In contrast, for services whose quality is difficult to measure, such as prisons (Useem & Goldstone, 2002; Cabral, Lazzarini, & de Azevedo, 2013), outsourcing seems to be detrimental. Contrary to primary health care, there is some evidence that contracting out advanced care (where quality is harder to measure) increases expenditure without increasing quality (Duggan, 2004). Some quality aspects of education are easy to measure (e.g., enrollment and basic learning metrics), but other are harder (e.g., socialization and selec-

---

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

tion). In our setting, while outsourcing management improves most indices of school quality on average, the effect varies across providers. In addition, some providers' actions likely generated negative spillovers for the broader education system, underscoring the importance of robust contracting and monitoring for this type of program.

2 Research design

2.1 The program

2.1.1 Context

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

A second broad feature of Liberia's education system, relevant for the PSL program, is its performance — not only are learning levels low, but access to basic education and progression through school remains inadequate. The Minister of Education has cited the perception that "Liberia's education system is in crisis" as the core justification for the PSL program (Werner, 2017). While the world has made great progress towards universal primary education in the past three decades (worldwide net enrollment was almost 90% in 2015), Liberia has been left behind. Net primary enrollment stood at only 38% in 2014 (World Bank, 2014). Low *net* enrollment is partially explained by an extraordinary backlog of over-age children due to the civil war (see Figure A.1 in Appendix A.1): The median student in early childhood

education is eight years old and over 60% of 15 years olds are still enrolled in early childhood or primary education (Liberia Institute of Statistics and Geo-Information Services, 2016). Learning levels are low: Only 25% of adult women (there is no information for men) who finish elementary school can read a complete sentence (Liberia Institute of Statistics and Geo-Information Services, 2014).

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 providers to run ninety-three existing public primary and pre-primary schools. There are nine grades per school: three early childhood education grades (Nursery, K1, and K2) and six primary grades (grade 1 - grade 6). Providers receive funding on a per-pupil basis. In exchange they are responsible for the daily management of the schools.

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

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

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

(about USD 38).

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

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

The intended differences between treated (PSL) and control (traditional public) schools are summarized in Table 1. First, PSL schools are managed by private organizations. Second, PSL schools are theoretically guaranteed (as per the contract) one teacher per grade in each school, plus extra funding. Third, private providers are authorized to cap class sizes. Finally, while both PSL and traditional public schools are free for primary students starting in first grade, public schools charge early-childhood education (ECE) fees.

[Table 1 about here.]

2.1.3 What do providers do?

Providers enjoy considerable flexibility in defining the intervention. They are free to choose their preferred mix of, say, new teaching materials, teacher training, and managerial oversight of the schools' day-to-day operations. Rather than relying on providers' own description of their model — where there may be incentives to exaggerate and activities may be defined in non-comparable ways across providers — we administered a survey module to teachers in treatment schools, asking if they had heard of the provider, and if so, what activities the provider had engaged in. We summarize teachers' responses in Figure 1,

---

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

which shows considerable variation in the specific activities and the total activity level of providers.

For instance, teachers reported that two providers (Omega and Bridge) provided computers to schools, which fits with the stated approach of these two international, for-profit firms. Other providers, such as BRAC and Street Child, put more focus on teacher training and observing teachers in the classroom, though these differences were not dramatic. In general, providers such as More than Me and Rising Academies showed high activity levels across dimensions, while teacher surveys confirmed administrative reports that Stella Maris conducted almost no activities in its assigned schools.

[Figure 1 about here.]

2.1.4 Cost data and assumptions

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

In the first year, providers spent far more than this amount. *Ex ante* per-pupil budgets submitted to the program secretariat before the school year started (on top of the Ministry's costs) ranged from a low of approximately USD 57 for Youth Movement for Collective Action to a high of USD 1,050 for Bridge International Academies (see Figure 2a). *Ex post* per-pupil expenditure submitted to the evaluation team at the end of the school year (on top of the Ministry's costs) ranged from a low of approximately USD 48 for Street Child to a high of USD 663 for Bridge International Academies (see Figure 2b). These differences in costs are large relative to differences in treatment effects on learning, implying that cost-effectiveness may be driven largely by cost assumptions.

In principle, the costs incurred by private providers would be irrelevant for policy evaluation in a public-private partnership with this structure. If the providers are willing to make an agreement in which the government pays USD 50 per pupil, providers' losses are inconsequential to the government

8

(philanthropic donors have stepped in to fund some providers' high costs under PSL).[7] Thus we present analyses using both the Ministry's USD 50 long-term cost target and providers' actual budgets.[8]

Providers' budgets for the first year of the program are likely a naïve measure of program cost, as they combine start-up costs, fixed costs, and variable costs. It is possible to distinguish start-up costs from other costs as shown in Figure 2, and these make up a small share of the first-year totals for most providers. It is not possible to distinguish fixed from variable costs in the budget data. In informal interviews, some providers (e.g., Street Child) profess operating a variable-cost model, implying that each additional school costs roughly the same amount to operate. Others (e.g., Bridge) report that their costs are almost entirely fixed, and unit costs would fall if scaled; however, we have no direct evidence of this. Our estimate is that Bridge's international operating cost, at scale, is between USD 191 and USD 220 per pupil annually.[9]

[Figure 2 about here.]

2.2 Experimental design

2.2.1 Sampling and random assignment

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

---

[7]Philanthropic money was raised by providers and given under the condition that schools be operated by private providers under the PSL program. These costs matter to the government under at least two scenarios. First, if providers are spending more during the first years of the program to prove effectiveness, they may lower expenditure (and quality) once they have locked in long-term contracts. Second, if private provider's are not financially sustainable, they may close schools and disrupt student learning.

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

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

school in Liberia has 1.8 permanent classrooms — the median school has zero permanent classrooms — while the average school in the RCT has 3.16 classrooms. Figure 3a shows all public schools in Liberia and those within our sample. Table A.1 in Appendix A.1 has details on the differences between schools in the experiment and other public schools.

[Figure 3 about here.]

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

Private providers did not manage all the schools originally assigned to treatment and we treat these schools as non-compliant, presenting results in an intention-to-treat framework. After providers visited their assigned schools to start preparing for the upcoming school year, two treatment schools turned out to be private schools that were incorrectly labeled in the government data as public schools. Two other schools had only two classrooms each. Of these four schools, two had originally been assigned to More Than Me and two had been assigned to Street Child. Omega Academies opted not to operate two of their assigned schools and Rising Academies opted not to operate one of their assigned schools. In total, there are 7 non-compliant treatment schools.[11] Figure 3b shows the treatment assignment.

---

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

[11]More than Me and Street Child were provided with replacement schools, presenting them with a new list of counterparts and informing them, as before, that they would operate one of each pair of schools (but not which one). Providers approved the list before we randomly assigned replacement schools from it. However, we do not use this list as our main sample since it is not fully experimental. We analyzed results for this "final" treatment and control school list, and they are almost identical to the results for the "original" list. Results for this final list of treatment and control schools are available upon request. Bridge International Academies is managing two extra demonstration schools that were not randomized and are not part of our sample. Rising Academies was given one non-randomly assigned school, which is not part of our sample either. Thus, the set of schools in our analysis is not identical to the set of schools actually managed by PSL providers. Table A.2 summarizes the overlap between schools in our main sample and the set of schools actually managed by PSL providers.

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

2.2.2 Timeline of research and intervention activities

We collected data in schools twice: At the beginning of the school year in September/October 2016 and at the end of the school year in May/June 2017.[12] We collected the first round of data 2 to 8 weeks after the beginning of treatment. While we intended the first survey wave to serve as a baseline, logistical delays led it to take place shortly after the beginning of the school year. We see evidence of treatment effects within this 1-2 month time frame and treat this early wave as a very short-term outcome survey. Hence, we do not control for test scores collected during the first wave of data collection.[13] We focus on time-invariant covariates and administrative data collected before the program began when checking balance between treatment and control schools (see Section 2.2.5).

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. In addition, purely school-based tests would be contaminated by shifts in enrollment and attendance due to treatment. For the math part of the test we provided students with scratch paper and a pencil. We designed the tests to capture a wide range of student abilities. To make the test scores comparable across grades, we constructed a single adaptive test for all students. The test has stop rules

---

[12]A third round of data collection will take place in March/April 2019 conditional on continuation of the project and preservation of the control group (see Figure A.2 in Appendix A.1 for a detailed timeline of intervention and research activities).

[13]Our pre-analysis plan was written on the assumption we would be able to collect baseline data. Hence, the pre-analysis plan includes a specification that controls for test scores collected during the first wave of data collection along with the main specifications used in this paper. We report these results in Table A.4 in Appendix A.1. We view the differences in short-term outcomes as treatment effects rather than "chance bias" in randomization for the following reasons. First, time-invariant student characteristics are balanced across treatment and control (see Table 2). Second, the effects on English and math test scores appear to materialize in the later weeks of the fieldwork, as shown in Figure A.3. Third, there is no significant effect on abstract reasoning, which is arguably less amenable to short-term improvements through teaching (although the difference between a significant English/math effect and an insignificant abstract reasoning effect here is not itself significant).

that skip higher-order skills if the student is not able to answer questions related to more basic skills. Appendix A.3 has details on the construction of the test.

We estimate an item response theory (IRT) model for each round of data collection. IRT models are the standard in the assessments literature for generating comparative test scores.[14] There are two relevant characteristics of IRT models in this setting: First, they simultaneously estimate the test taker's ability and the difficulty of the questions, which allows the contribution of "correct answers" to the ability measure to vary from question to question. Second, they provide a comparable measure of student ability across different grades and survey rounds, even if the question overlap is imperfect. A common scale across grades allows us to estimate treatment effects as additional years of schooling. Following standard practice, we normalize the IRT scores with respect to the control group.

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

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

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

---

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

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

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, the first wave of data was collected 2 to 8 weeks after the beginning of treatment; hence, we focus on time-invariant characteristics when checking balance across treatment and control. Observable (time-invariant) characteristics of students and schools are balanced across treatment and control (see Table 2). Eighty percent of schools in our sample are in rural areas, over an hour away from the nearest bank (which is usually located in the nearest urban center), and over 10% need to hold some classes outside due to insufficient classrooms. Boys make up 55% of our students and the students' average age is 12. According to pre-treatment administrative data (EMIS), the number of students, infrastructure, and resources available to students were not statistically different across treatment and control schools (for details, see Table A.3 in Appendix A.1).

We took great care to avoid attrition: enumerators conducting student assessments participated in extra training on tracking and its importance, and dedicated generous time to tracking. Students were tracked to their homes and tested there when not available at school. Attrition in the second wave of data collection from our original sample is balanced between treatment and control and is below 4% (see Panel C). Appendix A.2 has more details on the tracking and attrition that took place during data collection.

[Table 2 about here.]

3 Experimental results

In this section, we first explore how the PSL program affected access to and quality of education. We then turn to mechanisms, looking at changes in material inputs, staffing, and school management.

3.1 Test scores

Following our pre-analysis plan, we report treatment-effect estimates from two specifications:

$$Y_{isg} = \alpha_g + \beta_1 treat_s + \varepsilon_{isg} \tag{1}$$

$$Y_{isg} = \alpha_g + \beta_2 treat_s + \gamma_2 X_i + \delta_2 Z_s + \varepsilon_{isg} \tag{2}$$

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

Table 3 shows results from student tests. The first three columns show differences between control and treatment schools' test scores after 1-2 months of treatment (September/October 2016), while the last three columns show the difference after 9-10 months of treatment (May/June 2017). Columns 1, 2, 4, and 5 show intention-to-treat (ITT) treatment estimates, while Columns 3 and 6 show treatment-on-the-treated (ToT) estimates (i.e., the treatment effect for students that actually attended a PSL school in 2016/2017). The ToT is estimated using the assigned treatment as an instrument for whether the student is in fact enrolled in a PSL school during the 2016/2017 academic year.[17]

After 1-2 months of treatment, student test scores increase by 0.05σ in math (p-value=0.09) and 0.07σ in English (p-value=0.04). Part of these short-term improvements can be explained by the fact that most providers started the school year on time, while most traditional public schools began classes 1-4 weeks later. Hence, most students were already attending classes on a regular basis in treatment schools during our field visit, while their counterparts in control schools were not. We estimate the treatment effect separately for students tested during the first and the second half of the first round of data collection (see Figure A.3 in Appendix A.1), and show that the treatment effects fade in during the course of field work — further supporting our conclusion that these results represent early treatment effects as opposed to baseline imbalance.

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

---

[16]These controls were specified in the pre-analysis plan and are listed in Table A.5. We had committed in the pre-analysis plan to a specification that controlled for pre-treatment individual outcomes:

$$Y_{isg} = \alpha_g + \beta_3 treat_s + \gamma_3 X_i + \delta_3 Z_s + \zeta_3 Y_{isg,-1} + \varepsilon_{isg} \tag{3}$$

However, as mentioned before, the first wave of data was collected after the beginning of treatment, so we lack a true baseline of student test scores. We report this specification in Table A.4 in Appendix A.1. The results are still statistically significant, but mechanically downward biased.

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

14

different measures of student ability (see Table A.6 in Appendix A.1 for details).

[Table 3 about here.]

An important concern when interpreting these results is whether they represent real gains in learning or better test-taking skills resulting from "teaching to the test". We show suggestive evidence that these results represent real gains. First, the treatment effect is significant (.19σ, p-value < 0.001) for new modules that were not in the first wave test (and unknown to the providers or the teachers), and statistically indistinguishable from the treatment effect over all the items (.18σ, p-value < 0.001). Second, the treatment effect is positive and significant (.12σ, p-value .0014) for the conceptual questions (which do not resemble the format of standard textbook exercises). We cannot rule out that providers narrowed the curriculum by focusing on English and mathematics or, conversely, that they generated additional learning gains in other subjects that we did not test.[18] We find no evidence of heterogeneous treatment effects by students' socio-economic status, gender, or grade (see Table A.7 in Appendix A.1).

3.2 Enrollment, attendance, and student selection

The previous section showed that education quality, measured using test scores in an ITT framework, increases in PSL schools. We now ask whether the PSL program increases access to education. To explore this question we focus on three outcomes which were committed to in the pre-analysis plan: Enrollment, student attendance, and student selection. PSL increased enrollment overall, but in schools where enrollment was already high and classes were large, the program led to a significant decline in enrollment. This does not appear to be driven by selection of "better" students, but by providers capping class sizes and eliminating double shifts.[19] As shown in Section A.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 .094) more students on average before treatment. Table A.3 uses administrative data, while Table 4 uses data independently collected by our survey teams. While the difference in enrollment in the 2015/2016

---

[18]As shown in Table 7 PSL schools have longer school days. As a result, treatment schools spend about 45 minutes per week more in both English and math. However, they do not spend a larger fraction of the school day in English or math (see Table A.17). More broadly, we cannot rule out that PSL spent disproportionately more resources improving English and Math instruction.

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

academic year is only significant in the latter, the point estimates are similar across both tables. Second, PSL schools on average have 57 (p-value < 0.001) more students than control schools in the 2016/2017 academic year, which results in a net increase (after controlling for pre-treatment differences) of 25 (p-value .088) students per school.

Since provider compensation is based on the number of students enrolled rather than the number of students actively attending school, increases in enrollment may not translate into increases in student attendance. An independent measure of student attendance conducted by our enumerators during a spot check shows that students in treatment schools are 16 (p-value < 0.001) percentage points more likely to be in school during class time (see Panel A, Table 4).

Turning to the question of student selection, we find no evidence that any group of students is systematically excluded from PSL schools. The proportion of students with disabilities is not statistically different in PSL schools and control schools (Panel A, Table 4).[20] Among our sample of students (i.e., students sampled from the 2015/2016 enrollment log), students are equally likely across treatment and control to be enrolled in the same school in the 2016/2017 academic year as they were in 2015/2016, and equally likely to be enrolled in any school (see Panel B, Table 4). Finally, selection analysis using student-level data on wealth, gender, and age finds no evidence of systematic exclusions (see Table A.8 in Appendix A.1).

[Table 4 about here.]

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

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

---

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

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

Enrollment in constrained school-grades decreases, while enrollment in unconstrained school-grades increases (see Column 1 in Table 5). Thus, schools far below the cap have positive treatment effects on enrollment and schools near or above the cap offset it with declining enrollment. Our student data reveal this pattern as well: Columns 2 and 3 in Table 5 show the ITT effect on enrollment depending on whether students were enrolled in a constrained class in 2015/2016. In unconstrained classes students are more likely to be enrolled in the same school (and in any school). But in constrained classes students are less likely to be enrolled in the same school. While there is no effect on overall school enrollment, switching schools may be disruptive for children (Hanushek, Kain, & Rivkin, 2004). Finally, test-scores improve for students in constrained classes. This result is difficult to interpret as it includes the positive treatment effect over students who did not change schools (compounded by smaller class sizes) with the effect over students removed from their schools. These results are robust to excluding adjacent grades from the "constrained" measure (see Table A.9 in Appendix A.1).

[Table 5 about here.]

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 in treatment schools (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, replace those teachers, and provide additional ones. Ultimately, the extra teachers result in lower pupil-teacher ratios (despite increased student enrollment). This re-shuffling of teachers means that PSL schools have younger and less-experienced teachers, who are

more likely to have worked in private schools in the past and have higher test scores (we conducted simple memory, math, word association, and abstract thinking tests). Replacement and extra teachers are recent graduates from the Rural Teacher Training Institutes (see King, Korda, Nordstrum, and Edwards (2015) for details on this program). While the program's contracts made no provisions to pay teachers differently in treatment and control schools, teachers in PSL schools report higher wages. If higher salaries for teachers in PSL schools are conditional on them working in program schools, then this would create an incentive to perform well. However, we could not find an explanation for these higher salaries. Hence it is unclear whether higher salaries are tied to the program. But large unconditional increases in teacher salaries have been shown elsewhere to have no effect on student performance in the short run (de Ree, Muralidharan, Pradhan, & Rogers, 2018).

Our enumerators conducted a "materials" check during classroom observations (See Panels C - Table 6). Since we could not conduct classroom observations in schools that were out of session during our visit, Table A.10 in Appendix A.1 presents Lee (2009) bounds on these treatment effects (control schools are more likely to be out of session). Conditional on the school being in session during our visit, students in PSL schools are 23 percentage points (p-value < 0.001) more likely to have a textbook and 8.2 percentage points (p-value .049) more likely to have writing materials (both a pen and a copybook). However, we cannot rule out that there is no overall effect as zero is between the Lee (2009) bounds.

[Table 6 about here.]

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 (i.e., the school is open, students and teachers are on campus, and classes are taking place) during a regular school day (p-value .057), and have a longer school day that translates into 3.2 more hours per week of instructional time (p-value < 0.001). Although principals in PSL schools have scores in the "intuitive" and "time management profile" scale that are almost identical to their counterparts in traditional public schools, they spend more of their time on management-related activities (e.g., supporting other teachers, monitoring student progress, meeting with parents) than actually teaching, suggesting a change in the role of the principal in these schools — perhaps as a result of additional teachers, principals in PSL schools did not have to double as teachers. Additionally, management practices (as measured by a "good practices" PCA index normalized to a mean of zero and standard deviation of one in the control

group) are .4σ (p-value < 0.001) higher in PSL schools.[21] This effect size can be viewed as a boost for the average treated school from the 50th to the 66th percentile in management practices.

[Table 7 about here.]

3.3.3 Teacher behavior

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

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 7.5 percentage points (p-value < 0.001) less likely to have missed school the previous week. Students in PSL schools also report that teachers are 6.6 percentage points (p-value .0099) less likely to hit them.

Classroom observations also show changes in teacher behavior and pedagogical practices. Teachers in PSL schools are 15 percentage points (p-value .0023) more likely to engage in either active instruction (e.g., teacher engaging students through lecture or discussion) or passive instruction (e.g., students working in their seat while the teacher monitors progress) and 25 percentage points (p-value < 0.001) less likely to be off-task.[23] Although these are considerable improvements, the treatment group is still far off the Stallings

---

[21]The index includes whether the school has an enrollment log and what information is in it, whether the school has an official time table and whether it is posted, whether the school has a parent-teacher association (PTA) and whether the principal has the PTA head's number at hand, and whether the school keeps a record of expenditures and a written budget. Table A.11 has details on every component of the good practices index.

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

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

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).

[Table 8 about here.]

These estimates combine the effects on individual teacher behavior with changes to teacher composition. To estimate the treatment effect on teacher attendance over a fixed pool of teachers, we perform additional analyses in Appendix A.1 using administrative data (EMIS) to restrict our sample to teachers who worked at the school the year before the intervention began (2015/2016). We treat teachers who no longer worked at the school in the 2016/2017 school year as (non-random) attriters and estimate Lee (2009) bounds on the treatment effect. Table A.10 in Appendix A.1 shows an ITT treatment effect of 14 percentage points (p-value < 0.001) on teacher attendance. Importantly, zero is not part of the Lee (2009) bounds for this effect. This aligns with previous findings showing that management practices have significant effects on worker performance (N. Bloom, Liang, Roberts, & Ying, 2014; N. Bloom, Eifert, Mahajan, McKenzie, & Roberts, 2013; Bennedsen, Nielsen, Pérez-González, & Wolfenzon, 2007).

3.4   Other outcomes

Student data (Table 9, Panel C) and household data (Table 9, Panel A) show that the program also increases student and parental satisfaction. Students in PSL are more likely to think going to school is fun, and parents with children in PSL schools (enrolled in 2015/2016) are 7.5 percentage points (p-value .021) more likely to be satisfied with the education their children are receiving.

Providers are not allowed to charge fees and PSL should be free at all levels, including early-childhood education (ECE) for which fees are permitted in government schools. We interviewed both parents and principals regarding fees. In both treatment and control schools parents are more likely to report paying fees than schools are to report charging them. The amount parents claim to pay in school fees is much higher than the amount schools claim to charge (see Panel A and Panel B in Table 9). Since principals may be reluctant to disclose the full amount they charge parents, especially in primary school (which is nominally free), this discrepancy is normal. While the likelihood of charging fees decreases in PSL schools by 26 percentage points according to parents and by 19 percentage points according to principals, 48% of parents still report paying some fees in PSL schools.

Providers often provide textbooks and uniforms free of charge to students (see Section 2.1.3). Indeed, household expenditures on fees, textbooks, and uniforms drop (see Table A.12 for details). In total,

20

annual household expenditures on children's education decrease by 6.6 USD (p-value .11). A reduction in household expenditure in education reflects a crowding out response (i.e., parents decrease private investment in education as school investments increase). To explore whether crowding out goes beyond expenditure we ask parents about engagement in their child's education. However, we see no change in this margin (we summarize parental engagement using the first component from a principal component analysis across several measures of parental engagement; see Table A.13 for the effect on each component).

To complement the effect of the program on cognitive skills, we also look for changes in student attitudes and opinions (see Table 9, Panel C). Some of the control group rates are noteworthy: 50% of children use what they learn in class outside school, 69% think that boys are smarter than girls, and 79% think that some tribes in Liberia are bad. Turning to treatment effects, children in PSL schools are more likely to think school is useful, more likely to think elections are the best way to choose a president, and less likely to think some tribes in Liberia are bad. The effect on tribe perceptions is particularly important in light of the recent conflict in Liberia and the ethnic tensions that sparked it. Our results also align with previous findings from Andrabi, Bau, Das, and Khwaja (2010), who show that children in private schools in Pakistan are more "pro-democratic" and exhibit lower gender biases (although we do not find any evidence of lower gender biases in this setting). Note, however, that our treatment effects are small in magnitude. It is also impossible to tease out the effect of who is providing education (private providers vs regular public schools) from the effect of better education, and the effect of younger and better teachers. Hence, our results show the net change in students' opinions, and cannot be attributed to providers per se but rather to the program as a whole.

[Table 9 about here.]

4 Provider comparisons

4.1 Raw 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 providers. This is unsurprising since providers stated different preferences for locations and some volunteered to manage schools in more remote and marginalized areas. Therefore, the raw treatment effects for each individual provider are internally valid, but not comparable with each other without further assumptions (see Section

We show how the average school for each provider differs from the average public school in Liberia in Table A.14. We reject the null that providers' schools have similar characteristics on at least three margins: number of students, pupil/teacher ratio, and the number of permanent classrooms. Bridge International Academies is managing schools that were considerably bigger (in 2015/2016) than the average public school in Liberia (by over 150 students), and these schools are larger than those of other providers by over 100 students. Most providers have schools with better infrastructure than the average public school in the country, except for Omega and Stella Maris. Finally, while all providers have schools that are closer to a paved road than other public schools, Bridge's and BRAC's schools are about 2 km closer than other providers' schools. Overall, these results confirm that some providers were more willing to work in average Liberian schools, while others preferred schools with easier access and better infrastructure.

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

[Table 10 about here.]

The treatment effects on composite test scores are positive and significantly different from zero for three providers: Rising Academies, Bridge International Academies, and Street Child (Table 10 - Panel A). They are positive but statistically insignificant for Youth Movement for Collective Action, More Than Me, and BRAC. Non-compliance likely explains the negative (but statistically insignificant) effect for Stella Maris and Omega Schools. Stella Maris never took control of its assigned schools. Omega had not taken control of all its schools by the end of the school year. Our teacher interviews reflect these providers' absence. In 3 out of four Stella Maris schools, all the teachers reported that no one from Stella had been at the school in the previous week. In 6 out of 19 Omega schools all the teachers reported that no one from Omega had been at the school in the previous week. While we committed in the pre-analysis plan to compare for-profit to non-profit providers, this comparison yields no clear patterns.

To measure teacher selection, we study the number of teachers dismissed and the number of new teachers recruited (Table 10 - Panel B). As noted above, PSL led to the assignment of 2.6 additional teachers per school and 1.2 additional teachers exiting per school. However, large-scale dismissal of teachers was

unique to one provider (Bridge International Academies), while successful lobbying for additional teachers was common across several providers. Although weeding out bad teachers is important, a reshuffling of teachers is unlikely to raise average performance in the system as a whole. We are unable to verify whether the teachers dismissed from PSL schools were reassigned to other public schools.

While enrollment increased across all providers, the smallest treatment effect on this margin is for Bridge, which is consistent with that provider being the only one enforcing class size caps (see Panel C in Table 10 and Figure A.5 in Appendix A.1 for more details). As shown in Section 3.2, in classes where class-size caps were binding (10% of all classes holding 30% of students at baseline), enrollment fell by 12 students per grade.

4.2 Comparable treatment estimates

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

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

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

and estimating the treatment effect of each provider on the average school in our sample.[24]

We show the full set of results across providers after adjusting for baseline differences and "shrinking" the estimates using the Bayesian hierarchical model in Table A.15 in Appendix A.1. While the comparable effects are useful for comparisons, the raw experimental estimates remain useful for non-comparative statements (e.g., whether a provider had an effect or not). Figure A.4 in Appendix A.1 shows the effects on learning after adjusting for differences in school characteristics (before the Bayesian hierarchical model) and the effects after applying a Bayesian hierarchical model (but without adjusting for school differences). Qualitatively, the results do not change. The learning gains remain positive for the same providers, and even after "shrinking" Bridge remains the only provider with a high (and statistically significant) percentage of teacher dismissal and the only one with a negative (and statistically significant) effect on enrollment in constrained grades.

4.2.1 Excluding some providers

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

For example, setting aside any political economy considerations, the government could drop the two providers that did not make much effort to control their schools (Omega and Stella Maris). It could also drop any provider who is potentially generating negative externalities (e.g., Bridge) or who may fail on dimensions different from test scores such as protecting students from physical and sexual abuse (e.g., More than Me). We estimate these potential outcomes by taking an inverse-variance weighted average across providers. We do this using both the raw estimates and comparable treatment estimates for completeness; however, given that the comparable treatment estimates are meant to inform about the treatment estimates

---

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

from the operators in any school in the experiment we focus on those.[25] Dropping the worst performing providers (Omega and Stella Maris) increases the overall treatment effect by 40-50%, while taking off the providers that may generate negative externalities (Bridge) reduces the treatment effect by 20-35% (see Table A.18 in the Appendix for details). Dropping both the worst performers and Bridge leaves the overall treatment effect unchanged. While the political economy of provider selection is non-trivial, we see this as prima facie evidence that the program has the potential to improve outcomes further by selecting providers dynamically.

5 Cost analysis

To attempt an answer to this question we make two comparisons: a comparative cost effectiveness calculation comparing PSL to business-as-usual expansion of Liberia's public school system, and a cost-benefit calculation based on the net present value of the Mincerian earnings returns to the education provided by PSL. Both calculations require strong assumptions (Dhaliwal, Duflo, Glennerster, & Tulloch, 2013), and we discuss a range of plausible alternatives.

Our data on operator costs are imperfect (see Section 2.1.4), and projecting the long-term unit cost of the program is essentially guesswork. Therefore we assume the cost of the program to be USD 50 per pupil, which was the government's budget target for PSL and the transfer made to operators. While some operators spent more than this, others reported spending near this amount. When the cost of additional teachers is included the cost rises to approximately USD 70 per student, and when the actual cost reported by providers for the first year is included the average increases to USD 238 (see Section 2.1.4 for details).

Computing the benefits is more straightforward. The ToT effect is $.22\sigma$, implying test scores increased 0.44σ per USD 100 spent (assuming a linear-dose relationship).[26] In order to juxtapose these results with the learning from business-as-usual schooling, and to map them onto existing estimates of the Mincerian returns to schooling, we translate them into "equivalent years of schooling" (EYOS) following Evans and Yuan (2017). This calculation assumes the value of schooling is fully captured by test score gains. In the control group, an extra year of schooling is associated with average test score gains of $.3\sigma$ across subjects.

---

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

[26]We do not present a cost-effectiveness comparison of the effect of the program on access to schooling since the overall treatment effect on enrollment is not statistically different from zero. Given our design, we are also unable to take into account any test score changes associated with drawing new students into school.

Thus, the ToT treatment effect is roughly 0.74 EYOS.

For the comparative cost-effectiveness calculation, we compare the program to the cost of one additional year of schooling from expanding the school system. We make two further strong assumptions: 1) the Liberian government can increase learning (and thus EYOS) by expanding the school system (e.g., hiring more teachers, building more classrooms, buying more textbooks); 2) there is a linear-dose relationship between the average per-pupil-year cost to the Liberian government of educating children (USD 50) and learning (EYOS). At face value, this calculation suggests that PSL is not a cost-effective program to increase schooling. The program provides an additional 0.74 EYOS at a cost of (at least) USD 50 per student, compared to a full year at equivalent or lower cost under business as usual.

Turning to cost-benefit analysis, we estimate whether the lifetime benefits accruing to students outweigh the costs of the program by estimating the net present value (NPV) of the investment. We make four key assumptions: 1) every extra year of schooling increases income by 8.6% (*Liberia Constraint Analysis*, 2013); 2) the discount rate is 10%[27]; 3) Liberia's GDP per capita is USD 694.3 in 2017 (World Bank, 2017); and 4) students work for 40 years (starting in ten years time). We find that the NPV of investing USD 50 in the PSL program is positive, at ∼USD 110; if we take into account the cost of additional teachers, the NPV is ∼USD 90. Thus, while the program may not be cost effective compared to expanding the school system, it is worth the investment.

How sensitive are these estimates to our assumptions? NPV becomes negative when the investment is above ∼USD 160, which is the case for some providers in the short run. A lower discount rate of, say, 4%[28] would raise this threshold cost to achieve positive NPV to USD 570. Alternatively, assigning an opportunity cost of children's time of USD 18 per annum or above would render PSL more cost effective than business as usual schooling. For comparison, the minimum wage in Liberia is USD 5.5 per day for skilled workers. Finally, incorporating the reduction in fees paid by parents would lower the cost of the program by USD 3.

6 Conclusions

Public-private partnerships in education are controversial and receive a great deal of attention from policy makers. Yet, there is little evidence for or against them in developing countries (Aslam et al., 2017). A

---

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

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

typical argument in favor is that privately provided but publicly funded education is a means to inject cost-efficiency into education without compromising equity. A typical argument against is that outsourcing will lead to student selection and other negative, unintended consequences.

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

We find no evidence that providers engage in student selection — the probability of remaining in a treatment school is unrelated to age, gender, household wealth, or disability. However, costs were high, performance varied across providers, and the largest provider pushed excess pupils and under-performing teachers into other government schools or completely out of the system. In addition, while outside the scope of our experimental analysis, the program has been plagued by accusations some operators failed to prevent, or actively concealed sexual abuse in schools they managed.

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

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

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

from across the distribution of outsourced schools in this setting.

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

149–171). Springer International Publishing.

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.

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

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

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.

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.

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

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

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

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

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (in press). Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. *The Quarterly Journal of Economics*.

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Figures

Figure 1: What did providers do?

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

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

Figure 2: Budget and costs as reported by providers

(a) Ex ante budget per pupil

(b) Ex post cost per pupil
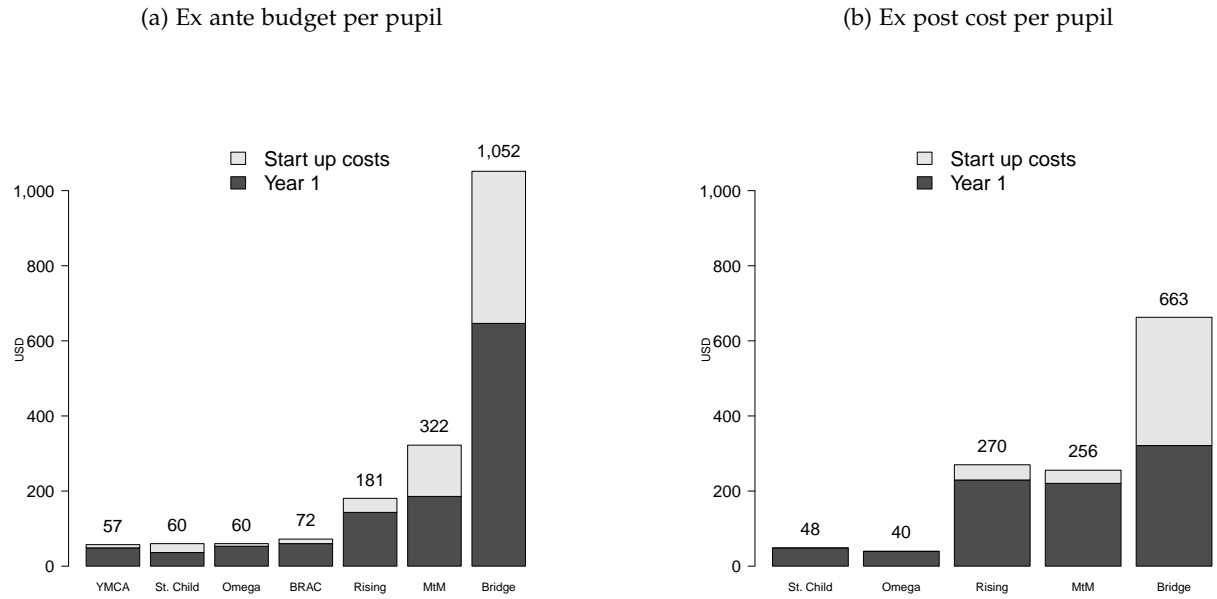
Note: Numbers in 2a are based on ex-ante budgets submitted to the program secretariat in a uniform template (inclusive of both fixed and variable costs). Stella Maris did not provide budget data. Numbers in 2b are based on self-reports on ex-post expenditures (inclusive of both fixed and variable costs) submitted to the evaluation team by five providers in various formats. Numbers do not include the cost of teaching staff borne by the Ministry of Education.

Figure 3: Public primary schools in Liberia

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

(b) Geographical distribution of treatment and control schools.
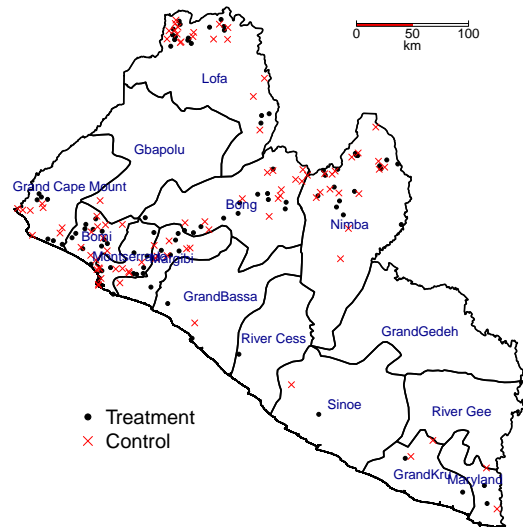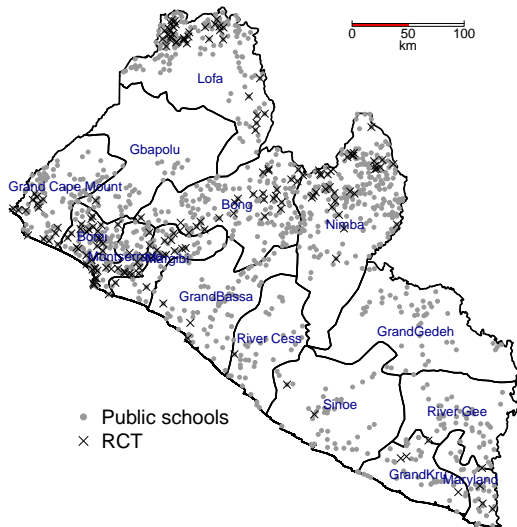
Tables

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

[a] Neither Bridge International Academies nor Stella Maris received the extra USD 50 per pupil.

[b] Bridge International Academies was authorized to cap class sizes at 55 (but in practice capped them at 45 in most cases as this was allowed by the MOU), while other providers were authorized to cap class sizes at 65.

[c] Bridge International Academies had first pick, before other providers, of the new teacher-training graduates.

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

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

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

Table 3: ITT treatment effects on learning

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

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

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

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Panel A: School level data (N = 175)				
Enrollment 2015/2016	298.45	264.11	34.34	34.18
	(169.74)	(109.91)	(21.00)	(20.28)
Enrollment 2016/2017	309.71	252.75	56.96	56.89
	(118.96)	(123.41)	(18.07)	(16.29)
15/16 to 16/17 enrollment change	11.55	-6.06	17.61	24.60
	(141.30)	(82.25)	(17.19)	(14.35)
Attendance % (spot check)	48.02	32.84	15.18	15.56
	(24.52)	(26.54)	(3.81)	(3.13)
% of students with disabilities	0.59	0.39	0.20	0.21
	(1.16)	(0.67)	(0.14)	(0.15)
Panel B: Student level data (N = 3,639)				
% enrolled in the same school	80.50	83.16	-2.66	0.71
	(39.63)	(37.43)	(3.66)	(2.06)
% enrolled in school	94.13	93.99	0.14	1.23
	(23.52)	(23.77)	(1.33)	(0.87)
Days missed, previous week	0.85	0.85	-0.00	-0.06
	(1.41)	(1.40)	(0.10)	(0.07)

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

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	5.30	4.04	1.64	0.15
	(1.11)	(1.39)	(0.73)	(0.034)
Constrained=1 × Treatment	-11.7	-12.8	0.070	0.35
	(6.47)	(7.74)	(4.11)	(0.11)
No. of obs.	1,635	3,625	3,485	3,490
Mean control (Unconstrained)	-0.75	82.09	93.38	0.13
Mean control (Constrained)	-7.73	84.38	94.81	-0.08
$\alpha_0 =$ Constrained - Unconstrained	-17.05	-16.79	-1.57	0.20
p-value ($H_0 : \alpha_0 = 0$)	0.01	0.03	0.71	0.07

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

Table 6: ITT treatment effects on inputs and resources

	(1) Treatment	(2) Control	(3) Difference	(4) Difference (F.E)
Panel A: School-level outcomes (N = 185)				
Number of teachers	9.62	7.02	2.60	2.61
	(2.82)	(3.12)	(0.44)	(0.37)
Pupil-teacher ratio (PTR)	32.20	39.95	-7.74	-7.82
	(12.29)	(18.27)	(2.31)	(2.12)
New teachers	4.81	1.77	3.03	3.01
	(2.56)	(2.03)	(0.34)	(0.35)
Teachers dismissed	3.35	2.17	1.18	1.16
	(3.82)	(2.64)	(0.48)	(0.47)
Panel B: Teacher-level outcomes (N = 1,167)				
Age in years	39.09	46.37	-7.28	-7.10
	(11.77)	(11.67)	(1.02)	(0.68)
Experience in years	10.59	15.79	-5.20	-5.26
	(9.20)	(10.77)	(0.76)	(0.51)
% has worked at a private school	47.12	37.50	9.62	10.20
	(49.95)	(48.46)	(3.76)	(2.42)
Test score in standard deviations	0.13	-0.01	0.14	0.14
	(1.02)	(0.99)	(0.07)	(0.06)
% certified (or tertiary education)	60.11	58.05	2.06	4.20
	(48.99)	(49.39)	(4.87)	(2.99)
Salary (USD/month)–Conditional on salary> 0	121.36	104.54	16.82	13.90
	(44.42)	(60.15)	(6.56)	(4.53)
Panel C: Classroom observation (N = 143)				
Number of seats	20.64	20.58	0.06	0.58
	(13.33)	(13.57)	(2.21)	(1.90)
% with students sitting on the floor	2.41	4.23	-1.82	-1.51
	(15.43)	(20.26)	(2.94)	(2.61)
% with chalk	96.39	78.87	17.51	16.58
	(18.78)	(41.11)	(5.29)	(5.50)
% of students with textbooks	37.08	17.60	19.48	22.60
	(43.22)	(35.25)	(6.33)	(6.32)
% of students with pens/pencils	88.55	79.67	8.88	8.16
	(19.84)	(30.13)	(4.19)	(4.10)

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Panel A has school level outcomes. Panel B presents teacher-level outcomes including their score in tests conducted by our survey teams. Panel C presents data on inputs measured during classroom observations. Since we could not conduct classroom observations in schools that were out of session during our visit, Table A.10 in Appendix A.1 presents Lee (2009) bounds on these treatment effects (control schools are more likely to be out of session). Standard errors are clustered at the school level.

Table 7: ITT treatment effects on school management

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

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Intuitive score is measured using Agor (1989)'s instrument and time management profile using Schermerhorn et al. (2011)'s instrument. The index of good practices is the first component of a principal component analysis of the variables in Table A.11. 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.

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

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Panel A presents data from spot checks conducted by our survey teams in the middle of a school day. Panel B presents data from our panel of students where we asked them about their teachers' behavior. Panel C presents data from classroom observations. If the school was not in session during a regular school day we mark all teachers not on campus as absent and teachers and students as off-task in the classroom observation. Table A.10 has the results without imputing values for schools not in session. Standard errors are clustered at the school level.

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

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

This table presents the mean and standard error of the mean (in parentheses) for the control (Column 1) and treatment (Column 2) groups, as well as the difference between treatment and control (Column 3), and the difference taking into account the randomization design (i.e., including "pair" fixed effects) in Column 4. Panel A presents data from household surveys. The index for parent engagement is the first component from a principal component analysis across several measures of parental engagement; see Table A.13 for details. Expenditure refers to the annual household expenditures on children's education. Panel B presents data from school principals on what fees schools charge. Panel C presents data on whether students agree or disagree with several statements. Standard errors are clustered at the school level.

Table 10: Raw (fully experimental) treatment effects by provider

	(1) BRAC	(2) Bridge	(3) MtM	(4) Omega	(5) Rising	(6) St. Child	(7) Stella M	(8) YMCA
Panel A: Student test scores (ITT)								
English (standard deviations)	0.19	0.28	0.19	-0.07	0.36	0.23	-0.23	0.58
	(0.10)	(0.09)	(0.22)	(0.11)	(0.24)	(0.13)	(0.23)	(0.26)
Math (standard deviations)	0.09	0.39	0.18	-0.06	0.42	0.28	-0.17	0.27
	(0.09)	(0.09)	(0.22)	(0.11)	(0.23)	(0.13)	(0.22)	(0.26)
Composite (standard deviations)	0.14	0.36	0.18	-0.08	0.42	0.27	-0.19	0.38
	(0.09)	(0.09)	(0.22)	(0.11)	(0.23)	(0.13)	(0.22)	(0.26)
Panel B: Changes to the pool of teachers								
% teachers dismissed	-6.83	49.98	15.86	-9.12	-5.73	-2.63	-3.51	20.96
	(6.51)	(6.36)	(11.90)	(6.89)	(12.88)	(8.59)	(14.52)	(14.52)
% new teachers	39.62	62.95	69.54	24.34	24.38	40.94	-21.93	62.20
	(12.29)	(12.02)	(22.46)	(13.01)	(24.31)	(16.21)	(27.41)	(27.41)
Age in years (teachers)	-5.03	-10.92	-11.20	-5.46	-10.75	-5.79	-4.53	3.25
	(1.93)	(2.01)	(3.52)	(2.03)	(3.82)	(2.54)	(4.30)	(4.30)
Test score in standard deviations (teachers)	0.03	0.36	0.48	0.18	0.18	0.32	0.16	-0.59
	(0.17)	(0.17)	(0.31)	(0.17)	(0.33)	(0.22)	(0.38)	(0.38)
Panel C: Enrollment and access								
Δ enrollment	38.02	-13.26	-25.98	51.27	19.31	44.86	-15.92	45.38
	(34.33)	(33.60)	(62.76)	(35.26)	(67.84)	(45.21)	(76.59)	(76.53)
Δ enrollment (constrained grades)	0.00	-23.85	0.00	0.28	0.00	32.15	0.00	0.00
	(0.00)	(11.19)	(0.00)	(37.16)	(0.00)	(61.95)	(0.00)	(0.00)
Student attendance (%)	20.09	5.25	37.81	18.01	28.76	19.56	9.71	13.53
	(9.02)	(9.05)	(16.50)	(9.53)	(17.82)	(11.88)	(23.32)	(20.11)
% students still attending any school	1.27	5.19	-3.12	4.71	2.82	3.64	5.98	4.48
	(4.45)	(4.22)	(10.25)	(4.99)	(11.03)	(6.11)	(10.57)	(12.21)
% students still attending same school	0.80	4.42	0.65	1.56	3.81	-0.82	1.03	-0.81
	(2.20)	(2.09)	(5.07)	(2.46)	(5.45)	(3.02)	(5.23)	(6.04)
Panel D: Satisfaction								
% satisfied with school (parents)	11.72	13.22	0.75	0.21	4.95	-4.96	29.49	18.02
	(7.30)	(7.14)	(13.34)	(7.53)	(14.44)	(9.62)	(16.28)	(16.27)
% students who think school is fun	5.83	2.11	0.50	4.86	9.44	2.84	-17.50	20.92
	(4.89)	(4.63)	(11.25)	(5.47)	(12.11)	(6.71)	(11.60)	(13.40)
Observations	40	45	8	12	38	10	24	8

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

A Online Appendix

A.1 Additional tables and figures

[Figure A.1 about here.]

[Figure A.2 about here.]

[Figure A.3 about here.]

[Figure A.4 about here.]

[Figure A.5 about here.]

[Table A.1 about here.]

[Table A.2 about here.]

[Table A.3 about here.]

[Table A.4 about here.]

[Table A.5 about here.]

[Table A.6 about here.]

[Table A.7 about here.]

[Table A.8 about here.]

[Table A.9 about here.]

[Table A.10 about here.]

[Table A.11 about here.]

[Table A.12 about here.]

[Table A.13 about here.]

[Table A.14 about here.]

[Table A.15 about here.]

[Table A.16 about here.]

A.2 Tracking and attrition

A potential issue with our sampling strategy is differential attrition at each round of data collection. In the first round, enumerators were instructed to sample 20 students from the 2015/2016 enrollment logs, track them, and test them. However, if a student had moved to another village, had died, or was impossible to track, the enumerators were instructed to sample another student. Thus, even at the first round an endogenous sampling problem arises if treatment makes students easier or harder to track in combination with enumerator shrinkage. To mitigate this issue, enumerators participated in additional training on tracking and its importance and were provided with a generous amount of tracking time. Students were tracked to their homes and tested there when not available at school. As Table A.19 shows, we have no reason to believe that this issue arose. The effort required to track students was different between treatment and control (it is easier to track students at the school), yet the total number of students sampled, to obtain a sample of 20 students, is balanced between treatment and control (see Table A.19).

[Table A.17 about here.]

A.3 Test design

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

from Uwezo), number identification (adapted from EGMA), addition (adapted from Uwezo and EGMA), subtraction (adapted from Uwezo and EGMA), multiplication (adapted from Uwezo and EGMA), division (adapted from Uwezo and EGMA), shape identification, fractions, and word problems in mathematics.

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

A.4   Provider's details

[Table A.18 about here.]

Figures

Figure A.1: Enrollment by age
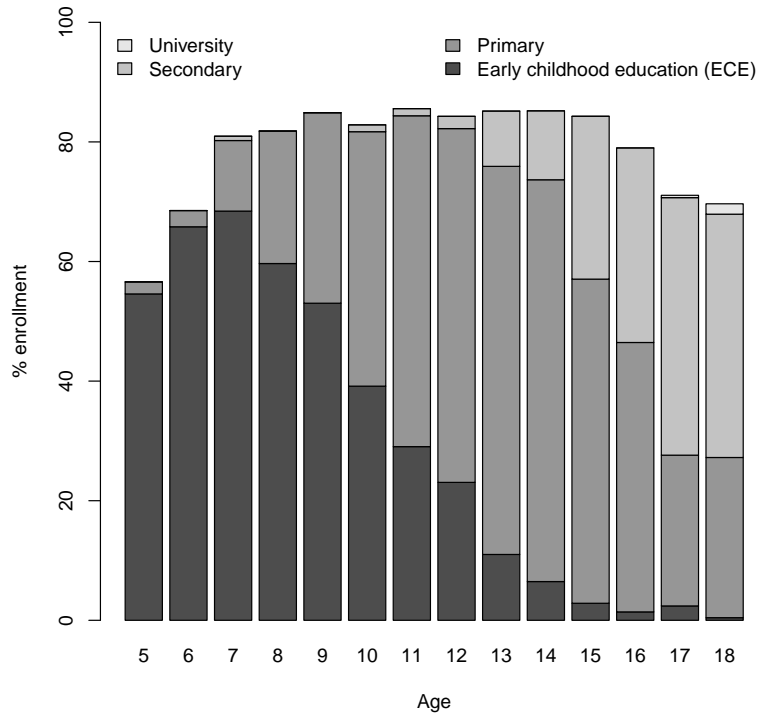
Note: This figure shows the percentage of students enrolled in early childhood education, primary schools, secondary schools, or universities by age. Authors' calculations based on 2014 Household Income and Expenditures Survey.

Figure A.2: Timeline

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

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

Figure A.3: Treatment effects by date tested during the first round of data collection

(a) Math

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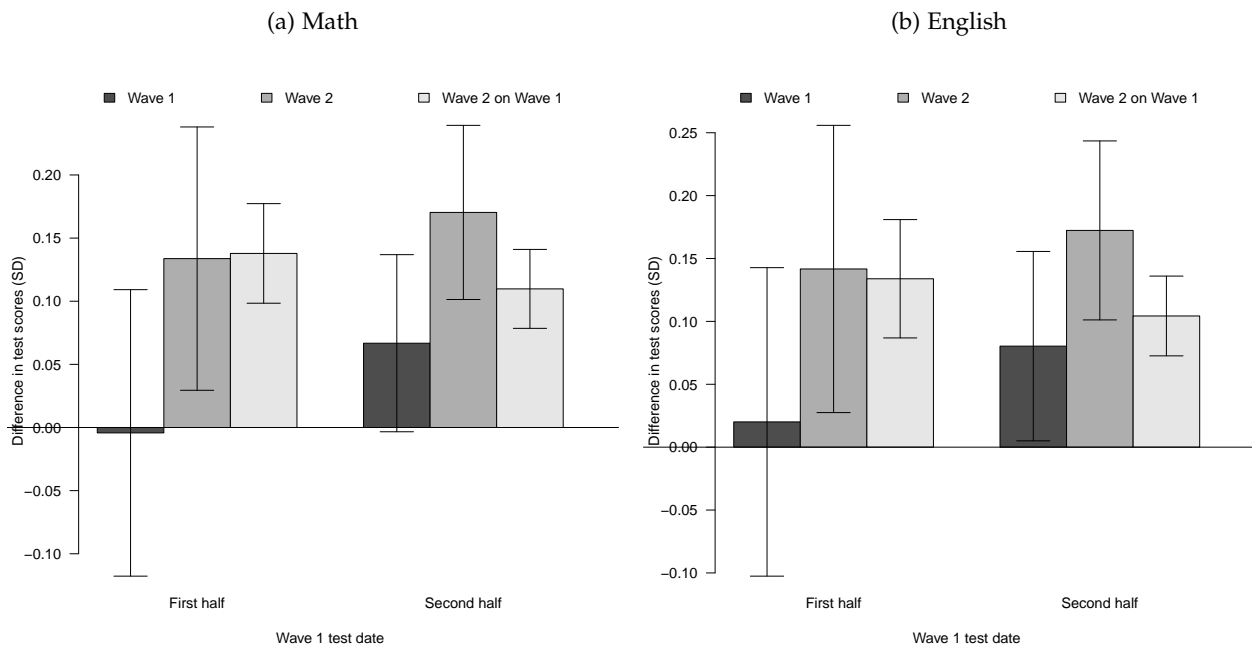
Note: These figures show the intention-to-treat treatment effects on test scores by whether during the first wave of data collection students were tested in the first or the second half of field work. Treatment effects are estimated using test scores from the first and the second wave of data collection.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

Figure A.4: Treatment effects by provider

(a) Intention-to-treat (ITT) effect   (b) Treatment-on-the-treated effect (ToT)
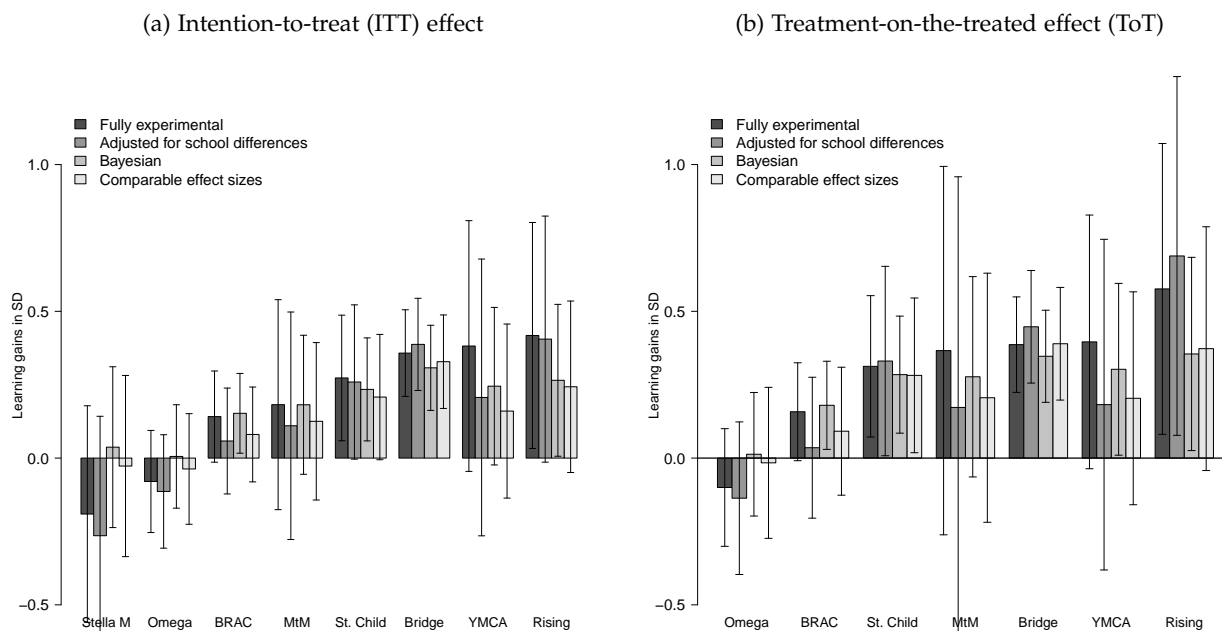
Note: These figures show the raw, fully experimental treatment effects, the effects after adjusting for differences in school characteristics (before the Bayesian hierarchical model), the effects after applying a Bayesian hierarchical model (but without adjusting for school differences), and the comparable treatment effects after adjusting for differences in school characteristics and applying a Bayesian hierarchical model. Figure A.4a shows the intention-to-treat (ITT) effect, while Figure A.4b shows the treatment-on-the-treated (ToT) effect. The ToT effects are larger than the ITT effects due to providers replacing schools that did not meet the eligibility criteria, providers refusing schools, or students leaving PSL schools. Stella Maris had full non-compliance at the school level and therefore there is no ToT effect for this provider.

Figure A.5: Class sizes and class caps

(a) BRAC

(b) Bridge

(c) YMCA

(d) More than Me

(e) Omega Schools

(f) Rising Academies

(g) St. Child

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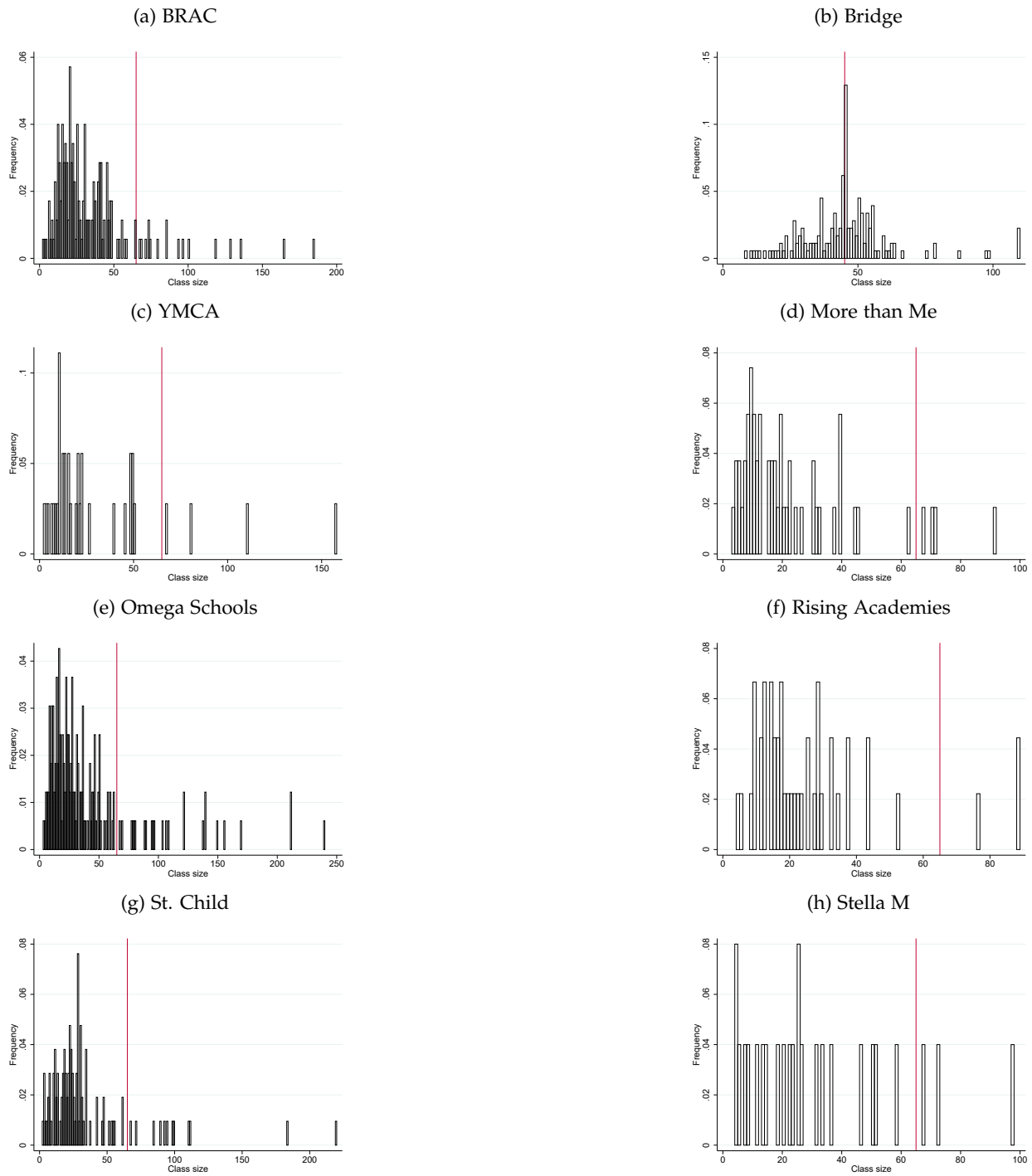
Note: These figures show the distribution of class sizes in treatment schools during the 2016/2017 academic year, as well as the class cap for each provider. The cap for all providers is 65 students, except for Bridge that has a cap of 45.

Tables

Table A.1: External validity: Differences in characteristics of schools in the RCT (treatment and control) and other public schools (based on EMIS data)

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

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

Table A.2: Number of schools by provider

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

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

Table A.3: Balance table: Differences in characteristics of treatment and control schools, pre-treatment year (2015/2016, EMIS data)

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

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

Table A.4: ITT and ToT effect after one year

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

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

Table A.5: Control variables

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

This table shows the control variables included in specification 2.

Table A.6: Treatment effects across various measures of difference in student ability

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

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

Table A.7: Heterogeneity by student characteristics

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

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

Table A.8: Student selection

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

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

Table A.9: ITT treatment effects, by whether class size caps are binding without including adjcent grades

	(1) Δ enrollment	(2) % same school	(3) % in school	(4) Test scores
Constrained=0 × Treatment	2.96	3.83	1.53	0.10
	(1.08)	(1.43)	(0.67)	(0.039)
Constrained=1 × Treatment	17.3	-12.5	-13.4	0.36
	(7.53)	(5.84)	(3.53)	(0.14)
No. of obs.	1,256	2,773	2,636	2,641
Mean control (Unconstrained)	-0.43	82.57	94.00	0.08
Mean control (Constrained)	-9.03	80.95	100.00	-0.33
$\alpha_0 =$ Constrained - Unconstrained	14.30	-16.34	-14.95	0.26
p-value ($H_0 : \alpha_0 = 0$)	0.07	0.01	0.00	0.07

This table mirrors Table 5, but adjacent grades are not included in the calculation of the constrained indicator. Column 1 uses school-grade level data. Columns 2 - 4 use student level data. The independent variable in Column 4 is the composite test score. The sample is the original treatment and control allocation. There were 216 constrained classes at baseline (holding 35% of students), and 1,448 unconstrained classes at baseline (holding 65% of students). Standard errors are clustered at the school level.

Table A.10: Intensive margin effect on teacher attendance and classroom observation with Lee bounds

	(1) Treatment	(2) Control	(3) Difference (F.E)	(4) Difference	(5) 90% CI (bounds)
Panel A: Spot check (N = 929)					
% on schools campus	68.15	52.40	15.75	14.17	2.51
	(46.64)	(50.00)	(4.45)	(3.75)	28.11
% in classroom	50.96	41.05	9.91	9.96	-1.34
	(50.04)	(49.25)	(4.78)	(3.86)	24.44
B: Classroom observation (N = 143)					
Active instruction (% class time)	38.12	30.13	7.98	7.62	-4.75
	(28.93)	(32.11)	(4.86)	(4.75)	19.92
Passive instruction (% class time)	16.24	12.80	3.44	4.72	-4.93
	(17.18)	(19.83)	(2.95)	(3.23)	9.62
Classroom management (% class time)	20.82	10.67	10.16	10.33	0.77
	(21.06)	(14.83)	(2.85)	(3.32)	16.99
Teacher off-task (% class time)	24.82	46.40	-21.58	-22.66	-40.24
	(32.65)	(41.09)	(5.92)	(6.26)	-10.32
Student off-task (% class time)	55.06	57.60	-2.54	-5.19	-16.05
	(31.23)	(34.87)	(5.26)	(4.88)	12.63
Panel C: Inputs (N = 143)					
Number of seats	20.64	20.58	0.06	0.58	-7.22
	(13.33)	(13.57)	(2.21)	(1.90)	5.36
% with students sitting on the floor	2.41	4.23	-1.82	-1.51	-7.48
	(15.43)	(20.26)	(2.94)	(2.61)	2.76
% with chalk	96.39	78.87	17.51	16.58	9.47
	(18.78)	(41.11)	(5.29)	(5.50)	27.85
% of students with textbooks	37.08	17.60	19.48	22.60	-1.21
	(43.22)	(35.25)	(6.33)	(6.32)	34.87
% of students with pens/pencils	88.55	79.67	8.88	8.16	1.36
	(19.84)	(30.13)	(4.19)	(4.10)	20.98

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

Table A.11: Treatment effect on schools' good practices

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

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

Table A.12: Treatment effect on household expenditure

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

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

Table A.13: Treatment effect on household engagement

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

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

Table A.14: Baseline differences between treatment schools and average public schools, by provider

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

This table presents the difference between public schools and the schools operated by each provider. The information for all schools is taken from the 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 providers had no schools with classes above the class caps, there is no data to estimate treatment effects over constrained classes. Standard errors are clustered at the school level.

Table A.15: Comparable ITT treatment effects by provider

	(1) BRAC	(2) Bridge	(3) MtM	(4) Omega	(5) Rising	(6) St. Child	(7) Stella M	(8) YMCA	(9) p-value
Panel A: Student test scores									
English (standard deviations)	0.14	0.26	0.17	0.02	0.24	0.21	0.03	0.24	0.10
	(0.09)	(0.09)	(0.15)	(0.11)	(0.16)	(0.12)	(0.18)	(0.17)	
Math (standard deviations)	0.04	0.35	0.10	-0.05	0.22	0.19	-0.05	0.10	0.010
	(0.10)	(0.10)	(0.17)	(0.11)	(0.18)	(0.13)	(0.19)	(0.18)	
Composite (standard deviations)	0.08	0.33	0.13	-0.04	0.24	0.21	-0.03	0.16	0.020
	(0.10)	(0.10)	(0.16)	(0.11)	(0.18)	(0.13)	(0.19)	(0.18)	
Panel B: Changes to the pool of teachers									
% teachers dismissed	-8.61	49.53	13.93	-6.24	0.51	-0.81	-1.65	11.96	<0.001
	(6.43)	(7.16)	(11.14)	(6.73)	(11.95)	(8.98)	(12.93)	(13.00)	
% new teachers	38.10	70.88	47.26	22.55	20.62	35.95	-9.86	35.77	0.0050
	(11.22)	(13.14)	(18.78)	(11.90)	(20.15)	(15.19)	(26.26)	(21.07)	
Age in years (teachers)	-5.50	-9.12	-7.80	-5.75	-8.07	-6.53	-5.99	-3.50	0.16
	(1.71)	(2.18)	(2.57)	(1.73)	(2.74)	(2.08)	(2.72)	(3.51)	
Test score in standard deviations (teachers)	0.12	0.24	0.23	0.17	0.17	0.23	0.17	0.05	0.46
	(0.13)	(0.14)	(0.18)	(0.13)	(0.18)	(0.16)	(0.18)	(0.23)	
Panel C: Enrollment and access									
Δ enrollment	31.91	7.80	12.63	28.93	16.24	25.38	15.79	27.52	0.49
	(25.45)	(26.78)	(32.57)	(25.01)	(32.75)	(28.81)	(33.86)	(34.02)	
Δ enrollment (constrained grades)	41.57	-29.66	41.64	-3.45	41.30	22.33	–	–	0.48
	(44.13)	(14.58)	(44.24)	(36.91)	(44.03)	(46.95)	(–)	(–)	
Student attendance (%)	18.44	12.80	20.74	17.54	19.00	19.37	16.67	17.47	0.48
	(6.61)	(7.55)	(9.19)	(6.69)	(8.93)	(7.94)	(9.51)	(9.08)	
% students still attending any school	-1.98	1.27	-4.78	-2.03	-3.82	-1.97	-3.21	-3.18	0.36
	(3.38)	(3.68)	(5.90)	(3.64)	(5.62)	(4.26)	(5.35)	(5.60)	
% students still attending same school	0.54	2.36	0.36	0.66	0.74	0.25	0.30	0.15	0.44
	(1.76)	(1.91)	(2.58)	(1.87)	(2.58)	(2.22)	(2.62)	(2.77)	
Panel D: Satisfaction									
% satisfied with school (parents)	11.75	10.97	3.72	1.72	2.59	-0.37	9.96	8.51	0.24
	(6.32)	(6.39)	(8.45)	(6.32)	(9.00)	(8.36)	(9.36)	(9.14)	
% students who think school is fun	4.08	2.73	2.49	3.25	3.47	2.59	0.01	4.75	0.59
	(3.90)	(3.66)	(5.42)	(4.05)	(5.59)	(4.65)	(6.67)	(6.12)	
Observations	40	45	8	12	38	10	24	8	

This table presents the ITT treatment effect for each provider, after adjusting for differences in baseline school characteristics, based on a Bayesian hierarchical model. Column 9 shows the p-value for testing $H_0 : \beta_{BRAC} = \beta_{Bridge} = \beta_{YMCA} = \beta_{MtM} = \beta_{Omega} = \beta_{Rising} = \beta_{St.Child} = \beta_{StellaM}$. Table 10 has the raw experimental treatment effects by provider. Standard errors are shown in parentheses. Estimation is conducted on collapsed, school-level data.

Table A.16: Descriptive statistics by provider and treatment

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

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

Table A.17: Treatment effect on instruction time by subject

	Hours per week			% time per week		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.69	0.33	0.33	-0.40	-2.36	-2.36
	(0.11)	(0.13)	(0.13)	(0.25)	(1.26)	(1.26)
Math or English	2.08	1.71		10.1	8.07	
	(0.13)	(0.16)		(1.16)	(1.76)	
Treatment $\times$ Math or English		0.72			3.91	
		(0.26)			(2.32)	
Math			0.34			4.27
			(0.076)			(1.96)
English			3.08			11.9
			(0.30)			(2.22)
Treatment $\times$ Math			0.66			3.22
			(0.17)			(2.64)
Treatment $\times$ English			0.78			4.60
			(0.47)			(2.97)
No. of obs.	4,299	4,299	4,299	4,299	4,299	4,299

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

Table A.18: Simulated treatment effects without some providers

	(1) Raw	(2) Comparable
All operators	.18	.14
	(.032)	(.055)
Without Stella M and Omega	.27	.2
	(.035)	(.055)
Without Bridge	.12	.11
	(.037)	(.05)
Without Stella M, Omega, Bridge	.21	.14
	(.042)	(.062)

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

Table A.19: Tracking and sampling in the first wave of data collection

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

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

Table A.20: Provider's characteristics

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

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

Table A.20 – continued from previous page

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

Table A.20 – continued from previous page

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

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