Does BRAC Provide Highly Effective Schooling in Developing Countries?

John C. Ham *
New York University, Abu Dhabi
NYU Wagner School of Public Service

Saima Khan †
New York University, Abu Dhabi

March 20, 2021

Abstract

BRAC, a pioneer of non-formal primary education, has been operating in Bangladesh for the past 30 years often filling the gap that the formal sector failed to address. Since the 2000s BRAC has taken its schooling model to international locations such as Afghanistan, Philippines, Uganda etc. Despite its long tenure and wide coverage, there are few independent rigorous evaluations of BRAC schools. Thus, in this paper, we compare BRAC to Government, JAAGO (an innovative new school-type) and other NGO schools in two slums of Dhaka city. We examine school-type impact on student achievement in mathematics, using high quality data that we collected ourselves. We find that in terms of fluid intelligence (IQ) and parents' education, 'weaker' students sort into BRAC schools as compared to Government, JAAGO and other NGO schools. We control for this selection using a rich set of conditioning variables. We use propensity score matching in the presence of choice-based sampling to estimate the treatment effects. We find that controlling for fluid intelligence (IQ) has, by far, the biggest effect on reducing selection bias, and hence indicates the importance of collecting such data when using non-experimental methods. We find that once we control for selection, BRAC students are worse off than their counterparts in Government and JAAGO schools; we find no difference between comparable students at BRAC versus NGO schools. We disaggregate by gender to find that the BRAC versus Government effect is being driven by boys and the BRAC versus JAAGO effect is being driven by girls.

Keywords: Marginalized Communities, Gender Heterogeneity, Selection, Propensity Score Matching, Choice-Based Sampling.

JEL Codes: J01, I25, I26, O1.

---

*Address: Social Science Building (A5), NYU Abu Dhabi, Saadiyat Campus P.O. Box 129188, Abu Dhabi, United Arab Emirates; e-mail: jch18@nyu.edu.

†Presenter; Address: New York University, Development Research Institute, 14A Washington Mews, Room no. 304, New York, NY 10003; e-mail: sk7720@nyu.edu.

1 Introduction

A major challenge for developing countries is making quality schooling available to a large proportion of the population, particularly in marginalized communities. Often the government, as the sole provider of primary education, cannot successfully cover the entire school going population. Hence, non-government organizations (NGOs) frequently serve a significant fraction of children. In a country like Bangladesh, BRAC schools teach a considerable number of students. BRAC states that it targets out-of-school children in marginalized communities, not covered by the formal schooling system.

BRAC started in Bangladesh in 1985 with 22 schools. Now it has more than 40,000 schools both in Bangladesh and in international locations such as Afghanistan, Philippines, South Sudan and Uganda. The BRAC model has even been replicated by the Government of Bangladesh under the banner of Reaching Out of School Children (ROSC) schools.

Given BRAC's coverage, both within Bangladesh and other developing countries, it is surprising that there are very few evaluations of BRAC schools, all of which have been conducted by BRAC researchers. Further, these studies compare mean performance on a standardized test in BRAC schools and in other schools (e.g., government schools.) But a comparison of these means is only valid if there is no selection into BRAC schools relative to the other teaching modes, and the above authors do not provide any evidence on such selection. In our data on urban students, BRAC serves the most-disadvantaged children, while, e.g., government and JAAG0 schools teach much more advantaged students. It is possible that this sorting does not occur in the rural settings that the above researchers focus on, but our findings suggest that it is very worthwhile to investigate selection in rural schools.

Ours is the first arm's length study of BRAC schools; even if just for appearances sake, it would be good to have more independent studies of BRAC's impacts. In terms of the research by BRAC-affiliated researchers, there are two studies of note. Nath et al. (2007) compares the impact of BRAC's non-formal primary schools (BPS henceforth) to the impact of government schools (GOV henceforth). They find that BRAC students do slightly better than GOV schools across a number of subjects. They do not consider differences in family background between the BRAC and GOV students or report any information on them. Alternatively, Nath and Hossain (2017) compare BPS students to a national sample of students, and find that the BRAC students generally do better than those in the national sample. The BRAC students in that study are at a disadvantage in terms of family background characteristics, but the authors do not generally control for family background in obtaining their parameter estimates.

Here we use data from a previous study, Ham and Khan (2021). This study used choice-based sampling to evaluate the impacts on math achievement from JAAGO schools, NGO schools, and GOV schools in 2015-2016. We collected data on approximately 600 students in each school type, in two slums of Dhaka city, the capital of Bangladesh. Note that choice-based sampling was crucial for our evaluation, since there would have been very few JAAGO students in a random sample of students in the Dhaka slums. These sample sizes allowed for precise estimation of these impacts in our original study. We found first that JAAGO and GOV students came from much better family backgrounds, and had considerably higher IQ scores than the NGO students. When we do not control for these differences, we found that both JAAGO and GOV students each did significantly better than NGO students by a wide margin. When we controlled for family background and IQ, the difference between GOV students, JAAGO students, and NGO schools shrunk, but were still significant and substantial.

After completing the above study, we realized that we could identify BRAC students from the non-BRAC NGO students for the NGO sample in our data set, and found that we had 239 BRAC students and 411 non-BRAC NGO students in our original NGO sample. This raised the question of whether we would be able to do precise inference with only 239 BRAC students. However, that was not a problem in our results. We first found that BRAC students came from significantly poorer families than JAAGO students, GOV students, and other-NGO students. Second, we found that when we did not control for selection, BRAC students had (significantly) much poorer scores than JAAGO students, GOV students, and other NGO students. When we controlled for selection, the differences between BRAC students and both JAAGO and GOV students fell by approximately half, but these differences were still quite substantial. But after controlling for selection we could not reject the null hypothesis that BRAC schools and other-NGO schools produce equal impacts. We also found that conditioning on a student's IQ, as opposed to conditioning on just family background, was by far the most important factor in controlling for selection.

Note that our results differ dramatically from those in Nath et al. (2007) and Nath and Hossain (2017). However, when considering these differences, readers should keep in mind that (i) we used an urban sample whereas they use a rural sample and (ii) we measure the schooling impacts on only math performance whereas they measure the impacts on a variety of subjects and (iii) we use a performance test widely used in the literature while they used tests developed for Bangladesh which have not been widely used in the literature.

As noted above, we used choice-based sampling data set in this study to collect data on approximately 600 JAAGO, Government and NGO students. We use propensity score matching to control for selection. Standard PSM approaches do not provide consistent estimates of treatment effects for a choice-based sample. However, drawing on Heckman and Todd (2009) and Hahn, Ham and Ridder (2021) we were able to use consistent estimators for the treatment effects. We would note that instrumental variables generally is not a consistent estimator in the presence of choice based sampling. Hahn, Ham and Ridder (2021) derive a consistent IV estimator for choice-based samples, but it is not applicable in our context.

In using our matching estimator, we address the issue of common support, i.e., for every student, can we sensibly find valid comparisons for her in the other group? This issue tended to receive more attention in earlier matching studies than in recent matching papers, but given the wide range of students across the different school types, it is essential that we take it seriously.

In summary, we find that students are significantly worse off at BRAC schools than at GOV or JAAGO schools. Moreover, the magnitude of these differences is substantial and the confidence intervals for these differences are small, so our results will indeed be useful for evaluating the different relative impacts of these various schools. We find no significant difference between students at BRAC and other NGO schools, but this may reflect our relatively small samples for BRAC and other NGO students. When we disaggregate by gender, we see that the significant BRAC vs GOV differences are being driven by boys; we find a significant difference for boys but cannot reject the null hypothesis that there is no difference between girls at both school-types.

When we compare BRAC to JAAGO schools, the difference is now being driven by girls, who do significantly worse at BRAC schools than at JAAGO schools; we find no significant difference between boys at BRAC versus boys at JAAGO schools. Finally, we do not see any significant differences between BRAC vs other NGO schools for boys or for girls.

The rest of the paper is organized as follows. Sections 2 gives some background on BRAC schools, and then focuses on the differences in approach across the different school-types. Here, Section 3 reviews the literature while Section 4 gives an overview of the data. We present formal evidence of the different allocation of students across the four school-types in our sample in Section 5. Section 6 gives an overview of propensity score matching in the presence of choice-based sampling. We present our empirical results in Section 7, while Section 8 concludes.

2 How Different School-types Operate - A Comparison

2.1 An Introduction to BRAC Schools

Established in 1972, BRAC is now one of the largest international NGOs in the world. BRAC has a host of programs ranging from providing micro loans to providing education, health care, skills development, legal and human rights services.[1] Our focus is on BRAC's Education Program; BRAC runs schools in both rural and urban areas.

Started in 1985, BRAC's Education Program focuses on five areas: non-formal primary education, the pre-primary schools program, the adolescent development program, multipurpose community learning centers and the mainstream secondary schools support initiative (Hanemann, 2017). For the purpose of this paper we focus on the first two areas.

The BRAC model of non-formal schooling has a one-teacher one-classroom set up where student intake occurs every four years. Essentially, a four-year session starts in grade 1, with a group of approximately 30 children who are paired with one teacher. These same group of children then progress through years 2, 3, and 4 with

[1] Further details on BRAC can be found on BRAC's website: https://www.bracinternational.nl/en/

the same teacher they were paired with initially. Thus, a BRAC non-formal schooling model does not operate multiple grades with multiple teachers as is the case in formal schools. Once the children complete four years at BRAC schools, they can proceed to formal schools at the secondary level.

Further, once the four years are complete with one batch, the same teacher/branch starts with a new cohort; however, if there is insufficient demand, then the BRAC branch relocates to a new area where there is demand.

BRAC schools cover a five-year curriculum in four years using BRAC's own textbooks in grades 1 to 3, and NCTB (National Curriculum and Textbook Board) textbooks in grade 4 and 5.[2] Once they finish 4 years at BRAC (i.e. grades 1-5), BRAC graduates go on to sit for the national exams, the Primary School Certificate Exams (PSCE), and can subsequently integrate into the government schooling system or to other NGO schools. BRAC schools are free, i.e. , they do not charge tuition fees.

2.2 Comparing the Different School-types

Since this paper compares performance across different school-types, it is pertinent for us to briefly discuss their characteristics. We examine how BRAC, GOV, JAAGO and other NGO schools vary along dimensions such as selection criterion, teacher qualification, share of female teachers, class-size etc. We summarize school-type characteristics in Table 1. Since selection into the different school-types is central to our paper, we first give an overview of the selection process adopted by BRAC, GOV, JAAGO and other NGO schools in our sample slums.

BRAC's goal is to provide education to children who hail from poor households, remote areas and minority groups (De Nooijer, 2011).

GOV schools fall under the jurisdiction of the Ministry of Primary and Mass Education (MoPME). Such schools start at grade 1 and continue to grade 5, with increasing number of schools including a pre-primary section since 2010. For state run primary schools, according to the Compulsory Primary Education Act of 1990, all children aged six or above and living in the school's catchment area should be admitted

[2]Textbooks are distributed by the National Curriculum and Textbook Board which is under the Ministry of Education.

to the nearest government school.[3] However, given the large number of students in any given area, it is doubtful that GOV schools can accommodate all eligible students in a catchment area.

JAAGO uses three eligibility criteria outlined below when admitting students. A child admitted to JAAGO must fulfil the following criteria:

- He or she must be from a household where each family member earns less than or equal to BDT 2,000 (USD 23.6) per month;
- He or she must not be enrolled in other schools simultaneously;
- A child from a single parent household is given preference.

Additionally, JAAGO interviews both the prospective students and their parents as a part of the admission process.

Since detailed information for NGO schools is not readily available in published sources, we conducted a survey on how NGO schools operate in 2018 (two years after our original survey).[4] Based on our 2018 school survey, the non-BRAC NGO schools in our sample employ a variety of selection criteria which includes minimum age of entry, family income, children coming from single parent households, admission tests and a host of other miscellaneous criteria. These schools tend to use a combination of these admission criteria; what stands out however is that for a majority of the schools in our sample, child's age is the most important admission criteria, followed by family income. Most of these schools also use interviews as a screening mechanism with 83% of the NGO schools interviewing both parents and child while the remaining schools interview only the parents, but not the child.

In Table 1 we summarize the characteristics of the 4 school-types. First, as mentioned earlier, BRAC provides non-formal schooling, while JAAGO and GOV schools provide formal schooling; in other words, the latter have a multi-grade setup with different teachers assigned to each grade and/or subject. The NGO schools in our sample are a more heterogeneous mix, with some of them following BRAC's non-formal setup while some adopt the formal structure. Second, except for JAAGO all other school-types are Bengali medium. This means that all subjects are taught in Bengali except for the English language class. In contrast, at JAAGO, all subjects are taught in

[3]The catchment area is defined as the geographical area served by a school (Hallak 1977).

[4]In 2018 we successfully interviewed 1 NGO schools (16 non-BRAC NGOs and 2 BRAC schools) out of the 28 NGO schools that are represented in our sample.

Table 1: Comparing School-types

Characteristics	BRAC	GOV	JAAGO	other NGOs
Formal Schooling	×	✓	✓	✓ /×
Instruction in English	×	×	✓	×
Minimum teacher qualification - Bachelors Degree	×	×	✓	×
Teachers require strong command over English	×	×	✓	×
High level in-service training	×	✓	×	×
High share of female teachers	✓	×	✓	✓
High administrative monitoring	✓	×	✓	✓
High teacher salary	×	✓	×	×
Small class size	✓	×	✓	✓
Longer school days	✓	×	✓	✓
Longer school year	✓	×	✓	NA
Corporal punishment	×	✓	×	×

Notes: We leave a "NA" for cases/ characteristics where we do not have enough information.

English, except for the Bengali language class.

In terms of teacher qualification, BRAC and other NGO schools tend to have a laxer requirement. BRAC requires its teachers to have ten years of schooling, be willing to teach on a part-time basis and live in the community being served by the school. Majority of the NGO schools require their teachers to have at least an HSC degree which is equivalent to completing high school, i.e., twelve years of schooling. In the case of GOV schools, until 2013, the requirement for female teachers was at least an SSC degree (equivalent to completing 10 years of schooling), while requirement for male teachers was at least an HSC degree.[5] JAAGO only employs teachers with at least a bachelor's degree.

The GOV schools invest more in teacher training than the other school-types. About 90% of GOV school teachers had some form of training such as the Certificate in Education (C-in-Ed), the Diploma in Education (DipEd) or the Bachelors in Education (B-in-Ed). [6] The C-in-Ed is a one-year in-service training program designed for teachers teaching at the primary level while the Bachelors in Education (B-in-Ed) is a one-year in-service training program designed for teachers teaching at the secondary level. The Diploma in Education (DipEd) is an 18 month long basic training program,

[5]See Nath et al. (2015) and circulars published by Directorate of Primary Education (DPE).
[6]See Nath et al. (2015).

created to replace the C-in-Ed.

In particular, about 70-75 percent of GOV teachers having undergone the Certificate-in-Education in-service training program. [7] Additionally, all GOV school teachers undergo 6 one-day refresher courses throughout the year. In contrast, BRAC, NGO and JAAGO schools have shorter training programs. Specifically, BRAC teachers usually undergo a training program of 12-15 days before they start teaching; the average NGO schools in our sample provides about 7-14 days of pre-service training, while JAAGO teachers undergo a month long pre-service training program. All three school-types include some form of refresher courses throughout the year.

In terms of gender balance, the share of female teachers is about: 60-66% at GOV schools; approximately 80% at JAAGO school ; 98-99% at BRAC schools.[8]Similarly, 50% of our sampled NGO schools had an all female teaching staff.

When it comes to administrative monitoring, often measured by looking at teacher and headmaster absenteeism, amongst the school-types we consider, JAAGO is the strictest. At JAAGO, Education Coordinators (similar to school principals) oversee the day-to-day operations of the schools, monitoring teacher presence, ensure teachers sign in and out and linking salary to unpunctuality. At BRAC, administrative monitoring involves program officers visiting randomly assigned school at least twice a week (Chabbott, 2006). On the day of the visit, the monitor attends all the classes during school hours in order to evaluate lessons delivery and teacher-student interactions (Hanemann, 2017). Meanwhile, the non-BRAC NGO schools in our sample, practice some degree of monitoring by having teachers sign-in (and in some cases sign-out as well); in only one-third of these schools is teacher pay linked to teacher absenteeism. On the other hand, GOV schools use lax monitoring practices which leads to high absenteeism and unpunctuality by both teacher and headmasters.[9] For instance, teacher absenteeism was approximately 12% in 2008 based on unannounced visit on the day of the survey, it fell slightly to 11% in 2014 (Nath et al., 2015).

Contact hours varies across school-types with the average GOV school holding lessons for only 2-3.5 hours.[10] BRAC and NGO schools have slightly longer hours

[7]Nath et al. (2015), APSC (2014), APSC (2015) and APSC (2018).

[8]See Bangladesh Bureau of Statistics (2017), APSC (2015) and De Nooijer (2011).

[9]FMRP (2006), Nath and Chowdhury (2008), and Nath et al. (2015).

[10]See ASPR (2017) and FMRP (2006).

ranging from 3-4 hours, while the typical school day at JAAGO lasts for 4.5 to 5 hours (Chabbott, 2006). In terms of days-per-year, BRAC and JAAGO schools are open for about 265 days (De Nooijer, 2011). In contrast, the average school year is 228 days at GOV schools.[11]

GOV schools tend to have the largest class size, with the average class size ranging from 46 to 52 between 2006 and 2014 (Nath et al., 2015; FMRP, 2006). The other school-types have smaller classes, with the class size being around 30 at BRAC schools, JAAGO and NGO schools.[12]

3 Review of Schooling Literature

Needless to say, the education literature is vast. For our paper, two strands of this literature are relevant. One group of studies compares different school-types (e.g., public, private, religious schools) in terms of learning outcomes (Angrist et al., 2002, 2006; Muralidharan and Sundararaman, 2015; Kingdon, 1996; Goyal, 2009; French and Kingdon, 2010; Singh, 2015; Pal and Saha, 2014; Beegle and Newhouse, 2006; Chudgar and Quin, 2012). Another branch examines how one isolated educational component (e.g., teacher qualification, administrative monitoring, class size) affects student outcomes (Muralidharan and Sheth, 2016; Aslam and Kingdon, 2011; Fehrler et al., 2009).

While it is useful to understand how individual school features affect student outcomes, when the government or NGO establish schools, they rarely tweak only one or two educational components. Instead, they combine a number of educational components to create an educational intervention such as combining small class size with high monitoring, well qualified teachers, remedial classes etc.

Thus, our study is closely aligned with the first group of studies which compares performance across different school-types. Since we study BRAC schools, which operate only in developing countries, we restrict our attention to the schooling literature on developing countries. We include a detailed review of this literature in our earlier paper examining the differential impact of government, NGO and JAAGO schools (Hahn et al, 2020). Thus, to avoid repetition, in this paper, we summarize all rele-

[11]Length of the school year is 242 days officially, but in practice it is 228 days.

[12]JAAGO and NGO class size are based on our school-survey and while Kielland (2015) reports BRAC's average class size.

vant schooling studies in Tables 7 and 8 in the appendix; note that, in the interest of completeness, we include both group of studies mentioned above.

3.1 Schooling Literature on BRAC schools

After sifting through BRAC's own reports, we found only three reports that compare BRAC to other school-types (Nath et al., 2007; Nath and Hossain, 2017; Nath et al., 2005). The remaining papers that deal with some school-type comparisons compare BRAC schools in different locations or different BRAC cohorts over time (Nath, 2003; Yasmin et al., 2018; Nath, 2018).

Most BRAC papers use the same locally developed test instrument called the Achievement of Basic Competencies (Education Watch, 2000).[13] This test comprises of 27 competencies (subtests) on Bangla, English, Mathematics, Social studies, General Science and Religious studies.

3.2 BRAC's In-house Reports: BRAC versus Other School-types

Nath et al (2007) compares 1,181 students across 30 BRAC non-formal primary schools (600 students) to 30 government primary schools (581 students), in rural Bangladesh, using the competency-based instrument mentioned above. Note that they only use means and percentages of learning outcomes to draw comparisons between the 2 school-types; they do not use any identification strategy that controls for the selection across the two school-types. First, they find that a higher percentage of BRAC students (2.2%) have completed all 27 competencies as compared to government school students (0.9%). Second, in terms of mean number of competencies achieved, in the absence of conditioning variables, BPS students are slightly ahead of GPS students with BPS students completing 19.1 competencies and GPS students completing 18.6 competencies. Note however, that the authors do not specify if the difference in means is statistically significant. In terms of raw means, boys outrank girls within each school-

[13]Note that Education Watch comprises of a series of studies, on literacy and basic education in Bangladesh, conducted by an advocacy and campaign network called Campaign for Popular Education (CAMPE) (CAMPE 2008, pg 3). CAMPE itself is comprised of NGO schools, researchers, educators and Civil Society Organizations; established in 1991, their stated goal is to "achieve equitable and quality education and lifelong learning for all" (CAMPE 2021).

type; this gender difference is statistically significant. However, it is possible that this within school gender difference could disappear once conditioning variables like parents' education, family expenditure, use of private tutor etc are included.

Nath and Hossain (2017) compare school completers of BRAC's NFPE (non-formal primary schools) to a national sample of students who were part of the national literacy survey conducted by Education Watch in 2016. This national sample most likely consists of primary level students from government schools, NGO schools, English medium kindergartens (private schools) and madrassahs.[14] The sample consisted of 472 students from 30 randomly selected BRAC schools and 328 students from the national sample, all aged 11 or more at the time of the survey. Using a locally developed Literacy Assessment (Education Watch, 2002) they first compare literacy rate in terms of percentages; they find that the literacy rate was 95.7% among BRAC students and 91% among the national sample. After controlling for the child and family background, using multivariate logistic regression, they find that BRAC students are 2.61 times more likely to be literate than their counterparts in the national sample.[15]

Nath et al. (2005) compare 1594 students across three school-types: BRAC Non-formal Primary Education schools (NFPEs), Community Schools and Formal BRAC schools. They study 845 students from 60 BRAC NFPEs, 406 students from 29 community schools and 343 students from 11 BRAC formal schools. Community schools are established and run by the local community with financial assistance from the GOV.[16] In stark contrast to BRAC's NFPEs, BRAC formal schools are primary schools, established in permanent structures with five classrooms and five teachers.

Note that, once again, all school-type comparisons are done in terms of means and percentages of student performance in the Achievement of Competencies test developed by Education Watch (2000). Out of 27 competencies, in the absence of any conditioning variables, mean number of competencies achieved is the highest for BRAC formal school students (21.2) followed by community school students (19.2) and BRAC

[14]The national sample most likely included students from a mixture of school-types based on trends in earlier Education Watch reports (CAMPE 1999, 2008); Education Watch 2016 does not explicitly mention what school-types the national sample includes.

[15]Their conditioning variables include gender, age, mother's education, father's education, ethnicity, religion, having electricity at home, labor sales status of households and household food security status.

[16]They are usually two shift schools going from grades 1 to 5, with 3-4 teachers. Community schools follow the national curriculum using NCTB textbooks. They usually have 3 classrooms, while average class size is below 40.

NFPEs (18.9). However, the study does not specify if this mean difference is statistically significant.

4 Data Collection and Choice of Survey Instruments

4.1 Data Collection

Between 2015-2016, we collected our own data on 1802 slum children (aged 5 - 12) attending school in two slums of Dhaka. Our aim was to collect a stratified sample of approximately 600 students in GOV schools, JAAGO schools and NGO schools. In actuality, we collected data on 576 JAAGO students, 586 GOV students, 640 NGO student. As noted above we used this data to measure the performance of JAAGO students, GOV students, and NGO students, on a standard mathematics test. We also collected data on the family background and IQ of each student. Our sampling scheme is an example of choice-based sampling, which is commonly used by researchers when at least one of the outcomes is relatively rare.[17]

We chose this sampling scheme over a random sampling schemes, since the latter would have included few JAAGO students, as JAAGO schools teach only a very small fraction of the overall students. We collected the data by 'streets', where a street is a very long road and it includes smaller roads off this main road. We started with a street with a JAAGO student, then collected NGO and GOV students on the same street. We collected from 26 of these 'streets', i.e., we have 26 clusters in our sample. Thus, we adjust the standard errors for this cluster sampling following Abadie et al. (2017).[18]

We avoided selecting students from different neighborhoods; if we had included students from different neighborhoods, we would have had to deal both with selection into the different neighborhoods and into the different school types, while with our current design we only have to worry about selection by school type. Here

[17] For example, in transportation studies, an equal number of subjects can be interviewed at the bus station, train station, airport and parking garages, even though commuters using these modes make up different fractions of the population. A seminal paper in this literature is Manski and Lerman (1977).

[18] Cluster sampling by street or small area is widely used in data collection, for example in the National Longitudinal Survey data collection.

we are being guided by Heckman et al. (1997, 1998) and Heckman et al. (1998) finding that matching, our preferred technique, 'works better' when treatments and comparisons are from a common economic environment.

We interviewed children outside of school hours. Since schools operate in morning and day shifts, children attending morning shifts were interviewed in the afternoon hours and those attending day shifts were interviewed in the morning hours. Children were not interviewed during mealtimes. Enumerators were randomly assigned to different slum sub-areas and as well as across the times of the day.

When conducting our survey, we took many steps to insure that our data was of high quality. We hired experienced enumerators and invested heavily in monitoring. We hired a team of audio auditors who listened to each of the interviews to identify mistakes and/or 'cheating' on the survey instruments. Enumerators who were caught cheating were immediately dismissed and there were a nontrivial number of such enumerators. All interviews by these enumerators were redone by trusted enumerators. When conducting the household surveys and IQ tests, along with on-field monitors, we hired a team of data editors who specialized in reviewing the information in household questionnaires and verifying contradictory or missing information with the enumerators and surveyed households. Additionally, all survey instruments were graded twice by two separate data editors. Further, to minimize errors in data entry, we used a double entry system for our data.

4.2 Survey Instruments

When choosing tests that would enable us to measure mathematics achievement and IQ, we had two requirements. Given that our sample includes children aged between 5 and 12, with younger children more likely to be illiterate, we needed a test that (i) would not require literacy (i.e., tests that could be orally administered) and that (ii)would be suitable for the above age range. Note that GOV, BRAC and NGO students are taught in Bengali, while JAAGO students are taught in English. We used only the mathematics subtests since they is not as dependent on language skills and medium of instruction. However, we administered the tests in Bengali to GOV, BRAC and NGO students and administered the same tests to JAAGO students in 'Banglish' (i.e., kept technical terms in English).

To determine the appropriate achievement test for our purposes, we reviewed tests used in both the developed and developing country literature. Tests used in the developing country studies are usually developed locally, based on local curriculum, and are non-standardized with regard to international tests (Beegle and Newhouse, 2006; French and Kingdon, 2010; Goyal and Pandey, 2009b; Goyal, 2009; Hsieh and Urquiola, 2006; Lucas and Mbiti, 2014; Pal and Saha, 2014). This can make it quite hard to compare results across countries since one does not know if cross-country differences are a real phenomenon or simply the result of using different tests. Note that the two BRAC studies cited above use Bangladesh-specific tests.

In the developed country context, there are some standardized tests frequently used in the Economics, Education, Child Psychology and Development Psychology literature; these include the Woodcock Johnson Test of Achievement III (WJ-III), the Peabody Individual Achievement Test (PIAT), the Wechsler Individual Achievement Test (WIAT) and the Wide Range Achievement Test (WRAT).

We selected the Woodcock Johnson Test of Achievement III, since it is most frequently used in the literature, is applicable for the age range of 2 to 90+ and hence fully covers our age range of 5-12, is suitable for grades ranging from pre-primary to grade 5, and can be administered orally.[19] We used three orally administered Maths subtests: Test 10 (Applied Problems), Test 18A (Quantitative Concepts) and Test 18B (Number Series). From these three tests, we calculate the Maths Reasoning cluster, which captures a child's problem solving, analysis, reasoning and math vocabulary skills.

In order to capture fluid (innate) intelligence or IQ, we use the Kaufman Brief Intelligence Test-2 (K-BIT henceforth) since it can be administer orally and satisfies our age and grade requirement. The K-BIT is extensively used in the psychology literature as a measure of fluid intelligence, i.e., intelligence that is not affected by schooling unless schools 'teach to the test'. But certainly, some readers will worry that school type affects the KBIT scores, with better schools raising the K-BIT score by more than

[19]See e.g. Burchinala et al. (2014), Cameron et al. (2012), Crosnoe et al. (2010), Davis-Kean (2005), Denton et al. (2013), Duncan et al. (2007), Espy et al. (2004), Ferrer and McArdle (2004), Fitzpatrick et al. (2014), Floyd et al. (2006), Gormley et al. (2005), Harlaar et al. (2012), Mazzocco and Kover (2007), McClelland et al. (2007), McCormick et al. (2013), El Nokali et al. (2010), Patall et al. (2010), Pong and Landale (2012).

poorer schools. We show below that our estimated treatment effects will be understated in such a case.

Since both the Woodcock Johnson Test and the K-BIT scores tend to increase with age, we account for this by normalizing each score by age. For both the tests, we use their age adjusted z-scores. In other words for student i in age group a, for both the tests, we calculate $Z_i = (X_i - X_a)/\sigma_a$ where X_a and σ_a is the mean and standard deviation in age group a.

5 Selection: Who ends up at BRAC schools?

The first question we ask is if there is any selection across school-types in terms of student characteristics. Specifically, what type of students are being drawn to and admitted into these 4 school-types.

We investigate selection across different school-types in terms of 4 key variables – monthly family expenditure (deflated by equivalence scale), father's schooling, mother's schooling and K-BIT (IQ/fluid intelligence).[20] We use family expenditure as a proxy for family income, since families are often reluctant to reveal their monthly earnings but more readily share their average monthly expenditure with enumerators. The second and third variables of interest are father's schooling and mother's schooling measured in terms of highest grade completed. Finally we use our IQ measure to capture student attributes that would usually be treated as unobserved heterogeneity in a standard study. Our study is somewhat unique in the developing country schooling literature in terms of use of IQ as a conditioning variable. We find below that having data on IQ allows us to do a much better job of controlling for selection, and dramatically lowers our estimated treatment effects (but which are still quite substantial).

We first present the means for the pooled sample (pooled over boys and girls) for the 4 key variables in Table 2. Based on this table, we find that the average monthly expenditure per capita for our sample households ranges from approximately BDT 5,200 to BDT 6,200 across school types. Average father's schooling by school type

[20] An equivalence scale adjusts for family size but notes that family expenses will rise less than proportionately as we add another family member.

Table 2: Means by School-Type (Pooled Sample)

	BRAC	GOV	JAAGO	Other NGOs
Monthly Family Expdt	5.1830	6.1546	5.8473	5.4777
(in BDT 1000 adjusted by equivalence scale)	(0.1565)	(0.1202)	(0.1299)	(0.1279)
Father's schooling	2.3096	3.7025	3.6687	3.3101
	(0.1745)	(0.2307)	(0.1984)	(0.3085)
Mother's schooling	1.9372	3.2503	3.7995	2.9451
	(0.1366)	(0.2268)	(0.1913)	(0.2281)
K-BIT (IQ)	-0.5356	0.0663	0.2646	-0.1508
	(0.0686)	(0.0672)	(0.0798)	(0.0900)
Observations	239	586	576	401

Notes: (a) Standard errors in parentheses clustered at the street level; (b) For the IQ score, we use age adjusted Z-scores. In other words for student i in age group a, we calculate, $Z_i = \frac{X_i - X_a}{\sigma_a}$, where X_a and σ_a is the mean and standard deviation in age group a.

ranges between 2.3 and 3.7 years of schooling, while average mother's schooling ranges between 1.9 to 3.8 years of schooling. Further, K-BIT Z-scores range from -0.15 to 0.26. The means of these 4 variables follow a similar pattern when we disaggregate by gender as shown in Tables 9 and 10 in the appendix. In the calculation of standard errors, for the tables in this section, we account for our cluster sampling design.

To draw more meaningful comparisons, we look at the difference in means across school types. In Table 3, column (1) is obtained by subtracting the estimated GOV mean from the estimated BRAC mean for our four variables of interest. Column (2) subtracts the estimated JAAGO means from the estimated BRAC means, while column (3) subtracts the estimated NGO means from the BRAC means. Thus, in column (2), a positive (negative) sign is interpreted as BRAC outranking (being outranked by) JAAGO. Since all four of our control variables would be expected to raise math performance, if there is adverse selection into BRAC schools vis-a-vis the other school types, we expect the entries in Table 3 to be positive and statistically significant. This in fact is what we see, with the differences between BRAC and JAAGO students and between BRAC and GOV students being larger than the differences between the BRAC students and non-BRAC NGO students. Further manipulation of the data (not shown) demonstrates that the quality of GOV and JAAGO students is roughly equal, and both school-types have higher quality students (in terms of our conditioning variables) than the non-BRAC NGO students. When we replicate Table 3 results for separate samples, disaggregated by gender, as shown in Tables 11 and 12, we find a similar pattern to the pooled sample. To the best of our knowledge, this is the first documentation of how disadvantaged BRAC students are compared to those in other school-types in an urban context.

We would note that sorting students into the different school types reflects both the preferences of the parents and of the school, and that this allocation process does not depend on school fees. Because of these facts, it would be quite difficult to run a randomized trial allocating students to schools since the schools care about which students they are getting, and one cannot reduce fees for some students at certain school types.

Table 3: Mean Differences Across School-types (Pooled Sample)

	(1) BRAC vs GOV	(2) BRAC vs JAAGO	(3) BRAC vs Other NGOs
Monthly Family Expdt	-0.9716***	-0.6643***	-0.2947
(in BDT 1000 adjusted by equivalence scale)	(0.2067)	(0.1929)	(0.2066)
Father's Schooling	-1.3929***	-1.3591***	-1.0004***
	(0.2919)	(0.3119)	(0.2761)
Mother's Schooling	-1.3130***	-1.8622***	-1.0079***
	(0.2622)	(0.2480)	(0.2209)
K-BIT (IQ)	-0.6019***	-0.8002***	-0.3848***
	(0.0861)	(0.1189)	(0.0943)
Observations	825	815	640

Notes: (a) Standard errors in parentheses clustered at the street level; (b) We report the difference in means at the 1 percent, 5 percent and 10 percent significance level denoted by *** , ** , and * respectively; (c) For KBIT (IQ score), we report age adjusted Z-scores.

6 Identification Strategy: Propensity Score Matching and Choice Based Sampling

One way to deal with this selection problem is to use the Instrumental Variable Approach where we treat school-type as endogenous, assuming that a variable such as distance to the closest school of each type is a strong instrument. However, we do not use this approach because IV estimates are inconsistent in the presence of choice-based sampling (Solon, Haider, and Wooldridge, 2015). Further, adjusting IV estimator to make it consistent is unfeasible given our sample (Hahn, Ham and Ridder (2020)).

We argue that the strong control variables we have allow us to achieve the 'matching on observables' assumption underlying propensity score matching (PSM). Specifically, we will compare each school type to the other school types individually. Standard PSM will not produce consistent estimates in the prescence of Choice Based Sampling. However, we will use results in Heckman and Todd (2009) and Ham and Khan (2021) to obtain consistent estimates here. Following Heckman and Todd (2009) we match on log odds ratio (LOR) of the estimated propensity score and following **?** we appropriately evaluate the relevant treatment effects.

6.1 Log Odds Ratio Matching

To understanding how matching works, let us take a simple example. For expository purposes, in what follows, we let the BRAC individuals be the treatment students and GOV individuals be the comparison students. Let $D_i = 1$ if child i goes to BRAC and $D_i = 0$ if child i goes to a GOV school. Each BRAC student has the potential outcome Y_{1i} if she goes to BRAC and Y_{0i} if she goes to a GOV school. Due to selection, we cannot simply compare the mean outcomes of BRAC and GOV students.

However there may be observable X's such that once we condition on X, who goes to BRAC and GOV schools is like a coin toss. If we pick child i who is in BRAC and place her in a GOV school, her achievement will be the same as the weighted average of her closest counterparts in the GOV school (and vice versa). That is,

$$E\left(Y_{0i}|D=1,X\right)=E\left(Y_{0i}|D=0,X\right)$$

or,

$$Y_{0i}\perp D|X.$$

This is called the'Matching on Observables' or Conditional Independence Assumption (CIA). In order to counter the dimensionality problem brought about by a large number of $X's$, following Rosenbaum and Rubin (1985), given that CIA and the common support assumption (discussed later) hold, we match on estimated propensity score - a single index $P(x) = Pr(D = 1|X)$. That is,

$$E\left(Y_{0i}|D=1,X\right) = E\left(Y_{0i}|D=0,X\right)$$

implies

$$E\left(Y_{0i}|D=1,P(X)\right) = E\left(Y_{0i}|D=0,P(X)\right)$$

The fundamental problem of estimating treatment effects is that of incomplete information. Even though the researcher observes whether the treatment occurs and the outcome of treated individuals conditional on treatment, he/she cannot observe the outcome had the treated individual not been treated. In other words, the researcher cannot observe the counterfactual.

Studies using PSM generally calculate three treatment effects. The first effect is the Average Treatment on the Treated (ATT) effect, which asks what would have happened to treatments if they had not received treatment. The second effect is the Average Treatment on the Untreated (ATTU) effect, which asks what would have happened to the nonntreatments if they had received treatment. Finally, the third effect is the Average Treatment Effect (ATE) which examines what would have happened to an average member of the population if they had received treatment. In the absence of choice-based sampling, the ATE is a weighted sample of the ATT and ATU. With choice-based sampling the ATU is again a weighted average of the ATT and ATU except that they weights will depend on the true proportions of students going to each school type. For, students going to JAAGO and BRAC, the true respective proportions are miniscule and the ATE and the ATU are essentially equal. Thus, for the rest of the paper we do not consider estimating ATEs.

Continuing the example of BRAC vs. GOV case, consider the Average Treatment on the Treated (ATT) which captures the gains, on achievement, of going to BRAC as opposed to GOV schools for BRAC students.

The Average Treatment Effect on the Treated (ATT) is:

$$\frac{1}{N_1}\sum_{D_i=1}(Y_{1i}-\widehat{Y_{0i}}) = \frac{1}{N_1}\sum_{D_i=1}(Y_{1i}) - \frac{1}{N_1}\sum_{D_i=1}(\widehat{Y_{0i}})$$

where Y_{1i} denotes observed test score of child i going to BRAC and $\widehat{Y_{0i}}$ denotes predicted test score of BRAC child i if s/he had gone to GOV.

When estimating the ATT, we thus face the following problem. For BRAC students, we can observe

$$\frac{1}{N_1} \sum_{D_i=1} (Y_{1i})$$

But, we cannot observe

$$\frac{1}{N_1} \sum_{D_i=1} (Y_{0i}).$$

Hence, we must use an imputed value of Y_{0i}, denoted by $\widehat{Y_{0i}}$, to calculate the ATT. Here, we follow much of the literature and use local linear regression based on the log odds ratio to construct this counterfactual. Specifically we minimize the following objective function, where

$$\hat{q}(x_j) = ln \left(\frac{\widehat{P_i}}{1 - \widehat{P_i}} \right)$$

$$\min_{\beta_0, \beta_1} \sum_{j=1}^{N_2} \left\{ Y_j - \beta_{0i} - \beta_{1i} \left[\hat{q}(x_j) - \hat{q}(x_i) \right] \right\}^2 K\Big(\frac{\hat{q}(x_j) - \hat{q}(x_i)}{h} \Big)$$

In the above equation, $K(.)$ is the kernel weighting function, h is the bandwidth, and j refers to GOV students whose total number is N_2. We then set $\widehat{Y_{0i}} = \beta_{0i}$ to obtain the cunterfactual.

6.2 Trimming Methods and Common Support

Next, we need to trim the data to achieve common support. What do we mean by that? Continuing with the example of BRAC vs. GOV students, essentially, we do not want to estimate the ATT for BRAC vs. GOV where there are where there are very few or no GOV students around each BRAC child in tersm of estimated log odds ratios.

We use 2 methods to obtain common support which we denote by Common Trim and 1:5 Trim. Common Trim, the default in psmatch2, imposes a common support by dropping treatment (control) observations whose LOR is higher than the maximum or less than the minimum LOR of the controls (treatments), when estimating the

ATT (ATU). [21] In contrast, in the 1:5 Trim, we impose common support by dropping treatment and control observations in an interval where the ratio of treated to control observations is more than 5 or less than 0.2. In both cases, this helps to ensure that there are comparable GOV students for each BRAC student, but the 1:5 Trim approach is more stringent.

6.3 Matching and Conditioning Variables

When it comes to choosing conditioning variables, we usually want to match on pre-baseline $X's$ that affect both school allocation [22] and student achievement; ones hope is that conditional on X, whether a child goes to the treatment or comparison school can be treated as random. It is important to stress that the causality must run from the pre-baseline $X's$ to the schooling outcome and not vice-versa. We want pre-baseline $X's$ i.e. covariates that are not affected by school choice. However, we could not collect pre-baseline covariates before BRAC (and JAAGO schools) started operations, we must use post-baseline $X's$ instead that are credibly not affected by the schooling outcome.

Specifically, we condition on a rich set of covariates which includes our fluid intelligence/IQ measure (K-BIT z-scores), family size, gender, child's age, father's absence, father's schooling, and mother's schooling. We cannot rule out the possibility that family expenditure may be affected by school choice because the school day differs across schooling type, and hence the school-type may affect the parents' earnings and thus may bias our estimates. Thus we exclude monthly family expenditure as a matching covariate.

One might also be concerned that the school type will affect our fluid IQ measure, although in principle it should not. But as long as better schools raise IQ by more than less able schools, students with a given IQ in the poorer schools will be of higher quality than students in the better students, and this will bias our treatment effects downward.

[21] The psmatch2 routine is a widely used in Stata to implement PSM

[22] We do not use the term 'school choice' here since the allocation of a child to a school is a joint decision between the parents and the school.

7 Matching Results

In this section we look at how BRAC schools fare against their GOV, JAAGO and NGO counterparts, once we control for selection. The ATTs for different comparisons and different conditioning variables are shown in Table 4. As we discuss below, this table highlights two important findings; first, it shows how important it is to control for IQ and second, it gives us an overview of school-type effects when BRAC is compared to other school-types in the slums.

Table 4: **Estimating ATT using Matching to Control for Selection (Data Driven Bandwidth using Epanechnikov Kernel)**

	Dependent Variable - Achievement Test Z-Score		
	(1) **Mean Difference (no controls)**	(2) **Only Family Background (no IQ)**	(3) **Family Background & K-BIT**
BRAC vs GOV	-0.5465***	-0.6118***	-0.2723**
	(0.1211)	(0.1220)	(0.1121)
p-value		0.0000	0.0151
bandwidth		0.53	0.55
BRAC vs JAAGO	-0.7311***	-0.8049***	-0.4483***
	(0.1292)	(0.1264)	(0.1156)
p-value		0.0000	0.0001
bandwidth		0.61	0.49
BRAC vs Other NGOs	-0.3175***	-0.2481**	-0.0365
	(0.1107)	(0.1120)	(0.1108)
p-value		0.0267	0.7419
bandwidth		0.49	0.32

Notes: (a) Bootstrapped standard errors in parentheses clustered at the street level; (b) We report Treatment Effects at the 1 percent, 5 percent and 10 percent significance level denoted by *** , ** , and * respectively; (c) Family background matching covariates consist of gender, child's age, family size, father absence dummy, father's schooling and mother's schooling; (d) For sample size, refer to Table 2.

7.1 Matching and Controlling for Selection

First, to see the importance of IQ in controlling for selection we compare columns (1), (2) and (3) by turn. Column (1) displays the difference in means in achievement z-scores (note that no controls are included here) while columns (2), displays the treatment effects when we include family background variables (i.e., gender, child's age, father's absence, family size, father's schooling and mother's schooling) as matching covariates. Finally, in columns (3), we add in our measure of IQ, the K-BIT test scores, to our standard set of matching covariates.

When we compare columns (1) and (2) we find a relatively small change in the estimated effects. However, once we compare columns (2), (3) we see that in most school-type comparison cases, inclusion of K-BIT (as a measure of IQ) leads to a larger fall in ATT. Thus, our results suggest that family background does not play much of a role in controlling for selection, but including IQ has a considerable impact on controlling for selection.

7.2 Matching and the impact of BRAC schools

When it comes to examining school-type effects, we refer to our preferred specification where the IQ measure is included as a matching covariate along with the family background variables (refer to column (3) of Table 4).

When we compare BRAC to GOV schools, we see that after controlling for selection, students are worse off at BRAC schools as compared to GOV schools. This is shown by the fact that achievement test scores falls by 0.2723 of a standard deviation in achievement (σ_a), with a standard error of 0.1121, from going to BRAC instead of GOV schools. Similarly, when we compare BRAC to JAAGO schools, we find that students are worse off at BRAC than at JAAGO; mean achievement falls by $0.4483\sigma_a$ (standard error: 0.1156).

Lastly, when we compare BRAC to the other NGO schools in the two slums, the negative but statistically insignificant ATT in column (3) shows that students are no worse off at BRAC than at NGO schools. This is interesting given that when we compared raw achievement before controlling for selection, BRAC students did worse than NGO students.

In an attempt to unpack these treatment effects, we separate our sample by gender and rerun the estimations. We summarize our results in Table **??**. When we compare BRAC to GOV schools, we see no significant difference in math achievement between girls at either of these school-types. However, boys are worse off at BRAC than at GOV schools. This suggests that the impact we saw in the BRAC vs GOV comparison in the pooled sample is driven mainly by the negative effect of BRAC on boys.

When we compare BRAC to JAAGO schools, we see that girls are better off at JAAGO than at BRAC schools in terms of their learning outcomes. In contrast, there is no significant difference for boys at either of these school-types.

Lastly, consistent with the pooled sample, even when we separate by gender, we cannot find any difference between BRAC and NGO schools for either boys or girls.

Table 5: **Estimating ATT using Matching to Control for Selection (Data Driven Bandwidth using Epanechnikov Kernel)**

	Dependent Variable - Achievement Test Z-Score		
	(1) **Mean Difference (no controls)**	(2) **Only Family Background (no IQ)**	(3) **Family Background & K-BIT**
BRAC vs GOV (girls)	-0.4026‘***	0.2666*	-0.0777
	(0.1242)	(0.1520)	(0.1924)
p-value		0.0793	0.6864
bandwidth		0.16	0.27
BRAC vs GOV (boys)	-0.7069***	-0.8244***	-0.5464**
	(0.1488)	(0.1790)	(0.2787)
p-value		0.0000	0.0499
bandwidth		0.18	0.3
BRAC vs JAAGO (girls)	-0.6509***	-0.6669***	-0.4111**
	(0.1210)	(0.1541)	(0.1652)
p-value		0.0000	0.0128
bandwidth		0.11	0.22
BRAC vs JAAGO (boys)	-0.8279***	-0.8664***	-0.2041
	(0.1678)	(0.1741)	(0.2187)
p-value		0.0000	0.3508
bandwidth		0.25	0.31
BRAC vs Other NGOs (girls)	-0.2430***	-0.1359	0.0520
	(0.1144)	(0.1362)	(0.1554)
p-value		0.3183	0.7381
bandwidth		0.09	0.17
BRAC vs Other NGOs (boys)	-0.4151***	-0.3831*	-0.1896
	(0.1605)	(0.2114)	(0.1973)
p-value		0.0700	0.3366
bandwidth		0.17	0.25

Notes: (a) Bootstrapped standard errors in parentheses clustered at the street level; (b) We report Treatment Effects at the 1 percent, 5 percent and 10 percent significance level denoted by *** , ** , and * respectively; (c) Family background matching covariates consist of child's age, family size, father absence dummy, father's schooling and mother's schooling; (d) For sample size, refer to Tables 9 and 10.

7.3 Balancing Tests

But, can we get a signal if matching is appropriate here? In other words, to determine if the CIA holds, we carry out balancing tests. Given the log odds ratio, we look for a treatment effect on the X's since we should not see one. There are many ways of doing balancing tests, see, e.g. Smith and Todd (2005), Dehija (2005) etc. Ours is another approach which has the advantages that it takes into account the fact that the propensity score is estimated.

Similar to obtaining the treatment effects for school-type comparisons, for the balancing tests, we use local linear matching and adjust for the choice based sampling by matching on log odds ratio of the estimated propensity score. Note that we use the same trimmed sample, bandwidth type and kernel-type in the balancing tests as in the matching exercises. The balancing tests again use bootstrapped standard errors clustered at the street level. Recall that, without matching, there are big differences in the conditioning variables, i.e. raw values do not balance.

We pass the balancing test for all variables when we use the full model to estimate the log odds ratio; on the other hand, if we use only family background variables (excluding IQ) to estimate the log odds ratio, we fail the balancing test for variables like K-BIT.

Table 6: **Balancing Tests at the 5% level for Matching Estimators using Data Driven Bandwidth and the Epanechnikov Kernel**

	(1) LOR estimated using correctly specified model (including IQ)	(2) LOR estimated using misspecified model (excluding IQ)
BRAC vs GOV	0	1
BRAC vs JAAGO	0	1
BRAC vs Other NGOs	0	1

Notes: (a) Log Odds Ratio (LOR) estimated using the misspecified model uses 5 covariates. This includes child's age, gender, family size, father absence dummy, father's schooling and mother's schooling; (b) Log Odds Ratio (LOR) estimated using correctly specified model uses all 7 matching covariates. This includes the standard set of 6 family background variables mentioned in (a) as well as IQ measures, i.e., K-BIT Z-scores.

8 Conclusion

BRAC, a Bangladeshi NGO, in recent years has become global sensation. With the aim of providing schooling to marginalized communities, BRAC's model of non-formal primary schooling has been replicated by many NGOs and even the Government of Bangladesh. But given its 30 year long tenure and its wide coverage, it is surprising that there are few independent evaluations of BRAC schools. Our is in fact one of the first arm's length evaluations of BRAC.

We use high quality data we collected ourselves in two slums of Dhaka city between 2015 and 2016. We use choice based sampling while collecting our data. This is common in evaluation studies both in terms of cost effectiveness and ensuring sufficiently high frequency of program participants (especially for rare events). To account for this choice based sampling, we estimate treatment effects using propensity score matching which can be adjusted for choice based sampling (see Heckman and Todd (2009)).

To our knowledge, our study is the first to highlight how disadvantaged BRAC students are in terms of different child and family characteristics. One of the key findings of this paper is that BRAC attracts 'weaker' students. Students with higher IQ and relatively better educated parents end up at JAAGO, GOV and other NGO schools. Furthermore, we the same sorting patterns when we separate by gender.

Additionally, we see that commonly used family background variables do not play much of a role in controlling for selection. On the other hand, fluid intelligence plays a crucial part in controlling for selection. Including fluid intelligence, as measured by K-BIT, which most developing country studies fail to account for, substantially reduces bias.

Using log odds matching to control for selection, we find that comparable students perform worse at BRAC than at GOV and JAAGO schools. We find no difference between students at BRAC and other NGO schools in terms of their math achievement. When we disaggregate by gender, we see that boys perform worse at BRAC than at GOV; but we find no significant difference between boys for the BRAC vs JAAGO and BRAC vs NGO comparisons. On other other hand, once we control for selection, we find that girls perform worse at BRAC than at JAAGO; but there is no significant difference between girls for the BRAC vs GOV and BRAC vs NGO comparisons.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2017). Sampling-based vs. design-based uncertainty in regression analysis. *Working Paper Stanford University*.

Andrabi, T., J. Das, A. Khwaja, T. Vishwanath, and T. Zajonc (2007). Learning and educational achievements in Punjab schools (LEAPS): Insights to inform the education policy debate. *World Bank*.

Angrist, J., E. Bettinger, E. Bloom, E. King, and M. Kremer (2002). Vouchers for private schooling in Colombia: evidence from a randomized natural experiment. *American Economic Review 92*(5), 1535–1558.

Angrist, J., E. Bettinger, and M. Kremer (2006). Long-term educational consequences of secondary school vouchers: evidence from administrative records in Colombia. *American Economic Review 96*(3), 847 – 862.

APSC (2014). Annual primary school census. *Ministry of Primary and Mass Eductaion, Government of Bangladesh*.

APSC (2015). Annual primary school census. *Ministry of Primary and Mass Eductaion, Government of Bangladesh*.

APSC (2018). Annual primary school census. *Ministry of Primary and Mass Eductaion, Government of Bangladesh*.

Aslam, M. and G. Kingdon (2011). What can teachers do to raise pupil achievement? *Economics of Education Review 30*(3), 559 – 574.

ASPR (2017). Bangladesh Primary Education Annual Sector Performance Report - 2017.

Banerjee, A., R. Banerji, E. Berry, J.and Duflo, H. Kannan, S. Mukherji, M. Shotland, and M. Walton (2015). Teaching at the right level: Evidence from randomized evaluations in India. *MIT*.

Banerjee, A., R. Banerji, E. Duflo, R. Gelnnerster, and S. Khemani (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy 2*(1), 1–30.

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

Banerjee, A., S. Jacob, M. Kremer, J. Lanjouw, and P. Lanjouw (2004). How much would universal primary education cost? Unpublished Manuscript.

Bangladesh Bureau of Statistics (2017). Education Scenario in Bangladesh:Gender Perspective.

Beegle, K. and D. Newhouse (2006). The effect of school type on academic achievement: evidence from Indonesia. *Journal of Human Resources 41*(3), 529–557.

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economics of Education Review 28*, 629–640.

Burchinala, Vernon-Feagansa, L., V. Vitiellob, M., M. Greenbergc, and T. F. L. P. K. Investigators (2014). Thresholds in the association between child care quality and child outcomes in rural preschool children. *Early Childhood Research Quarterly 29*, 41–51.

Cabezas, V., J. Cuesta, and F. Gallego (2011). Effects of short-term tutoring on cognitive and non-cognitive skills: Evidence from a randomized evaluation in Chile. *JPAL*.

Cameron, C. E., W. M. Murrah, D. Grissmer, L. L. Brock, L. H. Bell, S. L. Worzalla, and F. J. Morrison (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child Development 83*(4), 1229–1244.

Chabbott, C. (2006). Meeting EFA: Bangladesh Rural Advancement Committee (BRAC) Primary Schools.

Chin, A. (2005). Can redistributing teachers across schools raise educational attainment: Evidence from operation blackboard in India. *Journal of Development Economics 78*(2), 384–405.

Chudgar, A. and E. Quin (2012). Relationship between private schooling and achievement: Results from rural and urban India. *Economics of Education Review 31*, 376–90.

Crosnoe, R., R. J. Wirth, R. C. Pianta, T. Leventhal, K. M. Pierce, and N. E. C. C. R. Network (2010). Family socioeconomic status and consistent environmental stimulation in early childhood. *Child Development 81*(3), 972–987.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology 19*(2), 294–304.

De Nooijer, P. (2011). The two-pronged approach:Evaluation of Netherlands support to primary education in Bangladesh.

Denton, C. A., S. Vaughn, T. D. Tolar, J. M. Fletcher, A. E. Barth, and D. J. Francis (2013). Effects of tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology 105*(3), 633–648.

Dixon, P., A. J. Egalite, and P. J. Wolf (2015). Private school choice in developing countries: experimental results from Delhi, India. In P. Dixon, S. Humble, and C. Counihan (Eds.), *Handbook of International Development and Education*, pp. 456 – 471. MIT Press.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from randomized evaluation in Kenya. *American Economic Review 101*(5), 1739–1774.

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

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

Duncan, G. J., A. Claessens, A. C. Huston, L. S. Pagani, M. Engel, H. Sexton, C. J. Dowsett, K. Magnuson, P. Klebanov, L. Feinstein, J. Brooks-Gunn, K. Duckworth, and C. Japel (2007). School readiness and later achievement. *Developmental Psychology 43*(6), 1428–1446.

El Nokali, N. E., H. J. Bachman, and E. Votruba-Drzal (2010). Parent involvement and children's academic and social development in elementary school. *Child Development 81*(3), 988–1005.

Espy, K. A., M. M. McDiarmid, M. F. Cwik, M. M. Stalets, A. Hamby, and T. E. Senn (2004). The contribution of executive functions to emergent Mathematic skills in preschool children. *Developmental Neuropsychology 26*(1), 465–486.

Fehrler, S., K. Michaelowa, and A. Wechtler (2009). The effectiveness of inputs in primary education: Insights from recent student surveys in Sub-Saharan Africa. *Journal of Development Studies 45*(9), 1545–1578.

Ferrer, E. and J. J. McArdle (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology 40*(6), 935–952.

Fitzpatrick, C., R. D. McKinnon, C. B. Blair, and M. T. Willoughby (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children? *Learning and Instruction 30*, 25–3.

Floyd, R. G., R. Bergerson, and V. C. Alfonso (2006). Cattell–Horn–Carroll cognitive ability profiles of poor comprehenders. *Reading and Writing 19*, 427–456.

FMRP (2006). Social sector performance surveys: Primary education final report. *Financial Management Reform Programme*.

French, R. J. and G. Kingdon (2010). The relative effectiveness of private and government schools in rural India: Evidence from ASER data. *Institute of Education Working Paper*.

Glewwe, P., M. Kremer, and S. Moulin (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics 1*(1), 112–135.

Gormley, W. T., T. Gayer, D. Phillips, and D. Brittany (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology 41*(6), 872–884.

Goyal, S. (2009). Inside the house of learning: relative performance of public and private schools in Orissa. *Education Economics 17*(3), 315 – 327.

Goyal, S. and P. Pandey (2009a). Contract teachers. *South Asia Human Development Sector Series No. 28. World Bank*.

Goyal, S. and P. Pandey (2009b). How do government and private schools differ? Findings from two large Indian states. *World Bank Working Paper*.

Ham, J. and S. Khan (2021). Selection, Gender and the Impact of Schooling Type in the Dhaka Slums: Much Better Data Matters! *Working Paper*.

Hanemann, U. (2017). Brac education programme, Bangladesh.

Harlaar, N., P. S. Dale, Y. Kovas, and S. A. Petrill (2012). Mathematics is differentially related to reading comprehension and word decoding: Evidence from a genetically sensitive design. *Journal of Educational Psychology 104*(3), 622–635.

Heckman, J. J., H. Ichimura, J. Smith, and P. E. Todd (1998). Characterizing selection bias using experimental data. *Econometrica 66*(5), 1017 – 1098.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies 64*(4), 605 – 654.

Heckman, J. J., H. Ichimura, and P. E. Todd (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies 65*(2), 261–294.

Heckman, J. J. and P. E. Todd (2009). The estimation of choice probabilities from choice-based samples. *The Econometrics Journal 12*(S1), S230–S234.

Hsieh, C. and M. Urquiola (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics 90*(8-9), 1477–1503.

Kielland, A. (2015). Paper commissioned for the EFA Global Monitoring Report 2015, Education for All 2000-2015: achievements and challenges.

Kingdon, G. (1996). The quality and efficiency of private and public education: a case study of urban India. *Oxford Bulletin of Economics and Statistics 58*(1), 56–82.

Kingdon, G. and F. Teal (2010). Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach. *Journal of Development Economics 91*(2), 278 – 288.

Kingdon, G. G. (2006). Teacher characteristics and student performance in India: A pupil fixed effects approach. *Global Poverty Research Group (GPRG) Working Paper Series 059*.

Lakshminarayana, R., A. Eble, P. Bhakta, C. Frost, P. Boone, D. Elbourne, and V. Mann (2013). The Support to Rural India's Public Education System (STRIPES) trial: A cluster randomised controlled trial of supplementary teaching, learning material and material support. *PLOS ONE 8*(7).

Lucas, A. M. and M. M. Mbiti (2014). Effects of school quality on student achievement: discontinuity evidence from Kenya. *American Economic Journal: Applied Economics 6*(3), 234 – 263.

Manski, C. and S. Lerman (1977). The estimation of choice probabilities from choice-based samples. *Econometrica 45*(8), 1977–1988.

Mazzocco, M. M. and S. T. Kover (2007). A longitudinal assessment of executive function skills and their association with Math performance. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence 13*(1), 18–45.

McClelland, M. M., C. M. Connor, A. M. Jewkes, C. E. Cameron, C. L. Farris, and F. J. Morrison (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and Math skills. *Developmental Psychology 43*(4), 947–959.

McCormick, M. P., E. E. O'Connor, E. Cappella, and S. G. McClowry (2013).

Teacher–child relationships and academic achievement: A multilevel propensity score model approach. *Journal of School Psychology 51*, 611–624.

Muralidharan, K., J. Das, A. Holla, and A. Mohpal (2014). The fiscal costs of weak governance: Evidence from teacher absence in India. *NBER Working Paper 20299*.

Muralidharan, K. and K. Sheth (2016). Bridging education gender gaps in developing countries: The role of female teachers. *Journal of Human Resources 51*(2), 269–279.

Muralidharan, K. and V. Sundararaman (2010). The impact of diagnostic feedback to teachers on student learning: Experimental evidence from India. *Economic Journal 120*(546), F187–F203.

Muralidharan, K. and V. Sundararaman (2013). Contract teachers: Experimental evidence from India. *NBER Working Paper Series 19440*.

Muralidharan, K. and V. Sundararaman (2015). The aggregate effect of school choice: evidence from a two-stage experiment in India. *Quarterly Journal of Economics 130*(3), 1101–1166.

Nath, S. (2003). Achievement of competencies of the students of BRAC non-formal primary schools, 2000-2002. *BRAC Research and Evaluation Division*.

Nath, S. (2018). Students 'achievement of competencies at transitional period: A study on BRAC primary schools'. *BRAC Research and Evaluation Division Research Monograph Series 79*(September).

Nath, S. and A. Chowdhury (2008). State of primary education in Bangladesh. *Education Watch 2008 CAMPE*, 94;99–103.

Nath, S., A. Chowdhury, K. Ahmed, J. Rahman, M. Ahmed, and R. Choudhury (2015). Moving from MDG to SDG Accelerate Progress for Quality Primary Education. *Education Watch 2015 CAMPE*, 33–99.

Nath, S. and A. Hossain (2017). Literacy status of BRAC primary school completers. *BRAC Research and Evaluation Division Working Paper*.

Nath, S., G. Roy, N. Dutta, and A. Hossain (2007). Achievement of primary competencies: A comparison between government and BRAC schools. *BRAC Research and Evaluation Division Research Report*.

Nath, S., M. Shahjamal, R. Yasmin, Z. Tata, and M. Kabir (2005). Achievement of competencies of the BRAC school students. *BRAC Research and Evaluation Division Research Report*.

Orkin, K. (2013). The effect of lengthening the school day on children's achievement in Ethiopia. *YL-WP 119*.

Pal, S. and B. Saha (2014). In 'trusts' we trust: Socially motivated private schools in Nepal. *IZA Discussion Paper*.

Patall, E. A., H. Cooper, and A. B. Allen (2010). Extending the school day or school year: A systematic review of research (19852009). *Review of Educational Research 80*(3), 401–436.

Pong, S. and N. S. Landale (2012). Academic achievement of legal immigrants' children: The roles of parents' pre- and postmigration characteristics in origin-group differences. *Child Development 83*(5), 1543–1559.

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

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.

Smith, J. and P. Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Economterics 125*(1-2), 305 – 353.

Solon, G., S. Haider, and J. Wooldridge (2015). What are we weighting for? *The Journal of Human Resources 50*(2), 301–316.

Somers, M. A., P. J. McEwan, and J. D. Willms (2004). How effective are private schools in Latin America? *Comparative Education Review 48*(1), 48–69.

Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia. *The Review of Economics and Statistics 88*(1), 171–177.

Urquiola, M. and E. Verhoogen (2009). Class-size caps, sorting, and the regression discontinuity design. *American Economic Review 99*(1), 179–215.

Yasmin, N., S. Nath, and A. Hossain (2018). Leveraging urban inclusion through education: Study on BRAC Slum School Project. *BRAC Research and Evaluation Division Working Paper*.

A Tables: Literature Review of Educational Interventions

Table 7: Literature Review: Linking Schooling as Packages to Student Outcomes

Study	Country	Intervention	Methodology	Findings
		EXPERIMENTAL STUDIES		
Angrist et al. (2002)	Colombia	► Studies effect of PACES voucher program three years after application. ► Vouchers subsidized private school attendance, renewable annually, conditional on grade advancement.	► RCT; ► Maximum likelihood; ► Random Effects; ► OLS; ► IV.	► Lottery winners have higher test scores, higher schooling attainment and are more likely to complete eighth grade. ► Gender differential - girls benefit from the voucher program in terms of schooling attainment; boys indifferent.
Angrist et al. (2006)	Colombia	► Tests effect of PACES voucher program on high-school graduation rates seven years after vouchers first administered. ► Vouchers subsidized private school attendance, renewable annually, conditional on grade advancement.	► Maximum Likelihood; ► OLS; ► Modified Tobit (parametric approach); ► Nonparametric bounds.	► Raises high-school graduation rates and test scores in national university entrance exam. ► Significant test score gains at both upper and lower tail bounds of score distribution.

Study	Country	Intervention	Methodology	Findings
Muralidharan and Sundararaman (2015)	India	► Voucher program; allocate vouchers using 2-stage lottery system. ► Provide vouchers for students to finance attending private schools.	► Estimate ITT using OLS; ► Estimate ATT by scaling up ITT estimates with inverse of takeup rate; ► Use IV to distinguish between English-medium private schools and Telugu-medium private schools.	► No evidence that voucher program improves test scores for Telugu (native language), Math, Social Studies, English. ► Private schools achieve results comparable to public schools in Telugu and Math at lower cost and less instructional time. ► Students from public schools who switch to Telugu-medium schools score better (Telugu, Math and Social Studies) than students who switch to English-medium schools.
Dixon et al. (2015)	India	► Voucher program; subsidize cost of private schools. ► Vouchers allocated via lottery.	► RCT; ► OLS; ► IV.	► Positive private school effect on English; no impact on Math and Hindi. ► Program benefits girls more than boys.
Hsieh and Urquiola (2006)	Chile	► Voucher program; provide vouchers to any student wishing to attend private schools.	► OLS; ► IV.	► No evidence that voucher program improved test scores, repetition rates and years of schooling. ► Voucher program led to cream skimming. More well-off and smarter students sorted into private schools.
NON-EXPERIMENTAL STUDIES				

Study	Country	Intervention	Methodology	Findings
Beegle and Newhouse (2006)	Indonesia	► Compare performance of public school students with private school students on national exams. ► Compare public madrassah with different private schools, i.e., secular, Madrassah, other Muslim, non-Muslim religious schools.	► OLS; ► Fixed Effects; ► IV.	► Public school students outperform private school students in National Exams. ► Public Madrassahs and non-Muslim religious private schools outperform private secular and private Muslim schools. ► High ability students benefit the most from attending public schools and non-Muslim religious private schools.
Kingdon (1996)	India	► Compare achievement across privately unaided (PUA), privately aided (PA) and government (G) schools.	► OLS. ► Heckman's two-stage correction.	► Positive private school effect. ► PUA students outperform PA and G students in Mathematics, at less than half the cost for PA and G schools. ► G students outperform PA students in Mathematics. ► All schools perform at roughly the same level for reading.
Goyal (2009)	India	► Compare privately unaided schools with public schools;	► OLS; ► Specialized method by Altonji, Elder Taber (2005) to gauge bias in data.	► Private school students outperform public school students in reading and math.
French and Kingdon (2010)	India	► Determining private school effect relative to government schools.	► OLS; ► Fixed effects at the state, district, village and household level; ► Method by Altonji, Elder and Taber (2005).	► Positive private school effect on combined test scores in math and reading.

Study	Country	Intervention	Methodology	Findings
Singh (2015)	India	► Examine the effectiveness of various private schools (English-Medium and Telugu-Medium private schools) relative to government schools.	► Dynamic OLS Value-Added Model.	► For the 8-10 year old cohort (rural areas), positive private school effect in English and heterogeneous effects in Telugu (native language) test scores; no effect on math scores. ► For the 15 year old cohort (rural areas), positive impact of private schools on math and Telugu scores. ► No evidence of a significant private school effect in urban areas.
Pal and Saha (2014)	Nepal	► Compare achievement in socially motivated trust-run private schools, profit-motivated company-run private schools and traditional public schools;	► Heckman's two-stage correction.	► Use standardized test scores for the school leaving certificate (SLC) as the outcome variable. ► There is a positive private school effect; which school performs the best depends on expenditure level. ► Trust-run schools come at the top followed by company-run schools and public schools last, when they control for household and individual characteristics (but not expenditure).
Somers et al. (2004)	► 10 Latin American countries.[23]	► Compare Private Schools to Public Schools.	► Multilevel modeling for each country. ► Meta-analysis using multilevel modeling across countries to obtain the general private school effect for the entire region.	► Use math and language test scores as outcome variable. ► After controlling for peer-group characteristics, mean private school effect around zero across 10 countries.

[23] Argentina, Bolivia, Brazil, Chile, Colombia, Dominican Republic, Mexico, Paraguay, Peru and Venezuela.

Study	Country	Intervention	Methodology	Findings
Chudgar and Quin (2012)	India	► Assess private school effect (rural and urban settings). ► Assess effect for low-fee and high-fee private schools. ► Outcome variables - reading, writing and mathematics scores.	► OLS; ► Logit; ► Propensity score matching.	► No significant difference between public and private schools in terms of achievement using matched data. ► Distinguish between high and low fee private schools; using regression framework they find no difference between the latter and public school students.
Andrabi et al. (2007)	Pakistan	► Studies the private-public school difference. ► Focuses on grade 3 students.	► OLS.	► Private schools outperform government schools using test scores as outcome variable.
Andrabi et al. (2010)	Pakistan	► Studies the private-public school difference. ► Focuses on students from grades 3 to 5.	► OLS with village-fixed effects; ► IV.	► Private school students outperform public school students in terms of achievement. ► Private school students have better civic skills, pro-democratic values and less gender bias.
Andrabi et al. (2011)	Pakistan	► Examines the private-public school difference incorporating the central role of persistence in learning models. ► Focuses on students from grades 3 to 5.	► Lagged value-added model using OLS and 2SLS; ► Differenced dynamic panel using GMM; ► Difference-in-Differences Strategy.	► A fifth to a half of learning persists between grades. ► Private schools increase average achievement by 0.25 standard deviations each year.
Bold et al. (2013)	Kenya	► Compare public to private schools.	► OLS with district and year fixed effects; ► Aggregate data over public and private schools for each district, separated by gender.	► Uses scores from the Kenya Certificate of Primary Education (KCPE) examination as the outcome variable. ► Private school effect is large and statistically significant after controlling for selection.

Table 8: Literature Review: Linking Schooling Characteristics to Student Outcomes

Study	Country	Characteristic/Component	Methodology	Findings
		TEACHER QUALIFICATION & EFFORT		
Thornton et. al (2018)	Uganda	► Teacher effectiveness. ► Characteristics of classes by effective teachers. ► Teacher training and pedagogy program called Northern Uganda Literacy Program (NULP).	► RCT - students randomly assigned to classrooms; ► Value-added estimation used for classroom and teacher effects.	► Increasing teacher effectiveness by 1 SD, increases student learning by 0.09 SD. ► Classes by most effective teachers implement structured lesson plans and involve more student interactions.
Muralidharan and Sheth (2016)	India	► Effectiveness of female teachers (as opposed to male teachers) in reducing gender gap in achievement for grade 1-5 students.	► Value-added framework using school-grade fixed effects and grade fixed effects by student gender.	► Girls have higher annual test score when taught by female teachers as opposed to male teachers. ► No negative effects on boys when taught by female teachers.
Aslam and Kingdon (2011)	Pakistan	► Teacher certification ► Teaching pedagogy	► OLS; ► School-fixed effects; ► Pupil-fixed effects.	► Teacher certification has no impact on test scores; ► Pedagogy (lesson plans & encouraging student participation) raises student learning. ► Girls benefit from being taught by female teachers relative to male teachers.

Study	Country	Characteristic/Component	Methodology	Findings
Kingdon and Teal (2010)	India	► Teacher has MA or PhD. ► Pre-service training. ► Teacher's cognitive ability proxied by 1^{st} division in own Higher Secondary Exam. ► Teacher union membership.	► OLS; ► Student-fixed effects; ► School-fixed effects.	► Teacher credentials (and 1^{st} division in own Higher Secondary Education exam) increase student learning only in non-union schools. ► Pre-service training increases student learning at both unionized and non-unionized schools. ► Union membership negatively affects test scores.
Kingdon (2006)	India	► Teacher has MA or higher degree. ► Pre-service training.	► Fixed effects at state, school and pupil level.	► MA along with pre-service teacher training increase student achievement.
Fehrler et al. (2009)	Sub-Saharan Africa	► Teacher's academic attainment. ► In-service and pre-service training.	► Multilevel model with school random effects and country-fixed effects; ► Generalized least squares; ►Maximum likelihood estimation only when sampling weights are available.	► When teacher education and training are good quality, both enhance scores marginally.
Duflo et al. (2011)	Kenya	► Increased teacher effort and altered pedagogy (due to within school tracking). ► Teacher effort proxied by presence in school and in class on a random school day.	► RCT; ► OLS; ► Linear Probability Model; ► Regression discontinuity design.	► Raises test scores for all students along the achievement distribution.

Study	Country	Characteristic/Component	Methodology	Findings
Duflo et al. (2015)	Kenya	► Extra Teacher Program (ETP) - contract teachers hired and assigned to only one class (as opposed to civil service teachers who teach one subject across classes); ► School-Based Management (SBM) program where parents are empowered to manage and monitor school resources. ► Sustained good performance of contract teachers lead to renewed employment and ultimately civil service positions. 24	► RCT; ► OLS; ► Linear probability model.	► Student assigned to contract teachers under the Basic ETP have higher test scores and lower attrition than students in control schools. ► No statistically significant difference in test scores and achievement between students at control schools and students assigned to civil service teachers under Basic ETP. ► When combined with community empowerment (SBM), contract teacher program adds more to overall test score of students; civil service teachers less likely to be slacking.
Muralidharan and Sundararaman (2013)	India	► Extra Contract Teacher Program. ► Greater teacher effort (higher attendance, engagement in teaching and other) amongst contract teachers versus civil service teachers.	► RCT; ► OLS;	► Contract teachers improved learning outcomes and are as effective as regular teachers.
Goyal and Pandey (2009a)	India	► Contract teacher intervention program. ► Greater teacher effort (higher attendance and teacher engagement in teaching) amongst contract teachers versus civil service teachers.	► OLS with block fixed effects. .	► Contract teachers associated with higher test scores through higher teacher effort/engagement.
Chin (2005)	India	► Provide second teacher at one-teacher primary schools.	► Difference in difference estimation.	► Higher primary school completion rate for girls and poorer students.

[24]as opposed to civil service teachers who teach one subject across classes

Study	Country	Characteristic/Component	Methodology	Findings
Banerjee et al. (2004)	India	► Additional teacher in one teacher school (with preference for female teachers). ► Meal program.	► Panel data analysis; ► Random effects at the school level.	► Additional teacher increased attendance for girls but not boys. ► Meal program raised attendance for boys and girls. ► Ambiguous impact on test score.
		ADMINISTRATIVE MONITORING		
Duflo et al. (2012)	India	► Link teacher's pay to days absent; ► Monitor teacher's presence through time-stamped photographs;	►Randomized experiment. ►Linear regression framework.	► Linking teacher's pay to presence at school halves teacher absence and increases student test scores by 0.17σ. ► Both the mid-test and post-test suggests that girls gain more from the program than boys (the latter being statistically insignificant).
Muralidharan and Sundararaman (2010)	India	► Low-stakes monitoring of classroom processes; ► Teachers are provided with feedback (including diagnostic test results of their students' performance). ► Teachers are provided with performance-linked bonuses in some cases.	► RCT; ► OLS with *mandal* level fixed effects.	► Feedback and monitoring impact student test scores only when teachers are given incentives.
Muralidharan et al. (2014)	India	► Increased inspection and monitoring versus hiring additional teachers.	► Panel OLS regression with fixed effects at state and district level.	► Increased monitoring ten times more cost effective in reducing effective student-teacher ratio than hiring additional teachers. ► Suggestive evidence that teacher absence and student learning outcomes are negatively correlated.
		REMEDIAL/SUPPORT CLASSES		

Study	Country	Characteristic/Component	Methodology	Findings
Banerjee et al. (2007)	India	► Young women from the community are hired to teach students lagging behind in basic literacy and numeracy skills (*Balsakhi* program) for two hours each day. ► Math-focused computer-assisted learning (CAL) program for two hours every week; targets all children.	► RCT; ► OLS; ► IV; instrument actual treatment status with intention to treat dummy.	► During the 2 year intervention period: Balsakhi program had strong impact on math and language test scores. CAL program had strong impact on math scores. ► Both programs had sustained impact one year after completion on the bottom third of the distribution. ► For CAL, however, average effect declines but remains significant even a year after program completion.
Lakshminarayana et al. (2013)	India	► Community volunteers hold two-hour supplementary remedial lessons after school hours. ► For children in classes 2-4.	►Cluster randomized trial. ► Compare mean differences using unpaired t-test. ►Linear regression models.	► Increases average math and language scores in treatment schools. ► Learning gains are larger for girls than for boys; these gains disappear when baseline test scores are included.
Banerjee et al. (2010)	India	► Local youth volunteers hold reading camps.	► RCT; ► OLS; ► IV; instrumented attending a reading class with dummy for whether child was assigned to the reading camp (i.e. treatment status).	► Large improvements for students at all levels of reading capability.

Study	Country	Characteristic/Component	Methodology	Findings
Banerjee et al. (2015)	India	► Core pedagogy - 'Teaching at the Right Level' (TaRL). ► A variety of interventions across four states including remedial classes, provision of material, teacher training and local volunteer support.	► RCT; ► OLS.	► When implemented effectively (as in Haryana and Uttar Pradesh), TaRL has a strongly positive and statistically significant impact on Math and Language scores. ► Teaching students at their learning level only successful when implemented outside of regular school hours. [25]
Cabezas et al. (2011)	Chile	► Volunteers (college students) provided small group tutoring for students from vulnerable schools and poorer backgrounds.	► RCT; ► OLS; ► Ordered logit; ► IV; instrument 'number of tutoring sessions attended' with intention to treat dummy.	► Increases the 4^{th} grade reading scores, benefiting low achievers the most. ► Improves cognitive and non-cognitive measures of reading especially among more vulnerable and low achieving students.

[25] This is because volunteers had the freedom to properly implement TaRL activities outside of school hours.

Study	Country	Characteristic/Component	Methodology	Findings
		CLASS SIZE		
Urquiola (2006)	Bolivia	► Two intervention designs to understand effects of class size on achievement.	► OLS; ► IV; use predictive class size as instrument for actual class size.	► Both designs suggest larger class sizes negatively affect achievement.
Urquiola and Verhoogen (2009)	Chile	► Variation in class size. ► Class size cap set at 45; at the 46th student, schools have to introduce a new section with average class size falling to 23 (and similar discontinuities at 90, 135, etc.)	► Fuzzy RDD with IV; use 4 class size cut-offs as instrument for actual class size.	► Increase in class size lowers math and language test scores. ► This effect disappears when socioeconomic controls are included, due to endogenous sorting around class-size cut-offs.
Duflo et al. (2015)	Kenya	► Reduction in class size from 82 to 44 students, ceteris paribas.	► RCT; ► OLS; ► Linear probability model.	► No change in student performance; ► Only has an impact when combined with 'contract teacher' intervention.
		SCHOOL ACCESS, PEDAGOGICAL MATERIAL & CURRICULUM		
Orkin (2013)	Ethiopia	► Increase length of school day.	► Difference-in-difference specification with school and grade-fixed effects.	► Has positive and statistically significant impact on numeracy and writing. ► No significant effect on reading. ► Girls gain more than boys in numeracy due to longer school days.
Bellei (2009)	Chile	► Increase length of school day.	► Difference in differences specification.	► Improves performance in Mathematics and Language for all students across the achievement distribution. ► Students in the 75th percentile and higher benefit the most.

Study	Country	Characteristic/Component	Methodology	Findings
Glewwe et al. (2009)	Kenya	► Provision of free textbooks to schools.	► RCT; ► Generalized Least Squares; ► Interact pretest scores with textbook schools using Random Effects.	► No impact on dropout or repetition rate for lower grades. ► Grade 8 students more likely to enter secondary school. ► No impact on average test scores. ► When distinguish between high ability and low ability students, program seen to affect above average student in terms of test scores. ► Lack of impact stems from textbooks being at an inappropriate level in Kenya.
Sabarwal et al. (2014)	Sierra Leon	► Textbook intervention program for grade 4 to 5 students in non-private schools (govt., govt.-assisted and community schools). ► Textbooks provided to primary schools based on student enrollment numbers.	► RCT; ► OLS; ► IV; use treatment assignment as instrument for actual school treatment.	► No impact on literacy scores, math scores and student enrollment. ► Limited impact on attendance. Grade 5 girls' attendance increases and is significant. ► Teachers are more likely to be found in class teaching and are more likely to have a lesson plan.
		PEER EFFECTS		
Duflo et al. (2011)	Kenya	► Peer-effect (in the context of with-in school tracking). ► High achievement students tracked into upper-track classes, while low achievement students are tracked into lower track classes.	► Reduced form; ► IV; [26]	► Tracking raises scores for all students along the ability distribution, even those assigned to classes for low achieving students; ► Tracking affects student achievement through both the direct channel of peer-to-peer effects and the indirect channel of increased teacher effort.

[26]This paper appears under two different categories: Teacher effort and Peer effects. Thus appears twice in this table.

B Tables: Means and Mean Differences

Table 9: Means by School-Type (Boys)

	BRAC	GOV	JAAGO	Other NGOs
Monthly Family Expdt	5.3479	6.2623	5.8478	5.4961
(in BDT 1000 adjusted by equivalence scale)	(0.2256)	(0.1460)	(0.1873)	(0.1596)
Father's schooling	2.7273	3.8987	4.0212	3.3333
	(0.3315)	(0.2598)	(0.2672)	(0.4036)
Mother's schooling	2.1712	3.2368	3.7327	2.9211
	(0.2489)	(0.2213)	(0.2504)	(0.2690)
K-BIT (IQ)	-0.5550	0.0439	0.3582	-0.1480
	(0.1125)	(0.0520)	(0.0792)	(0.1177)
Observations	110	278	260	171

Notes: (a) Standard errors in parentheses clustered at the street level; ((b) For the IQ score, we use age adjusted Z-scores. In other words for student i in age group a, we calculate, $Z_i = \dfrac{X_i - X_a}{\sigma_a}$, where X_a and σ_a is the mean and standard deviation in age group a.

Table 10: Means by School-Type (Girls)

	BRAC	GOV	JAAGO	Other NGOs
Monthly Family Expdt	5.0424	6.0575	5.8470	5.4639
(in BDT 1000 adjusted by equivalence scale)	(0.1519)	(0.1281)	(0.1277)	(0.1230)
Father's schooling	1.9535	3.5254	3.3787	3.2928
	(0.1800)	(0.2501)	(0.2050)	(0.2994)
Mother's schooling	1.7377	3.2624	3.8544	2.9630
	(0.1837)	(0.2877)	(0.1913)	(0.2428)
K-BIT (IQ)	-0.5190	0.0865	0.1876	-0.1529
	(0.0570)	(0.0936)	(0.1016)	(0.0868)
Observations	129	308	316	230

Notes: (a) Standard errors in parentheses clustered at the street level; (b) For the IQ score, we use age adjusted Z-scores. In other words for student i in age group a, we calculate, $Z_i = \dfrac{X_i - X_a}{\sigma_a}$, where X_a and σ_a is the mean and standard deviation in age group a.

Table 11: Mean Differences Across School-types (Boys)

	(1) BRAC vs GOV	(2) BRAC vs JAAGO	(3) BRAC vs Other NGOs
Monthly Family Expdt	-0.9144***	-0.4999	-0.1483
(in BDT 1000 adjusted by equivalence scale)	(0.2829)	(0.3160)	(0.2910)
Father's Schooling	-1.1714**	-1.2939**	-0.6061
	(0.4336)	(0.5207)	(0.4363)
Mother's Schooling	-1.0656***	-1.5615***	-0.7498**
	(0.3100)	(0.3409)	(0.2932)
K-BIT (IQ)	-0.5989***	-0.9132***	-0.4070***
	(0.1185)	(0.1628)	(0.1394)
Observations	388	270	281

Notes: (a) Standard errors in parentheses clustered at the street level; (b) We report the difference in means at the 1 percent, 5 percent and 10 percent significance level denoted by *** , ** , and * respectively; (c) For KBIT (IQ score), we report age adjusted Z-scores.

Table 12: Mean Differences Across School-types (Girls)

	(1) BRAC vs GOV	(2) BRAC vs JAAGO	(3) BRAC vs Other NGOs
Monthly Family Expdt	-1.0150***	-0.8045***	-0.4215**
(in BDT 1000 adjusted by equivalence scale)	(0.2082)	(0.1426)	(0.1976)
Father's Schooling	-1.5719***	-1.4252***	-1.3393***
	(0.3102)	(0.2984)	(0.2964)
Mother's Schooling	-1.5247***	-2.1167***	-1.2253***
	(0.3569)	(0.3095)	(0.3216)
K-BIT (IQ)	-0.6056***	-0.7066***	-0.3661***
	(0.1069)	(0.1259)	(0.0892)
Observations	437	445	359

Notes: (a) Standard errors in parentheses clustered at the street level; (b) We report the difference in means at the 1 percent, 5 percent and 10 percent significance level denoted by *** , ** , and * respectively; (c) For KBIT (IQ score), we report age adjusted Z-scores.