KDI SCHOOL

WORKING PAPER SERIES

KDI 국제정책대학원

KDI School of Public Policy and Management

Teachers' Monitoring and Schools' Performance: Evidence from Public Schools in Pakistan

Hee-Seung Yang
KDI School of Public Policy and Management

Booyuel Kim
KDI School of Public Policy and Management

Inayat Ullah
COMSATS University Islamabad

January, 2020
Working Paper 20-02

KDI 국제정책대학원
KDI School of Public Policy and Management

* We are grateful to the KDI School of Public Policy and Management for providing financial support.

Teachers' Monitoring and Schools' Performance: Evidence from Public Schools in Pakistan

Abstract

This paper evaluates the impact of an innovative monitoring system on teacher attendance and school performance in Pakistan. In 2014, the government in Khyber Pakhtunkhwa province introduced the Independent Monitoring Project aiming at increasing teacher attendance in primary and secondary public schools by distributing to the government-hired monitors smart phones with a special data collection software installed. Our analysis is based on a difference-in-differences approach using the country wide Annual Status of Education Report from 2012 to 2016. Our findings suggest that monitoring of government schools has increased teacher attendance by 7.5 percentage points in the first year of intervention. But the positive effect wears off to 2.7 percentage points in the second year. Child attendance and test scores also increased in the first year, but in the second year they disappeared. Especially, in the first year, the monitoring system improved students' math, reading, and English test scores by 0.13, 0.14, and 0.15 standard deviation, respectively, if they are grades 1-5. This result suggests that teacher attendance has an important role in delivering better student outcomes, but that monitoring should be coupled with appropriate incentive mechanism in order to have a lasting impact.

Key Words: Schools Monitoring, Teachers Attendance, Learning Achievement

1. INTRODUCTION

The recently developed “Sustainable Development Goals (SDGs)” emphasizes the need for more rigorous efforts through empirical findings that suggest feasible courses of actions to improve teaching quality and children learning achievement (UN SDGs, 2016). Despite some success in children enrollment, the overall quality of education especially at primary and secondary levels has remained the lowest in South Asia (e.g. India, Pakistan, Bangladesh etc.). Recently, a countrywide survey on educational attainment in India finds 44 percent of the children aged 7–12 cannot read a basic paragraph, and 50 percent cannot do simple subtraction despite increased school enrollment(Banerjee et al., 2007). According to Annual Status of Education Report (ASER-Pakistan) which reveals important trends each year covering over 255,000 children from 144 districts, Pakistan continues to be in a state of education emergency and learning lies at the heart of it. This is evident from its recent report showing that 52% children in grade 5 could read at story level dipping from 55% in 2015. Similarly, for English it was 46% (49% in 2015) and for arithmetic it was 48% compared to 50% in 2015 (ASER, 2016). In similar circumstances, as Banerjee (2007) suggests, policies that only increase school enrollment may not guarantee learning outcomes. Recent evidence also supports the idea that interventions that only focus on school participation might not improve test scores for the average student (Abdulkadiroğlu et al. , 2018; Attaullah & Malik, 2015; Burde & Linden, 2013; Duflo et al., 2007; Munene, 2015).

One important component of school environment is the presence of teachers that influence the overall performance of children (Banerjee & Duflo, 2009; Glewwe & Kremer, 2006). Teachers’ absence has been a widespread problem in developing countries particularly in far-flung rural areas. Recent studies in education research document evidence that increased absence rate of teachers is strongly related with school and children performance(Banerjee & Duflo, 2006; Banerji et al., 2013; Kremer et al., 2006; Duflo & Hanna, 2005). A number of factors can be found responsible for increased absenteeism such as distance from school, lack of appropriate incentives(Scott & Wimbush, 1991), ineffective monitoring(Duflo & Hanna, 2005) and other socio-economic factors(Alcázar et al., 2006). One of the important sources of differential teachers and schools performance is the type of monitoring and administrative oversight of schools and the resulting reward and penalty system. For example resources may be spent on hiring and payment to teachers who are absent from their schools such as the presence of *ghost* schools(Glewwe & Kremer, 2006). According to ASER (2015), teachers' presence was one of the big factors to account for differences across learning outcomes across public and private schools in Pakistan. Also, there has been increasing focus by practitioners and development researchers on the teaching quality and punctuality that has significant direct and indirect effects on children performance(Duflo, 2007; Munene, 2015). Literature on teacher’s performance indicates that teacher incentives and other interventions have larger impact in low performance settings (Murname and Ganimian, 2014). However, considering the high absenteeism in developing countries, incentives alone may not work unless coupled with effective supervision of teaching staff particularly in rural areas. For example, in Pakistan’s Punjab province, a public-private partnership program that offered bonus for teachers, had limited effect on children’s test score because such incentives were not effectively linked with students performance (Barrera-Osorio and Raju, 2010). Similarly, incentivizing the administrative staff such as headmasters in schools without effective monitoring mechanism may not improve teachers attendance and children learning (Kremer and Chen, 2001; CDPR, 2014). With regard to effectiveness of monitoring methods, previous studies suggest different ways of supervision such as strengthening administrative oversight and community-based supervision to ensure better teachers' attendance(CDPR, 2014; Muralidharan et al, 2014). Teachers failure to attend schools is mainly due to the lack of capacity of administration(e.g principle) and the beneficiary(children or local community) to monitor and penalize absence(Duflo & Hanna, 2005). Although, the headmasters have power to penalize absence by rules, nevertheless, by virtue of their close relationships with teachers (who generally belong to the local community), they are unable to enforce penalty or report absence to the higher authorities. Resultantly, the higher authorities in governments who are responsible for decision making, lack the real reporting of data from far-flung rural areas or get manipulated record about schools and teachers presence.

A number of reforms initiatives have been proposed for developing countries that can maximize the quality of learning of enrolled children, reduce dropout ratio and attract out-of-school children(Robert, 2005). The main focus of these studies remains both on the demand and supply side of education such as provision of educational facilities, widening access to education and increasing enrollment in schools etc.(Banerjee & Duflo, 2009; Jones, & Rajani, 2014; Raikes, 2016). With regard to teachers' availability in schools in developing countries, few studies have attempted to investigate the effectiveness of different policies that are targeted at schools or teachers' supervision. These include teachers' incentive programs such as providing incentives based on exam score of children, direct monitoring of teachers performance through camera coupled with high-powered incentives and community-controlled interventions etc.(Alcázar et al., 2006; Duflo & Hanna, 2005; Scott & Wimbush, 1991). The World Development Report suggests the expansion of community-based monitoring of schools that might strengthen the flow of information between community and school administration and effectively involving community in hiring, firing and payment or transfer of teachers (WDR, 2018). However, contextual evidence on community-based monitoring reflect less effectiveness of such programs particularly in rural areas (Banerjee and Duflo, 2005; Kremer & Vermeersch, 2005). One important factor in this cases is the awareness of local community or average education level that might influence the community response to teachers' unavailability. In other words, given the overall low education level in the community (more often in developing countries), it is less likely that local people will realize the consequences of teachers' absence and its effect on children learning. While much has been researched about significance of teacher's availability and school facilities, less is known about how to increase teacher's attendance especially in rural and remote areas in an effective and cost efficient way.

This paper takes advantage of data collected by the Annual Status of Education Report (ASER)-which is similar to *Pratham* in India and *Uwezo* in Africa-, to attempt a natural experiment on a recently introduced government-schools monitoring project by the KP government in Pakistan. We attempt to find a comparable administrative unit that has not been affected by the policy yet shares similar socio-economic and demographic characteristics across the border with the treated administrative unit.

The results discussed in this research suggests a number of practical and methodological insights. First, school performance in terms of teacher's attendance and school facilities can be increased by increasing monitoring of schools using professionally trained monitors and adaptation of latest technology. Second, evidence support the idea that improving schools performance affect parents and children behavior in terms of sending children to schools and attending schools respectively. Earlier studies based on natural experiments and randomized evaluations find mixed results on the effectiveness of monitoring vis-à-vis indirect incentives and rewards system in government policies on children's learning outcomes in developing countries. Third, given the weak public education system in developing countries, monitoring of schools and teachers should be coupled with appropriate incentive/punishment mechanism in order to have a lasting impact on children performance. Finally, we argue that there is scope for the use of nationally representative surveys in conducting natural experiments for assessing the impact of education programs carried out by sub-national governments in developing countries.

The following section gives a brief account of the education system in Pakistan, its short history and major problems that hinder the road to achieving quality education. The 3rd section provides a detailed description of the monitoring program and its implementation procedures. Section 4 outlines theoretical framework in the light of previous works. Experimental design and its key conditions are discussed in section 5. Section 6 describes the data, Section 7 details the empirical strategy followed by results and discussion in section 8. The last section concludes.

2. GAPS IN PAKISTAN'S EDUCATION SYSTEM

Being the sixth largest country in the World, Pakistan inhabits population of around 210 million of which 64% is below the age of 30 (UNDP, 2018). Despite significant decline in the fertility level in recent years, Pakistan's population is still growing at a rate of 2% per year, highest in South Asia (WB, 2018). According to Burki (2005), those less than 18 years old will account for about 50% of total population in 2030. This represents a big challenge as a significant proportion of young people will be poorly educated and inadequately skilled in case the successive governments fail to launch and implement ambitious education reforms.

To understand the structure of education system in Pakistan, it is important to dig into its history that started in the late 1940s. For the first 25 years (1947 to 1970), Pakistan's education system was relatively efficient, not much different from its neighboring India. Dominated by public sector, education departments in provinces were responsible for administering primary and secondary schools and colleges with a public sector teachers training schools and colleges. For several decade, the number of private schools was not much within the system of education. However, after the denationalization in 1990s, the private schooling become another major source of education at the lower level particularly for the elite class of society. Currently, the large public education system starts with primary schools at the lower level (0 to 5 grades), then secondary and high schools, and autonomous public funded universities at the higher level. Over the years, the amount of budget spent on public education has been one the lowest compared to other countries for various reasons. According to the World Bank's latest estimates, Pakistan spends nearly 4.9% of its GDP on education with about 30% spending on primary education (WB, 2016). According to Pakistan's Economic Survey, the overall literacy rate is 58% with male 70% and female 48% (MOF, 2017). In other words, nearly one-half of the women cannot read or write while this gap is much higher in rural areas. Solutions proposed for reforming the public education include incentives for parents and children, increasing the proportion of public resources going into education sector, diversion of more funds towards primary schooling and investment in teachers' training and improving the quality of schools and curriculum(Robert, 2005).

Pakistan continues to suffer from slower growth in key socio-economic indicators reflected by the human development report as compared to its neighboring countries such as India and Bangladesh (UNDP, 2016). Low education quality, both at primary and secondary level is at the centre of many problems that the country face in almost all regions. According to a study by International growth Centre (ICG), in Khyber Pakhtunkhwa(KP) province (the focus of this paper) in 2012-13, only 63% of 4-9 years old children were enrolled in schools with a much lower (56%) female enrollment(CDPR, 2014). For higher grade, the net enrollment is even worst. For example, for middle schools, the net enrollment was hardly 40% reflecting a significant dropout or no-enrollment during the middle school age group (11 to 15 years). Similarly, teacher's absenteeism rate was 16% for primary, 21% for middle, and 17% for high schools indicating unavailability of teaching service at a critical school age. With regard to learning achievements, the entire country including KP province faces alarmingly low level. Out of surveyed enrolled children, only 40% of grade-5 children could answer the second-grade level mathematics and language questions. From the supply side of education, the KP province employs nearly 55% of the civil servants in education department with a significant number of teachers. For example, teachers make up at around 75% of the 180,000 employees overall in elementary and secondary education department. To what extent this chunk of employment has been effective is the policy question that motivates this study.

Recently, as part of the constitutional amendments, Pakistan has devolved most of administrative and fiscal decision making to the provinces. In this devolved setting, provinces are autonomous in reforming their education sectors to improve the dismal conditions of schools and teachers quality and children learning. The establishment of an Independent Monitoring Unit (IMU) is one such initiative taken by the provincial government of Khyber Pakhtunkhwa (KP) province that aims at monitoring teachers and schools performance through trained monitors equipped with smart-phone aided facility(section 3 provide more details on IMU). According to ICG's analysis on the IMU school level data in 2014, there was significant variation in teacher's attendance and student attendance rates at the primary and secondary level. Also, large variation in school size measured as enrollment of children and teachers-students ratio were identified. Exploiting this variation, the same study by applying a statistical model, finds significantly positive effect of teachers attendance and school infrastructure on the children enrollment rates. With the exception of seven districts-in *hard areas*[1]- where

[1] Currently, seven districts e.g., Kohistan, Battagram, Tor Ghar, Dir Lower, Dir Upper, Shangla and Tank have been identified as "hard areas" for girls' schools(CDPR, 2014)

additional incentives are offered, the KP government has a uniform incentive structure for teachers similar to other provinces of Pakistan. Moreover, to improve girl's education, the KP government gives additional allowances for female education supervisors to increase their inspections to schools. Similarly, to attract girls enrollment, the KP government offers stipend program for secondary students for selected districts[2] with low enrollment. Also, in two districts, special scholarships are offered for girls for their enrollment in schools (e.g Kohistan and Torghar). A detailed review of the KP government civil service rules carried out by ICG's research shows the presence of a number of direct and indirect incentives for improvement in teacher's attendance and students learning(CDPR, 2014). However, these incentives were not properly linked with government objectives of improving education outcomes. The review further finds that promotion and up-gradation procedures, performance evaluation and transfer policies were not realistically linked with teacher's attendance measurements or student performance in exams, suggesting the need for a more objective criteria for measuring teacher's performance.

3. PROGRAM DESCRIPTION

In struggle for quality improvement in education sector, in 2014, the Khyber Pakhtunkhwa (KP) provincial government took an important initiative of establishing a landmark project, Independent Monitoring Unit (IMU) for monitoring teachers and schools performance through trained monitors equipped with smart-phone aided facility. The project was aimed at monitoring and data collection for over 28000 public sector primary and secondary schools in the province. The basic objective of the IMU was to ensure presence of teachers through effective monitoring besides collection and compilation of data on basic schools facilities such as electricity, boundary wall, toilets, and furniture etc. The specific objectives of the project included, collection of data on the presence of teachers in school, number of children enrolled, schools facilities, availability of school administration and other school related information.

Lunched formally in April 2014, the IMU's mandate was to monitor over 28000 schools with over 121,618 government appointed teachers across the province. The implementation of IMU project needed quite laborious work as the KP province is geographically characterized with rugged terrain and dispersed population in rural areas. Also, over the last 18 years, the education sectors in KP province and it's neighboring federally administered tribal areas, have been a direct target of terrorism resulting into destruction of hundreds of schools particularly girls school and killing of several teachers including female teachers. The IMU program conducts monitoring using both human efforts and technology for keeping external control while dealing with shirking teachers and school administration.

The IMU hired 550 Data Collection and Monitoring Assistants (DCMAs or monitors) and subsequently appointed them in every district of KP province. Their job is to visit randomly to government schools located within the assigned administrative clusters[3] (at least one time per month to each school). The assignment of clusters rotate clock-wise on monthly basis to

[2] These low enrollment districts of KP include Hangu, Peshawar, Bannu, Lakki, D.I Khan, Shangla and Nowshera
[3] Generally, a district is divided into 10 to 30 clusters(depending on the population of schools and gender)

minimize the possibility of relationship-bias. For example, the monitor who inspected cluster-A in January, will inspect cluster-B in February and so on. Each DCMAs are required to visit at least 3 to 4 schools every day in schooling-hour to collect data. They are not allowed to share any prior information with schools or teachers about their scheduled visits. Upon inspection of the school, DCMAs are required to send attendance status of teachers (confirmed with their thumb-impression) to the central office through GPRS system installed in their smart-phone. The performance of DCMAs is in turn supervised by the District Monitoring Officers (DMOs) appointed one for each district across the province (H. Altaf[4], interview, October 2018). The IMU operation is based on IT application by trained monitors following a structured protocol provided by the provincial independent monitoring authority. The DCMAs collect data by physically verifying various school-based indicators after visiting the school in his/her designated area. The DCMAs then upload the information directly to the database of IMU using a prescribed questionnaire designed by the Elementary and Secondary Education Department (E& SED) of the KP province. The DCMAs use a special android application for conducting various checks and filter techniques to ensure provision of accurate data. The data sent by DCMAs to the database is further analyzed by IMU's IT team using various statistical tools to help make incentive (reward and punishment) decisions and take other necessary actions. So far, according to IMU officials, prizes worth 220 million Rupees have been distributed under the Teachers Incentive Program (TIP) among teachers that have higher attendance record. The IMU data was utilized in deciding on TIP criteria. However, with regard to penalty of low performing teachers, there is no such record of punishment or any decision whatsoever.

[4] A personal Interview was conducted online with Mr. Ataf Hussain, IMU official at District Shangla of KP Province to obtain information about the organizational structure and job description of IMU monitors and their appointment methodology.

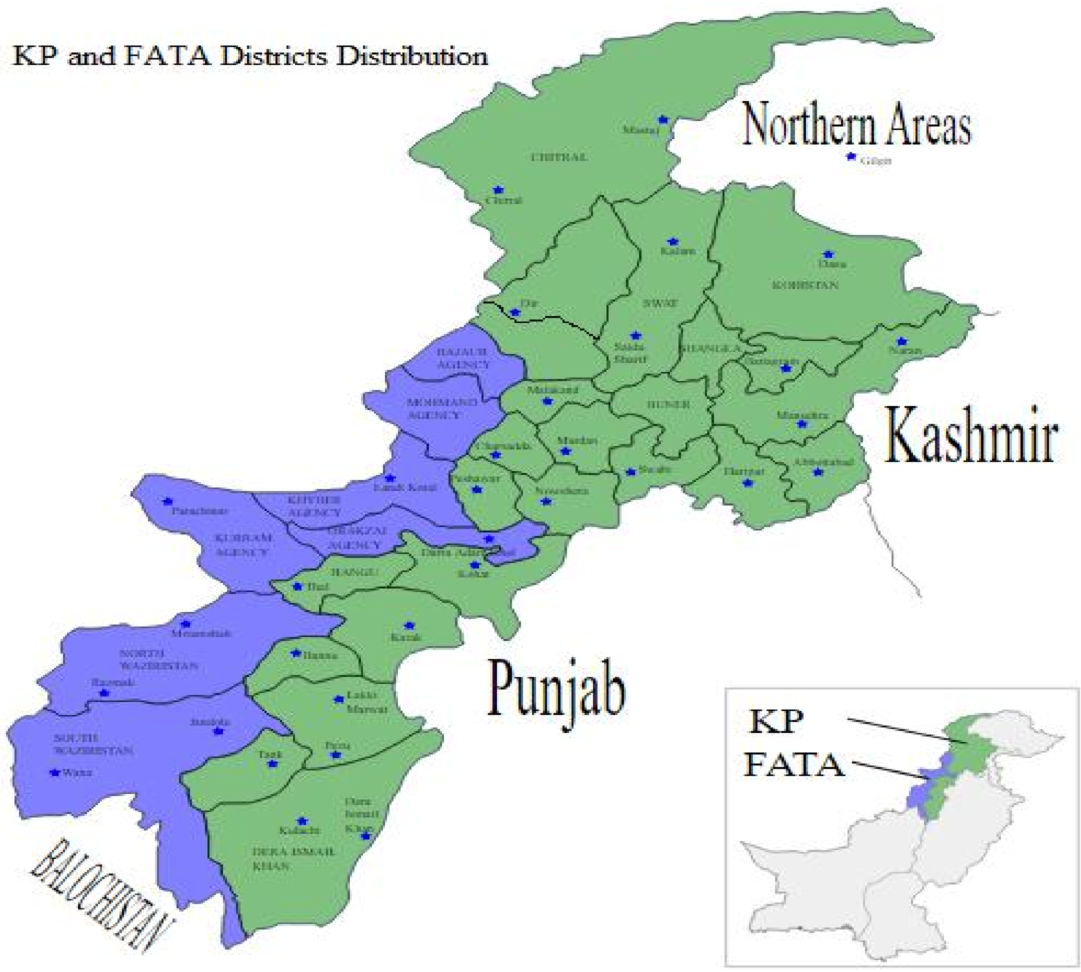

Figure 1: District wise Map of KP and FATA (Edited from commons.wikimedia.org)

So far, the government reports suggests that teachers attendance and punctuality have improved significantly ever-since the launch of the IMU, however, there is no empirical evidence about the impact of the extent to which the IMU has increased teachers attendance and the students' academic performance. This proposed research therefore will be a significant contribution towards genuine evaluation of this project.

4. EXPERIMENTAL DESIGN

This study aims at estimating the effect of the IMU program introduced by the KP provincial government in Pakistan on the school quality measured in the form of teacher's attendance on one hand, and on children learning outcomes measured through ability tests in three subjects e.g reading, math and English, on the other hand. To do so, it is important to use a model that truly identifies the causal effect of the IMU program. Literature on impact evaluation methodologies suggests several tools to estimate the impact of a policy intervention in education sector on student's achievement and school quality(Abdulkadiroğlu et al., 2018; Alcott & Rose, 2015; Attaullah & Malik, 2015; Burde & Linden, 2013; Card & Krueger, 2000; Croke, 2014; Duflo, 2007; Duflo et al., 2007; Munene, 2015). The focus of these studies is to know the likely impacts of various policy interventions on students' academic achievements.

Recently, randomized control trials (RCTs) have been considered the most effective design to find causal effect particularly in poor developing countries. For example incentive program linked with teachers presence measured through camera photograph with children in randomly selected schools in India by Duflo & Hanna (2005) finds reduced teachers attendance significantly and improved test score. Similarly, in a randomized trial in Nicaragua, radio instructions had significant impacts on pupils' math score (Jamison et al., 1981). In Kenya, randomized experiment of provision of school meals were was found to have positive impact on test score as long as teachers were well trained (Vermeersch & Kremer; 2004). In a remedial education program in urban India that focused on improving learning environment in government schools, increased test scores was observed at a reasonably low cost (Banerjee et al., 2004). Also in India, a computer-assisted learning program suggest potential positive impacts on students' learning achievement (Banerjee et al., 2004). However, besides other challenges such as implementation etc., one of the big limitations associated with such experiments is their high cost of implementation.

The second most credible design in recent impact evaluation literature is natural experiment. In the absence of random assignment of subjects, one can exploit variation caused by any policy change that is exogenous in nature. In such cases, the simplest way of calculating the causal effect is using "difference-in-difference" (DiD) method, by comparing pre-program difference with the post-program difference between treated and untreated groups. Evidence from recent natural experiments in low and middle income countries suggests a positive impact of increasing school quality on students' academic performance, despite extensive variation in different contexts. These experiments include(but are not limited to) impact evaluation of primary school environments on secondary school outcomes using data on Ethiopian Jews by Gould, Lavy & Paserman (2004) and impact of class size on student academic performance in Israel using Maimonides' Rule by Angrist & Lavy(1999) etc. Results of natural experiments vary by context and by subjects owing to a number of reasons. For example, a natural experiment using Israeli data shows reducing class size raises reading score but not math score, while providing computers has no effect on academic performance (Angrist & Lavy 2002). One big challenge of such experimental designs is the availability of control (untreated) group that satisfies all conditions for an ideal comparison. For example, in the context, of school' monitoring program, one needs to have schools that are not directly or indirectly affected by the policy targeted for specific treated schools. Another challenge is to find schools that share similar characteristics with the treated schools before the intervention. In cases where the outcome variables between the treated and untreated subjects vary before the interventions, studies attempt to mitigate this challenge by adopting the common trend assumption conditional with availability of data. Recently, the two stage least square (2SLS) or instrumental variables (IV) is adopted as an alternative approach to estimating the impact of education policy interventions. According to this approach, a variable is used as an instrument which may or may not arise from natural experiment, but is correlated with the endogenous variable and uncorrelated with the unobserved factors that might affect the outcome variable (e.g child's learning). In IV estimation, the common variation between the instrument and the endogenous variable is exploited in determining the estimate of the effect of certain variable of interest (Angrist & Pischke, 2009; Wooldridge, 2013). Despite its convincing power in explaining education production function, finding a good instrument is often a challenge.

While natural experiments (and randomized trails) are meant to create a pool of such results that are less likely to suffer from estimation problems, development economists stress the need for a much larger set of results on a more representative sample of population before reaching a general conclusion. Nevertheless, in many developing countries, natural experiments and randomized control tails are considered the most effective means for improving school quality through addressing the problems associated with weak teachers or teachers' behavior(Glewwe & Kremer, 2006). Understanding the impact of policies that affect teachers' behaviors is critical particularly in the context of developing countries that suffer from higher absenteeism. Considering the exogenous nature of IMU program introduction in KP province, Pakistan, we attempt to exploit an annually representative survey data produced by the Annual Status of Education Report (ASER) to conduct a natural experiment. Note that the purpose of collection of ASER data is unrelated with the IMU program in all aspects whatsoever. We attempt to find a comparable administrative unit that has not been affected by the policy yet shares similar socio-economic and demographic characteristics across the border with the treated administrative unit. We test this by conducting a pre-program trend analysis on all variables used in our estimations. We also supplement our results by adopting variation in our endogenous variable (teachers' attendance caused by increased monitoring) as an instrument to estimate the effect on children test performance. In doing so, one face the difficulty of factors that affect teachers' quality such as punctuality, might also affect child' test performance at home. However, we take advantage of the clear exogenous nature of the monitoring project. In other words, controlling for the effect of monitoring on teachers attendance, the IMU is less likely to affect children learning achievements through any other channel.

5. DATA

Out main data source is the 5 years country wide Annual Status of Education Reports (ASER), Pakistan survey, from 2012 to 2016. The ASER[5] is frequently cited in reference to teachers attendance, children enrollment and attendance, learning ability, private school enrollment, and other key education indicators by renowned researchers (Jones et al., 2014; Banerji et al., 2013; Zaka & Maheen, 2010; French, Kingdon, & others, 2010). ASER is the large scale citizen-led, household based initiative managed by *Idara-e-Taleem-Aagahi* (ITA)-Pakistan in partnership with a number of governmental and non-governmental organizations, to provide reliable data on the status of primary and secondary education in all rural and few urban districts of Pakistan. Each year, ASER conducts a comprehensive assessment of the state of learning, school performance, and other indicators of primary and secondary education throughout rural Pakistan. Mobilizing more than 10,000 volunteers each year, the survey covers 600 household in each of Pakistan's 136 districts yielding a large national dataset of 81600 households and around 286,000 children per year. Table 1 provides year wise coverage of ASER data for KP province and FATA (the target of our study). The ASER household survey include learning

[5] ASER survey is similar to Pratham in India and the Uwezo surveys in Africa.

tests performed by children at home while a separate survey of the government and private schools is conducted in the sample villages.

The ASER sampling framework is systematic and well designed. For example, each district is provided with a village list with population information given by the National Bureau of Statistics (NBS). In view of the variability of the key variables, population distribution and field resources, ASER selects a sample of 600 households from each district. Each district is further divided into 30 villages whereas 20 household are selected from each village. The ASER adopts tow stage sampling design. In the first stage 30 villages are selected using probability proportional to size (PPS) method. In the second stage, 20 household are selected[6] from each of the 30 selected villages. Village is considered as the primary sampling unit, while household is treated as secondary sampling unit. Every year, the ASER retains 20 villages from previous year, 10 new villages are added and 10 villages are dropped from the previous year. In this way the ASER survey give us a "rotating panel" of villages for better estimates. With regard to schools selection, ASER choose at least one government school which is mandatory (could be more than one) and one private school form each selected village. The later ASER surveys also include urban regions in Pakistan (ASER, 2016, 2015, 2013).

Table 1: ASER Survey Coverage (2012 to 2016) for KP and FATA

Survey Coverage	**2012**		**2013**		**2014**		**2015**		**2016**	
	KP	**FATA**	**KP**	**FATA**	**KP**	**FATA**	**KP**	**FATA**	**KP**	**FATA**
No. of Districts	23	9	25	9	27	9	26	11	24	9
No. of Villages	688	270	763	265	789	270	769	330	704	270
No. of Households	13,702	5,375	15,144	5,271	15,663	5,369	15,032	6,544	13,807	5,390
No. of Children	41,003	18,529	46,877	18,722	49,473	18,743	46,045	22,890	41,804	17,753

Notes: The number of districts covered each year in KP and FATA are not equal because of two reasons. First, coverage in districts which were affected by military operation against terrorist such as Mohmand Agency was skipped in 2012. Secondly, districts where the ASER team couldn't reach due to other administrative difficulties such as district Kohistan were also skipped. However, the number of missing district each year ranges between 1 and 4.

The primary strength of ASER dataset is its enormous sample size of children aged 5 to 16 years, households, government schools and private school related information across all districts in rural Pakistan that provides a clear picture of the state of schooling across the country. Secondly, the ASER learning tests which are well organized and carefully designed and conducted at home provide an opportunity to analyze children's ability without any potential school bias. Testing at school often carries a potential bias when teachers push more competent students forward during the survey. This feature of ASER testing allows us to be more confident about the validity and findings on learning tests. Moreover, ASER household survey collects data on all potential child-related and household related socio-economic variables that might affect learning ability such as age, gender, enrollment status, school

6 ASER divides each selected village into four parts: Surveyors are required to start from the central location and pick every 5th household in a circular fashion till 5 households are selected from each part (ASER, 2016).

status(government or private), current grade, tuition facility, house-condition and ownership and parents' education etc. Table 2 (a) and (b) shows the summary statistics of the 5 years ASER surveys annual data pooled form 2012 to 2016. The third important feature of ASER survey is its systematic coding of districts, villages, households, and children identification (IDs) that allows us to apply fixed effect models to control for any group-specific unobserved characteristics. Finally, the ASER provides sufficient baseline datasets on government and private schools information that enables us to conduct pre-treatment and falsification test on all relevant factors affecting school based and children related outcome.

Table-2(a) Government school summary (2012-16) pooled

	Government Schools		*Private Schools*	
Variables	KP	FATA	KP	FATA
Primary School(1 to 5)	0.655	0.789	0.272	0.208
Middle School Type A(1 to 8)	0.048	0.093	0.286	0.283
Middle School Type B(6 to 8)	0.095	0.003	-	-
High School Type A(1 to 10)	0.089	0.107	0.397	0.487
High School Type B(6-to-10)	0.157	0.005	-	-
All other school types	0.006	0.004	0.042	0.021
Average Enrollment of Children	230.755	155.404	293.698	386.779
Average Children Attendance	153.279	131.903	261.715	342.863
Average No. of Teachers Appointed	7.724	5.019	12.885	11.696
Average No. of Teachers Present	6.687	4.477	11.145	10.788
Student teacher ratio	38.468	39.145	25.434	33.56
Teachers-Attendance Ratio	0.875	0.897	0.919	0.906
Children Attendance Ratio	0.844	0.826	0.867	0.889
Laboratory Available(yes=1)	0.208	0.086	0.405	0.346
Compute Lab Available(Yes=1)	0.065	0.035	0.263	0.096
Internet Availability	0.03	0.007	0.19	0.05
N (No. of Schools surveyed)	3618	1386	1718	240

Notes: Table 2(a) use data from ASER government and private school surveys (pooled from 2012 to 2016). Values on school types and facilities represent the mean percentage of the surveyed schools. Student-teachers ratio, teacher's attendance ratio and children attendance ratio represents average ratio on corresponding variables. E.g. Teachers Attendance Ratio is calculated as no. of teachers present/total appointed teachers. Similarly, Children-Attendance Ratio is calculated as no. of children present/total enrollment in the surveyed school. KP stands for Khyber Pakhtunkhwa Province representing the treatment group while and FATA represents the control group called Federally Administered Tribal Areas. Middle schools type B and Higher schools type B do not apply for private schools.

Table-2 (b) Children Related Summary-2012-16(Pooled)

Variables	KP	FATA
Demographic Characteristics		
Child Age	9.038	8.438
Gender(Female=1)	0.397	0.37
Child Enrollment Status		
Child Enrollment Status(Yes=1)	0.755	0.675
Child Dropped Out(Yes=1)	0.034	0.033
Child School Type		
Child Enrolled in Government School(Yes=1)	0.518	0.481
Child Enrolled in Private School(Yes=1)	0.218	0.168
Child Enrolled in Other Schools(Yes=1)	0.014	0.024
Household Socio-Economic Conditions		
Private Tutoring(Yes=1)	0.072	0.05
House Ownership(Yes=1)	0.896	0.917
House Construction Weak(Yes=1)	0.348	0.544
House Construction Semi-Strong(Yes=1)	0.329	0.297
House Construction Strong(Yes=1)	0.323	0.158
Electricity Connection Available(Yes=1)	0.892	0.882
Mobile service Available(Yes=1)	0.841	0.687
TV Available(Yes=1)	0.512	0.406
Parents Information		
Father Age	41.004	39.38
Father Ever Attended the School	0.585	0.51
Father Years of Education	5.847	4.57
Mother Age	35.635	35.252
Mother Ever Attended the School	0.274	0.117
Mother Years of Education	2.202	0.77
N (No. of Children surveyed aged 3-16 years)	225202	96637

Notes: Table 2(b) use data from ASER- household survey (pooled from 2012-to-2016). All values represent the average percentages of the surveyed children. KP stands for Khyber Pakhtunkhwa Province representing the treatment group while and FATA represents the control group called Federally Administered Tribal Areas.

6. EMPIRICAL STRATEGY

The unique setting of the study area, the launching of monitoring program and ASER survey give us an opportunity to conduct a form of natural experiment. It is known that the monitoring project, IMU, was launched in the middle of April, 2014 across all districts of KP province. In Pakistan, two months summer vacations are observed every year from mid-June to mid-August. During the vacations, teachers are not required to attend the schools. The ASER collects data in September each year. In this way, considering the starting date of the program and summer vacations, it is less likely that the ASER data collected in September, 2014 has captured the program impact after September. During the first two months at the outset of the program (from mid-April to mid-June), a large scale program is less likely to be fully operationalized. The figure 2 shows the timeline and ASER data collection from 2012 to 2016. Therefore we do not have reason to consider year 2014 as a post-program period and expect the effect to take place in 2015. Given this context, our treatment period will consist of two years (2015 till 2016) in the selected districts. By the same token, considering 2014 as pre-program period is also likely to be bias our estimate, given the launch of the program in April, 2014. Although, we present results of 2014 as pre-program in for checking any possible difference (see annexures), we rely on 2012 to 2013 as pre-program in our main results.

Figure-2: Timeline of ASER data-collection and Launching of IMU

6.1. The Model

Our main outcome variables in the first stage is whether the intervention program has increased the teacher's attendance in the government schools in the KP province.

We hold the following assumptions to carry out diff-in-diff analysis in the given settings:

- The primary, and secondary education system in FATA is same as the KP due to the Exam Systems conducted by designated Education Boards[7].
- There is no significant difference in teacher's attendance and children performance between KP and FATA before the IMU introduction.
- FATA and KP share similar characteristics in terms of social, economic, geographic, and cultural conditions and population density etc.
- Our treatment period consists of two years (2015 till 2016) in the KP while the Pre-Treatment period consists of three years from 2012 to 2014.

We estimate the effect of monitoring program on school outcomes using the following equation:

$$Y_{idt} = \beta_0 + \beta_1 Monitoring_{idt} + \beta_2 X_{idt} + \propto_d + T_t + \varepsilon_{idt}, \qquad (1)$$

where Y_{idt} represents outcome on surveyed government school i in district d in time t; $Monitoring_{idt}$ is an interaction of treatment districts and post-year t (e.g. $Monitoring_{idt}$=1 if school i belongs to district d of KP province & t = 2015 or 2016 and $Monitoring_{idt}$=0 otherwise; X_{idt} *is* Vector of School level controls; $\propto_d$ is the district fixed effect; T_t is year fixed effect; and ε_{idt} is an error term clustered at village (=school) level.

[7] Education boards are constitutional bodies responsible for implementing school curriculum, conducting and supervising annual examinations and declaring results of government and private schools under the jurisdiction. All boards are located in KP province but consists of districts under its jurisdiction both in KP and FATA. In total, there are 8 Education Boards in KP province.

In a similar fashion the children test performance is estimated by the following equation:

$$Y_{igdt} = \beta_0 + \beta_1 Monitoring_{idt} + \beta_2 X_{igdt} + \propto_d + T_t + G_g + \varepsilon_{igdt}, \qquad (2)$$

where Y_{igdt} represents normalized test score of surveyed child i in district d in grade g at time t; $Monitoring_{igdt}$ is an interaction of treatment district d and post-year t; X_{igdt} is Vector of individual child related controls; $\propto_d$ is the district fixed effect; T_t is year fixed effect; G is individual grades' fixed effect; and ε_{igdt} is an error term clustered at village level.

6.2. Pre-Program Trend in KP and FATA

We take the advantage of the pre-program data to test the common trend assumption - the outcome in treatment and control group would follow the same trend in the absence of the treatment. The results suggest that teacher's attendance on average did not vary significantly between treatment and control before the policy was introduced. The same is true for children test performance. Table 3 (a) & (b) present the pre-program trends between KP and FATA on our main outcome variables, teacher's attendance and children standardized test scores respectively. The coefficient for interaction term(*pre-program diff*) shows that after controlling for observed factors such as school existing teaching quality, training quality, school age and size, the difference between KP and FATA in terms of teachers attendance ratio is not statistically significant in 2013 as well as in 2014. A similar common trend was observed between KP and FATA on normalized test score of children as shown in table 3 (b). We observe that, on average, coefficient of the interaction term of the normalized score for Reading, Math and English in lower grades (0 to 5) is not statistically significant indicating similar performance of KP children with FATA children in terms of these subjects. This is in line with previous studies that indicated lower performance of both KP province and FATA compared to the country-average in terms of basic reading ability at lower grades. With regard to education sector reforms, a close analysis of the recent government decisions in KP and FATA shows that during these five years period, there was no significant policy intervention other than education reforms that mainly focused on teachers attendance, school infrastructure and oversight(CDPR, 2014; Zaka & Maheen, 2010). In conducting pre-program analysis of children test performance, we control for all possible observed child-specific characteristics such as age, gender, parents education, household size and dummies for house ownership and facilities(see table 2(b) for description of control variables). We also conduct a pre-treatment analysis on upper grade children and including 2014 as pre-program (see Appendix III and IV for results). Overall, the trend is similar in all subjects except lower performance in normalized English score of children belonging to treatment province in upper grades.

Table 3 (a): Pre-Program Trend, Teachers Attendance Ratio[Equation(7)]

Dep. Var: Teachers Attendance Ratio	Post=2013	Post=2014
Pre-Program Diff (Treatment*Post)	0.0264	-0.0201
	(0.0230)	(0.0173)
School Teaching Quality	0.0359	0.0327**
	(0.0225)	(0.0166)
School Training Quality	0.00330	0.0223
	(0.0268)	(0.0185)
Urban	0.160***	0.0953*
	(0.0587)	(0.0554)
Old schools	0.00565	-0.00121
	(0.0138)	(0.0103)
School Size	0.0861	0.128***
	(0.0579)	(0.0423)
School Facilities	YES	YES
District FE	YES	YES
Year FE	YES	YES
Constant	0.706***	0.745***
	(0.0510)	(0.0426)
Observations	1,933	2,967
R-squared	0.074	0.060

Notes: Table-1 reports Pre-Program difference between KP province (treatment) and FATA (control) on teacher's attendance. Column (1) represent Post=2013 vs Pre=2012 while column (2) represent Post=2014 vs Pre=2012-13. The outcome variable is the ratio of teachers present in school to the total appointed teachers. Variable Pre-Program Diff is a typical diff-in-diff interaction of *to-be-treated province* (KP) and Post (year =2013 in column (1) and year=2014 in column (2)). Due to District and year fixed effect applied in each regression, we do not include variables for treatment and posts. Variables *School Teaching Quality* and *School Training Quality* are continuous variables showing the ratio of teachers with master's degree and specific training level to the total appointed teachers in each school. *School Facilities controls* include availability of water, boundary, toilet, library, playground, laboratory, computer and internet. *School Size* is a continuous variable representing the ratio of children enrolled in surveyed school to the school with highest number of enrolled children. The data is taken from the ASER-Pakistan School Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is the surveyed government school. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

Table 3 (b): Pre-Program Difference, Normalized Test Score [Equation (8)]

	Normalized Test Score		
	Lower Grades-(0-to -5)		
	Reading	Math	English
Pre-Program Difference (KP*Post)	-0.0354	-0.0442	-0.0435
	(0.148)	(0.148)	(0.155)
Child -Related Controls	Yes	Yes	Yes
Parents Education Controls	Yes	Yes	Yes
Household Characteristics Dummies	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes
Constant	-0.662***	-0.562***	-0.452***
	(0.119)	(0.125)	(0.130)
Observations	19,757	19,659	19,608
R-squared	0.023	0.022	0.019

Notes: Table 3(b) reports the pre-program difference using diff-in-diff estimates on the *children test performance* for Post=2013 vs Pre=2012 using the ASER Household Survey data. Standard errors clustered at village level are shown in parentheses. The unit of observation is surveyed 3 to 16 year's old child enrolled in government school from Grade-0 to grade-5. The dependent variable is the *test score normalized by grade*. The pre-program difference is a typical diff-in-diff interaction of *to-be-treated province* (KP) and Post (which is equal to 1 if year==2013 and 0 if year=2012 or 2013). Fixed Effect on individual grade, District and year applied in each regression. Child-related controls include age, private tuition; parent's education controls include, mother and father highest education in years; household characteristics include ownership, house condition, and availability of electricity, mobile and television facilities. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

7. RESULTS

7.1. Program Impact on Government Schools Outcomes

Table 4 (a) reports the main results of the monitoring program on the ratio of present teachers to total appointed teachers using basic OLS model in equation (7). We check the program effect using different post and pre-program-years to see any difference during post-program two years. Since most of the KP province and FATA contains rural areas, time-invariant district-specific factors such as school density (schools per km^2) and district administration offices etc., might affect the outcome variable(see appendix-V for list of districts in KP and FATA). To overcome any time-invariant district-specific unobserved characteristics and time trend, we use district fixed effect and year fixed effect respectively throughout our regressions. Also considering the potential variation in teacher's behaviors, we control for schools teaching and training quality, urban districts, history, size and a vector of school-related facilities. School teaching and training quality is measured as a ratio of teachers with master's degree and professional training certificate to the total appointed teachers in the surveyed school. We represent schools' history as a dummy of old schools with more than 50 years of establishment equals to one. As suggested by previous studies, enrollment of children in schools might affect teachers attendance behavior (Koedel & Betts, 2007), we therefore control for school-size represented by enrollment. The role of school infrastructure in creating better teaching environment is well documented in education literature (Abhijit Banerjee & Duflo, 2006; Robert, 2005). We control for all school-related facilities surveyed by ASER (e.g. availability of water, boundary wall, toilet, library, playground, laboratory, computer and internet).

Table 4(a) column (1) shows a significantly positive effect of the program on teachers' attendance ratio in the year immediately following the program (e.g in 2015). Controlling for observable covariates such as existing school teaching and training quality, location, history, school size, and a vector of school facilities, the coefficient of the interaction term shows an increase of .067 percentage points in teachers' attendance ratio in the KP province as compared to FATA. In other words, being exposed to the monitoring program, on average, teacher's attendance in government schools is likely to increase by nearly 8 % in the first year of program implementation. This effect is larger given the mean value of the dependent variable (.881). Table 4(a) does not include 2014 data, considering it a transition period. (See Appendix VI) for results on 2014 as pre-treatment period). Column (2) adds year 2016 as post-program period into our analysis. It can be observed that the program effect is not significant and has been decreased by nearly half after two years of program implementation. The effect is however statistically significant at 5% when we include year 2014 in our analysis. Appendix Table 4(a) reports results after including year 2014 as post-program period.

There could be several reason for decreasing effect of the program. First, the expected penalty (or reward) as a result of IMU was not strictly observed despite certain absenteeism reporting by IMU. Secondly, as other studies observe, there could be a *learning effect*(Banerjee & Duflo, 2006), from the perspective of teachers such as, teachers might have learnt sources of shirking by establishing contacts with people who might observe visiting monitors on their way to schools. This can happen more likely in far-flung rural areas, where distance between schools and monitors' place of residence is large. In their paper on addressing absence in India using a

Table -4 (a) : Teachers Attendance Ratio[Equation(7)]

	Post=2015	Post=2015+2016
Dep. Var: Teachers Attendance	(1)	(2)
Monitoring (Treatment*Post)	0.0665***	0.0256
	(0.0172)	(0.0162)
School Teaching Quality	0.0375**	0.0301**
	(0.0150)	(0.0127)
School Training Quality	-0.00375	0.00607
	(0.0182)	(0.0147)
urban	0.0620	0.0159
	(0.0408)	(0.0346)
old-school	0.000548	-0.00469
	(0.00919)	(0.00863)
School Size(enrollment)	0.0460	0.0368
	(0.0448)	(0.0433)
Schools Facilities Controls	YES	YES
District FE	YES	YES
Year FE	YES	YES
Constant	0.880***	0.839***
	(0.0302)	(0.0350)
Observations	3,019	3,919
R-squared	0.075	0.055
Mean of the dep. Var:	.886	.883

Notes: Table-4(a) shows the main effect of the monitoring program on teacher's attendance. Column (1) represent Post=2015 and Pre=2012-2013. Column (2) represent Post=2015-2016 while Pre=2012-13. The outcome variable is the ratio of teachers present in school to the total appointed teachers. Variable Monitoring is a typical diff-in-diff interaction of treatment (KP) and Post (for corresponding year). Due to District and year fixed effect applied in each regression, we do not include variables for treatment and posts. Variables *School Teaching Quality* and *School Training Quality* are continuous variables showing the ratio of teachers with master's degree and specific training level to the total appointed teachers in each school. *School Facilities dummies* include availability of water, boundary, toilet, library, playground, laboratory, computer and internet. *School Size* is a continuous variable representing the ratio of children enrolled in surveyed school to the school with highest number of enrolled children. The data is taken from the ASER-Pakistan School Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is the surveyed government school. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

camera photograph, Banerjee & Duflo (2006) contend the external control of monitoring by someone within the institutional hierarchy such as headmaster or principle due to possible collusion with teachers. Although the case of KP monitoring program does not have this problem of external control (e.g. monitors do not belong to schools, rather they are externally appointed and their jobs are rotated), yet we cannot rule out the possibility of shirking by teachers in areas where teachers' distance from school is small.

Although, the effect decreased in the second year, the overall impact of IMU program appears to bring immediate improvement in the teacher's attendance over a large area. We check the robustness of our model [equation (7)] on various sub-samples of school levels such as primary schools (0-to-5 grades) and high schools (6-to-10 grades) and a reduced sample of districts bordering with FATA. There are sixteen districts in KP province which share border with any districts (agency) in FATA. The results (shown in section 8.3) are similar and follow the same pattern as observed in table 4(a). Also, we conduct a falsification test using the private schools data on post-program period by running the same regression as table 4(a). Result of falsification test (shown in Appendix I) reflect no systematic difference in the teacher's attendance pattern in private schools suggesting evidence in favor of IMU effect on government schools.

7.2. Learning Achievements

Even if monitoring increased teacher's presence in schools, it is not clear whether increased teachers presence affect learning achievements. In other words, whether teachers teach once they decide to be in the schools is the question of our interest in this section. Several factors can be considered in explaining the mechanisms through which any potential impact of increased oversight of teachers and schools might influence the learning capacity of children. The basic theory behind hypothesizing the direct effect of teachers monitoring on children performance is the marginal cost of teaching after a teacher is present in school. Especially at lower level such as primary schools where the subject contents usually are not much difficult and where few teachers are appointed per school. We assume that after being present in school, at lower level, teachers generally tend to teach (they don't want to shirk), hence children get benefited of the increased presence(Duflo & Hanna, 2005). In other words, getting teachers to schools may work effectively at the lower level schools. At higher level however, the marginal cost of teachers after being present in school might be higher given the subject contents difficulty at of higher grades such as maths, english and science subjects of 9th or 10th grade. Previous studies support the idea that developing countries such Pakistan and India, are suffering from the low teachers' capacity at higher level(Robert, 2005). Secondly, parents might positively respond to a large scale oversight program in rural areas in terms of sending children to schools. Although, in many poor societies the opportunity cost of sending children to school is greater than the benefits of educating them, however, recent evidence on education status in South Asia confirm the slackness of parents towards sending children to school due to school quality or teachers absence rather than economic reasons(Banerjee & Duflo, 2006; Glewwe & Kremer, 2006). At higher grade level such as grade 9th and 10th, teachers' absence from schools might affect parent's response. For example the potential financial incentives for teachers when they (deliberately) avoid teaching at schools in order to increase the chances of private tutoring, might pose a financial challenge for parents (Glewwe & Kremer, 2006). The third source of monitoring effect on children performance might be the link between teacher's attendance and children attendance. We check the program impact on children attendance measured as number of present children on the date of survey to the total enrollment in the school. Results shown in Appendix II suggest a slight increase (1.7% with 10 % significance level) in children attendance in year 2015, however, the magnitude is small indicating a subtle effect on the children attendance. The program effect on children attendance is not significant when we add 2016 as a post-program year. In either of our specifications, children attendance appears to be less affected (or unaffected) during the year immediately after the program. This is surprising as a number of studies document a strong association of teachers attendance with school participation and hence children academic performance. However, Glewwe & Kremer (2006) differentiate school participation from children attendance and argue that increasing teachers attendance and school quality might increase participation which means giving more time to school related tasks rather than mere attendance. Finally, governance reforms such as monitoring that target school quality appear to hold more promise than simply providing monetary incentives to teachers based on test scores. For example, threat of a top-down audit significantly reduces corruption (Olken, 2004) and teachers at schools that were inspected more often resulted in reduced absence (Chaudhury et al., 2005b). However, there are limited evidence that externally controlled monitoring when coupled with clear and credible threat of punishment induces “good” teaching behavior at school.

We turn to our second outcome of interest, children test performance to see the direct effect of the monitoring program on the test performed by enrolled children at home. We follow Glewwe & Kremer (2006) to obtain the reduced form relationship using model (8) [equation (6)] in estimating normalized test performance in three different subjects e.g Reading, Mathematics and English. With regard to the level of difficulty, the ASER test questions[8] for each subject are designed to measure the very basic Learning, English and Math ability in view of achieving SDG indicator 4.2.1(ASER, 2016). According to ASER reports, the survey is pitched to grades 2 and 3 competencies only, corresponding with the SDG indicators for tracking learning at the lower primary level. The survey procedure in ASER annual publications also confirms the low difficulty levels of tests. In addition to that, ASER data survey also include three additional questions(called bonus questions) for reading, two bonus questions related to math and one additional question related to english. Although, these additional questions might still be easier, we attempt to utilize them to construct normalized test variable for upper grade children (See Appendix X for details on the procedure of ASER test questionnaire). In their paper on ASER-(Pratham), India, Banerji et al., (2013) describe that children of grade 3 onwards have no difficulties in completing all questions asked by ASER survey. Nevertheless, in view of the extremely discouraging learning status reported by different organizations in Pakistan over the last few years, we rely on the ASER’s basic test questionnaires (five questions each subject) for lower grade children to gauge the ability level of enrolled children. We aggregate the individual dummies for each of five questions in each subject to construct a raw score for each surveyed child and subsequently normalize[9] by year, district and individual grades to obtain a reliable measure of test score. A similar procedure was adopted for ASER bonus questions to create normalized test score for children enrolled in higher grade children (see Appendix VI and VII) for upper grade children and including year 2014 as pre-treatment.

Table 4(b) reports the direct program effect on the normalized test score for lower grade (0 to 5) enrolled children using 2012 and 2013 as Pre-Program. For simplicity purpose, we only report coefficients of the interaction term of KP and Post to show the differential effect of the treatment after the program. Previous literature on learning outcomes documents effects of factors such as individual characteristics, parent’s education and household characteristics on the learning performance of children(Abdulkadiroğlu et al., 2018; Azam et al., 2016; Banerjee

---

[8] The ASER HH survey contains five basic questions ranging from low difficulty to higher difficulty. For example, for reading five test dummies are whether the surveyed child is at Beginners level, can read letters, can read words can read sentences, can read story. Similar procedure is adopted for mathematics and English questions.

[9] After constructing the raw score, we standardize the score as: $z=\frac{(x-\bar{x})}{\sigma}$ where, $\bar{x}$ and σ are the mean and standard deviation of the test score by each individual grade.

et al., 2007; Croke, 2014; Jackson, 2009; Raikes, 2016). We therefore control for individual child-specific characteristics, parents education and household characteristics along with district fixed effect and year fixed effect. The first three columns report the program effect on the Reading, Maths and English test scores normalized by year, district and individual grade for year 2015 as post-program. The last three columns report the two years (2015 & 2016) program effect on normalized test score of lower grade children. We observe a significantly positive effect of the IMU program on the enrolled children performance in maths and English while positive (but not significant) effect in reading. Conditional on child-specific controls, parent's education and household characteristics, on average, being in the KP province increases a child's normalized test performance by 0.07 stand deviations (SD) points in Reading, 0.13 SD points in Maths and 0.11 SD points in English. Adding 2016 s post-program year into analysis shows that there is not significant direct effect on IMU on children test performance. We also check the direct effect of the program on higher grade (6 to 10) children. The results are reported in Appendix table A4 (b). Since, data on the higher grade related questions was not available in year 2012, therefore, we report the results of higher grade children in table A4 (b) which include 2014 as pre-program period. Though significant at 10% level, the program effect is positive for higher grade children in Reading bonus question and English bonus questions. This decreasing effect of program on higher grade children is consistent with earlier findings by Banerji et al., (2013) on the difficulty level of the ASER-India test questions. In estimating results for table 4(b) and table A4 (b), we only include children that are currently enrolled in government schools and for whom information on covariates were available.

After adding 2016 as post-program year, the direct program effect on lower grade children normalized test score is positive, but not significant indicating a decrease in the program effect during the year 2016. Nevertheless, for higher grade children (as we show in A4 (b), the program effect persisted, though slightly reduced. Controlling for child-specific factors, parents and household characteristics, and the district and year fixed effects, the IMU increases the ability of higher grade children to answer bonus-test questions by 0.127 SD points for Reading, 0.136 for English at 5% significance level. This decrease in effect of children test performance coincides with the decrease in teachers' attendance in 2016 as reflected in Table 4(a) giving more weight to the possibility of direct effect of the monitoring program on children test performance. One way of linking the decreasing effect on children performance might be the reducing efforts of teachers even though they are present in the schools. Previous evidence also does not rule out this possibility. In estimating the effect of teacher's incentive program in Kenya, Glewwe, Ilias and Kremer (2003) find a short run increase in learning score and argue that gains in learning were only temporary and were not accompanied by increases in teaching efforts.

Our results on the children test score provide evidence in support of the idea that absence of teachers at lower grades schools causes low learning achievements in developing countries. Thus addressing teacher's absence at lower level could be a key policy direction that can positively affect learning achievements of lower grade children. Such a policy direction might

Table 4 (b): Program Effect, Normalized Test Score [Equation (8)]

	Normalized Test Score					
	Post=2015			*Post=2015 & 16(pooled)*		
	Reading	Math	English	Reading	Math	English
Monitoring (KP*Post)	0.0722	0.137**	0.119**	-0.00987	0.00398	0.0210
	(0.0624)	(0.0538)	(0.0588)	(0.0591)	(0.0504)	(0.0556)
Child -Related Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parents Education Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Characteristics Dummies	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.390***	-0.358***	-0.151	-0.233***	-0.264***	-0.000539
	(0.0934)	(0.0866)	(0.0932)	(0.0793)	(0.0780)	(0.0812)
Observations	41,142	40,923	40,922	58,678	58,476	58,475
R-squared	0.096	0.097	0.093	0.065	0.078	0.081

Notes: Table 4(b) reports the Post-program difference using diff-in-diff estimates on the *children test performance* for Post=2015 vs Pre=12-to-2013(pooled). The data is from the ASER Household Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is surveyed 3 to 16 year's old child enrolled in government school from Grade-0 to grade-5. The dependent variable is the *test score normalized by year, district and grade*. Variable Monitoring is an interaction of *treated province* (KP) and Post (which is equal to 1 if year==2015 and 0 if year=2012 or 2013). Fixed Effect on individual grade, District and year applied in each regression. Child-related controls include age, private tuition; parents education controls include, mother and father highest education in years; household characteristics include ownership, house condition, and availability if electricity, mobile and television facilities. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively. (See Appendix table A4 (b) for complete regression results.)

combine external control monitoring tools such as IMU with appropriate incentives mechanisms to maintain the quality of schools on sustainable basis. With regard to higher grade children, besides increased oversight, teacher's education or training quality may be coupled with efforts of increasing their attendance to ensure learning achievements.

7.3.Enrollment Status

Enrollment has been widely used as a key indicator for achieving sustainable development goals particularly children of age 5 to 16 in developing countries. A large number of out-of-school children in rural areas of Pakistan has been a persisting issue that requires effective solution. According to recent reports, Pakistan continue to suffer for low enrollment and high dropout rate at primary and middle level schooling(Gouleta, 2015). A review by the International growth Centre (ICG), in Khyber Pakhtunkhwa(KP) province in 2012-13 shows only 63% of 4-9 years old children were enrolled in schools with a much lower (56%) female enrollment(CDPR, 2014). For higher grade, the net enrollment is even worst. For example, for middle schools, the net enrollment was hardly 40% reflecting a significant dropout or no-enrollment during the middle school age group (11 to 15 years).

To investigate the overall effect of the monitoring program on the enrollment status of children surveyed at home, we analyze ASER household survey data from 2012 to 2016. The ASER household survey include a variable on the status of children of age 5 to 16 asking whether they are enrolled in schools or not. We drop all those children enrolled in private school, madrassas or any other school to obtain reduced sample of children either enrolled in government schools or not enrolled. We attempt our diff-in-diff model for post-program year as 2015 only and 2015 and 2016 together to see the two years post program effect. Results reported in table 4(d) are suggestive of the positive direct effect of monitoring program on gross government school enrollment. Since enrollment status is a binary variable, in addition to simple OLS, we also compare Probit model while controlling for all household and child related characteristic. The OLS estimates show that conditional on household characteristics, compared to FATA, the probability of a schooling age child to be enrolled in government school increases in the KP province by 3.1% in 2015 while this effect is not significant in 2016. The Probit marginal effects imply that children in KP province have a 4% higher probability of getting enrolled in government schools compared to FATA. Both OLS and probit results points to the similar drop in the gross enrollment of children in 2016 consistent with a similar trend in the children's test outcomes and teachers attendance. However this effect should be interpreted carefully due to two reasons. First, children enrollment mainly depends on school density. In other words, if the government schools (e.g per village) increases, it might increase the gross enrollment per village. Secondly, each year, there might be a linear trend in population growth coupled with increasing awareness campaigns by government and non-government organizations. While we are applying year and district fixed effect which controls for any district and year specific characteristics, we believe this effect may come through parents whose behavior might be affected by the government's monitoring programs. Earlier studies also support the idea that parents positively respond to increasing school quality in terms of enrolling their children in schools(Berman et al., 2013; Glewwe & Kremer, 2006; Jones et al., 2014). Although these effects seems small, considering the status of out-of-school children in developing countries particularly Pakistan, the implication of these results is worth noticing. If a government policy targeted at one aspect of schooling such as teachers' attendance, affect the children enrollment and test performance simultaneously besides increasing school quality, then the cost of such policies should be evaluated in terms all three outcomes of education; school quality, learning outcomes and enrollment.

Table-4(d): Program Effect on Children Enrollment Status

Dep. Var: Enrollment Status[0,1]	Post=2015		Post=2016 & 2016	
	OLS	Probit	OLS	Probit
Monitoring(treatment*Post)	0.0317**	0.040*	0.00105	0.004
	(0.0152)	(0.017)	(0.0133)	(0.015)
Child Age	0.0352***	0.039**	0.0363***	0.040**
	(0.000610)	(0.001)	(0.000547)	(0.001)
Gender(Female=1)	-0.196***	-0.217**	-0.191***	-0.212**
	(0.00494)	(0.005)	(0.00427)	(0.005)
Mother Highest Education	-0.00161***	-0.002**	-0.00189***	-0.002**
	(0.000555)	(0.001)	(0.000501)	(0.001)
Father Highest Education	0.00553***	0.007**	0.00587***	0.007**
	(0.000414)	(0.000)	(0.000366)	(0.000)
House-ownership	0.0134*	0.016	0.00772	0.010
	(0.00805)	(0.009)	(0.00735)	(0.008)
HH- Size	-0.00161***	-0.002**	-0.00175***	-0.002
	(0.000498)	(0.001)	(0.000487)	(0.001)**
Urban Districts	0.0751**	0.095*	0.0655*	0.083
	(0.0355)	(0.044)	(0.0352)	(0.044)
HH-Facilities Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Constant	0.506***		0.517***	
	(0.0212)		(0.0179)	
Observations	144,988	144,988	188,579	188,579
R-squared	0.195		0.190	

Notes: Table 4 (d) reports the Post-program difference using diff-in-diff OLS coefficients and Probit marginal effects on the *enrollment status of surveyed children.* The first two columns reports results on the 2015 as post-program only while the last two columns reports post-program period as 2015 & 2016. The pre-program period in all columns is 2012 to 2014 pooled. The dependent variable is a binary which child is enrolled in government school and zero otherwise. The sample does not include children that are enrolled in private or other schools. Variable Monitoring is an interaction of *treated province* (KP) and Post-program period. District and year fixed effect are applied throughout regression while controls for household facilities are also included. The data is from the ASER Household Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is surveyed 3 to 16 year's old child. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively

7.4. Robustness Checks

Table -5 : Program Effect on Only Primary Schools[grade0 to 5]

Dep. Var: Teachers Attendance	(1)	(2)	(3)	(4)
Monitoring (Treatment*Post)	0.0657***	0.0567***	0.0243	0.0141
	(0.0182)	(0.0209)	(0.0165)	(0.0194)
School Teaching Quality	0.0384**	0.0396**	0.0298**	0.0316**
	(0.0162)	(0.0194)	(0.0140)	(0.0160)
School Training Quality	0.0243	0.00995	0.0263*	0.0175
	(0.0165)	(0.0208)	(0.0147)	(0.0172)
urban	-0.0164	-0.0244	-0.0521	-0.0255
	(0.0379)	(0.0640)	(0.0440)	(0.0393)
old-school	-0.00946	0.000805	-0.0104	-0.00386
	(0.00986)	(0.0115)	(0.00936)	(0.0107)
School Size(enrollment)	0.199***	0.159**	0.145***	0.105
	(0.0550)	(0.0693)	(0.0549)	(0.0683)
Schools Facilities Controls	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Constant	0.785***	0.764***	0.827***	0.819***
	(0.0355)	(0.0405)	(0.0307)	(0.0342)
Observations	2,765	2,087	3,429	2,751
R-squared	0.082	0.090	0.066	0.065
Mean of the dep. Var:	0.887	0.887	0.887	0.887

Notes: Table-5 shows the main effect of the monitoring program on teacher's attendance in government run primary schools only. Column (1) & (2) represent Post=2015 while Pre=2012-2014 & Pre=2012-2013 respectively. Similarly Column (3) & (4) represent Post=2015-2016 while Pre=2012-14(1) & Pre=2012-13 (2) respectively. The outcome variable is the ratio of teachers present in school to the total appointed teachers. Variable Monitoring is a typical diff-in-diff interaction of treatment (KP) and Post (for corresponding year). Due to District and year fixed effect applied in each regression, we do not include variables for treatment and posts. Variables *School Teaching Quality* and *School Training Quality* are continuous variables showing the ratio of teachers with master's degree and specific training level to the total appointed teachers in each school. *School Facilities dummies* include availability of water, boundary, toilet, library, playground, laboratory, computer and internet. *School Size* is a continuous variable representing the ratio of children enrolled in surveyed school to the school with highest number of enrolled children. The data is taken from the ASER-Pakistan School Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is the surveyed government primary school where children from grade0 to 5 are taught. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

Table -6 : Program Effect on Reduced Sample of Bordering Districts

Dep. Var: Teachers Attendance	(1)	(2)	(3)	(4)
Monitoring (Treatment*Post)	0.0800***	0.0779***	-0.00384	-0.000501
	(0.0176)	(0.0178)	(0.0220)	(0.0230)
School Teaching Quality	0.0462**	0.0313	0.0356**	0.0257
	(0.0218)	(0.0255)	(0.0178)	(0.0199)
School Training Quality	0.0127	0.000746	0.0123	0.00587
	(0.0252)	(0.0322)	(0.0217)	(0.0255)
urban	-0.0478	0.0393	-0.0905**	-0.0637
	(0.0481)	(0.0456)	(0.0454)	(0.0497)
old-school	-0.00143	0.00716	-0.00723	-0.00390
	(0.0130)	(0.0151)	(0.0122)	(0.0140)
School Size(enrollment)	0.127**	0.0733	0.138**	0.100
	(0.0556)	(0.0634)	(0.0542)	(0.0627)
Schools Facilities Controls	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Constant	0.759***	0.738***	0.789***	0.775***
	(0.0385)	(0.0427)	(0.0348)	(0.0378)
Observations	1,515	1,123	1,845	1,453
R-squared	0.056	0.070	0.050	0.056
Mean of the dep. Var:	0.871	0.871	0.871	0.871

Notes: Table-6 shows the main effect of the monitoring program on teacher's attendance in government run schools using the reduced sample of districts bordering with FATA and FATA. Column (1) & (2) represent Post=2015 while Pre=2012-2014 & Pre=2012-2013 respectively. Similarly Column (3) & (4) represent Post=2015-2016 while Pre=2012-14(1) & Pre=2012-13 (2) respectively. The outcome variable is the ratio of teachers present in school to the total appointed teachers. Variable Monitoring is a typical diff-in-diff interaction of treatment (KP) and Post (for corresponding year). Due to District and year fixed effect applied in each regression, we do not include variables for treatment and posts. Variables *School Teaching Quality* and *School Training Quality* are continuous variables showing the ratio of teachers with master's degree and specific training level to the total appointed teachers in each school. *School Facilities dummies* include availability of water, boundary, toilet, library, playground, laboratory, computer and internet. *School Size* is a continuous variable representing the ratio of children enrolled in surveyed school to the school with highest number of enrolled children. The data is taken from the ASER-Pakistan School Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is the surveyed government school. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

8. CONCLUSION

Initiatives to reduce teachers' absenteeism in public schools range from offering incentives to instituting school committees to decentralizing of education to local government to externally controlled monitoring etc., however, to what extent such initiatives persist their effect and how much they affect children learning performance is rarely understood. In this paper, we examine the effect of a large scale public schools monitoring program featured by the use of smart-phone aided facility through professionally trained monitors in the KP province, Pakistan. We use five years data from a country wide annual representative survey to compare treated region with a neighboring untreated region that share similar characteristics in all aspects except the program. Our data consists a rich set of variables that allow estimation of education production function in the context of a purely exogenous intervention. Our findings suggest that monitoring of government schools through trained monitors equipped with smart-phone-aided biometric facility improved teacher's attendance by nearly 8% in the year immediately following the program. However, this effect decreases by nearly half after two years of the program introduction.

We also find the program's direct effect on the enrolled children's test performance at home. Enrolled children's standardized Reading, Math and English ability in the monitored schools has improved significantly by 0.07, 0.13 and 0.11 standard deviations points respectively at the lower (0-5) grades. There is slight improvement in the standardized test performance of higher grade children. We also find a positive immediate effect of the program on the likelihood of school-aged children enrollment into government schools suggesting responsiveness of parents towards a large scale program.

Our results on the children performance provide evidence in support of the idea that absence of teachers at lower grades schools causes low learning achievements in developing countries. Thus addressing teacher's absence at lower level could be a key policy direction that can positively affect learning achievements of lower grade children. Such a policy direction might be combined with external control monitoring tools such as IMU with appropriate incentives mechanisms to maintain the quality of schools on sustainable basis. With regard to higher grade children, besides increased oversight, teacher's education or training quality may be coupled with efforts for increasing their attendance to ensure learning achievements.

Two broad implications can be derived from our results. First, incorporation of advanced technology in schools monitoring has a stronger effect on the teachers and children performance simultaneously. Such initiatives might have wide rage effects than the targeted outcomes. Secondly, how long such effects sustain, depends on complementary measures that links teachers performance with children performance.

9. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Despite having a clear identification strategy, our work is subject to certain limitations. First, we use survey data that is collected on annual basis, and only capture the yearly inspections of schools. Using monthly data on teacher's attendance might be more useful in evaluating any differential effect between KP and FATA schools performance. Secondly, we couldn't access more detailed administrative data on the characteristics of monitors employed by IMU for more in-depth analysis of the program. Data collected by IMU staff on teacher's attendance and school performance might be useful for comparison of ASER data and IMU data. Thirdly, the test questions for higher-grade children might weakly represent their performance because of low standard of the questions designed by ASER. ASER's test questions mainly target low grade children as shown in Appendix X. Although we utilize the bonus questions to create normalize test score for higher grade children, a more standardized design of test taken at home for higher grade children would be more useful in gauging children performance. Finally, establishing a systematic channel between teacher's attendance and children performance is important despite our findings that monitoring program has directly affected children test score. Given the differential effect in 2015, future research might utilize two stage least square (2SLS) approach to for establishing a clear link between teacher's attendance and children test score.

References

Abdulkadiroğlu, A., Pathak, P. A., & Walters, C. R. (2018). Free to choose: Can school choice reduce student achievement. *American Economic Journal: Applied Economics*, *10*(1), 175–206.

Alcázar, L., Rogers, F. H., Chaudhury, N., Hammer, J., Kremer, M., & Muralidharan, K. (2006). Why are teachers absent? Probing service delivery in Peruvian primary schools. *International Journal of Educational Research*, *45*(3), 117–136.

Alcott, B., & Rose, P. (2015). Schools and learning in rural India and Pakistan: Who goes where, and how much are they learning? *Prospects*, *45*(3), 345–363.

Attaullah, ., & Malik, S. Y. (2015). MDGs on Universalization of Primary Education: Analysis of Gaps Between Commitment and Implementation in Pakistan. *Mediterranean Journal of Social Sciences*, *6*(6), 224–231.

Azam, M., Kingdon, G., & Wu, K. B. (2016). Impact of private secondary schooling on cognitive skills: evidence from India†. *Education Economics*, *24*(5), 465–480.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, *122*(3), 1235–1264.

Banerjee, A., & Duflo, E. (2006). Addressing Absence. *Journal of Economic Perspectives*, *20*(1), 117–132.

Banerjee, A. V., & Duflo, E. (2009). The Experimental Approach to Development Economics. *Annual Reviews of Economics*, *1*, 151–178.

Banerji, R., Bhattacharjea, S., & Wadhwa, W. (2013). The annual status of education report (ASER). *Research in Comparative and International Education*, *8*(3), 387–396.

Berman, E., Bhalotra, S., Bharadwaj, P., Chin, A., Clemens, J., Cullen, J., … Gupta, S. (2013). Cycling to School: Increasing Secondary School Enrollment for Girls in India. *American Economic Journal: Applied Economics*, *9*(3), 321–350.

Bukowitz, W. R., Williams, R. L., & Mactas, E. S. (2004). Human capital measurement. *Research-Technology Management*, *47*(3), 43–49.

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

Card, B. D., & Krueger, A. B. (2000). Minimum Wages and Employment : A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, *90*(5), 1397–1420.

CDPR. (2014). Teacher and School Administrator Incentives for Improving Education Delivery in Khyber. In *Consortium for Development Policy Research*.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H. (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries. *Journal of Economic Perspectives*, *20*(1), 91–116.

Croke, K. (2014). The long run effects of early childhood deworming on literacy and numeracy : Evidence from Uganda. *Havard School of Public Health*, 1–25.

Duflo, E. (2007). *Using Randomization in Development Economics: A toolkit.*

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. *Handbook of Development Economics*, *4*(07), 3895–3962.

Duflo, E., & Hanna, R. (2005). *Monitoring Works: Getting Teachers to Come to School* (NBER WORKING PAPER SERIES).

French, R., Kingdon, G., & others. (2010). The relative effectiveness of private and government schools in Rural India: Evidence from ASER data. *London: Institute of Education*, (January).

Glewwe, P., & Kremer, M. (2006). *Schools, Teachers, and Education Outcomes in Developing Countries*. *Handbook of the Economics of Education* (Vol. 2).

Gouleta, E. (2015). Educational assessment in Khyber Pakhtunkhwa Pakistan's North-West Frontier Province: Practices, issues, and challenges for educating culturally linguistically diverse and exceptional children. *Global Education Review*, *2*(4), 19–39.

Jackson, C. K. (2009). Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation. *Journal of Labor Economics*, *27*(2), 213–256.

Jones, S., Schipper, Y., Ruto, S., & Rajani, R. (2014). Can your child read and count? Measuring learning outcomes in East Africa. *Journal of African Economies*, *23*(5), 643–672.

Koedel, C., & Betts, J. (2007). Re-Examining the Role of Teacher Quality In the Educational Production Function. *Working Papers*.

Munene, I. I. (2015). Providing basic education for all in Africa: What we have learned. *Africa Education Review*, *12*(1), 1–6.

Press, W. (1969). Determinants of School Enrollment and School Performance Author (s): John Conlisk Source : The Journal of Human Resources , Vol . 4 , No . 2 (Spring , 1969), pp . 140-157 Published by : University of Wisconsin Press Stable.

Raikes, H. (2016). Measuring of child development and learning: Background paper prepared for the 2016 Global Education Monitoring Report, 1–18. Retrieved from

Robert, H. (2005). *Education Reform in Pakistan*. *Education*. Woodrow Wilson International Center for Scholars, Washington, DC.

Scott, D., & Wimbush, J. (1991). Teachers Absenteeism in Secondary Education. *Education Administration Quaterly*, *27*(4), 206–529.

Zaka, F., & Maheen. (2010). *2013-2018 Five Years of Education Reforms.Wins, Losses and challenges for 2018-2023*. *Alif-Ailan*. IslamabaD.

Appendix I

Table –A1: Falsification Test on Private School Data

Dep. Var: Teachers Attendance Ratio	Post=2015(a)	Post=2015(b)	Post(a)	Post(b)
Monitoring (Treatment*Post)	0.000348	-0.0196	-0.0244	-0.0473
	(0.0335)	(0.0375)	(0.0257)	(0.0298)
School Teaching Quality	0.0348*	0.0239	0.0364*	0.0292
	(0.0190)	(0.0225)	(0.0198)	(0.0234)
School Training Quality	-0.00510	0.00115	-0.00739	-0.00506
	(0.0244)	(0.0322)	(0.0239)	(0.0304)
urban	0.0166	0.0408	0.00401	-0.0332
	(0.0297)	(0.0311)	(0.0276)	(0.0407)
old schools	-0.0232	-0.0260	-0.0262	-0.0276
	(0.0198)	(0.0220)	(0.0193)	(0.0214)
enrollment	0.0768*	0.0530	0.0874**	0.0718
	(0.0402)	(0.0507)	(0.0395)	(0.0488)
Schools Facilities Controls	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Constant	0.511***	0.576***	0.776***	0.545***
	(0.0267)	(0.189)	(0.0346)	(0.0297)
Observations	1,674	1,292	1,944	1,562
R-squared	0.064	0.100	0.057	0.081

Notes: Table-A1 reports the falsification test of the monitoring program on teacher's attendance using private school data. We run the same specification of our main effect on the private school data to see any systematic trend in the teacher's attendance of private school data. Column (1) & (2) represent Post=2015 while Pre=2012-2014(1) & Pre=2012-2013 respectively. Similarly column (3) & (4) represent Post=2015-2016 while Pre=2012-14(1) & Pre=2012-13(2) respectively. The outcome variable is the ratio of teachers present in school to the total appointed teachers. Variable Monitoring is an interaction of treatment (KP) and Post (for corresponding year). Due to District and year fixed effect applied in each regression, we do not include variables for treatment and posts. Variables *School Teaching Quality* and *School Training Quality* are continuous variables which show the ratio of teachers with master's degree and specific training level to the total appointed teachers in each school. *School Facilities controls* include availability of water, boundary, toilet, library, playground, laboratory, computer and internet. *Enrollment* is a continuous variable representing the ratio of children enrolled in surveyed school to the school with highest number of enrolled children. The data is taken from the ASER-Pakistan School Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is the surveyed private school. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

Appendix II

Appendix Table –A2: Program Impact on Children Attendance in Government Schools

Dep. Var: Children Attendance Ratio	Post=2015(a)	Post1=2015(b)	Post(a)	Post(b)
Monitoring (Treatment*Post)	0.0177*	-0.00579	0.00873	-0.0168
	(0.00973)	(0.0116)	(0.0103)	(0.0121)
School Teaching Quality	0.0104	0.0166*	-0.000394	0.00357
	(0.00817)	(0.00959)	(0.00800)	(0.00912)
School Training Quality	-0.00269	0.00140	0.00139	0.00773
	(0.00952)	(0.0112)	(0.00897)	(0.0103)
urban	0.00428	-0.0510*	-0.0215	-0.0434
	(0.0258)	(0.0293)	(0.0229)	(0.0276)
Old schools	-0.00862	-0.00905	-0.00790	-0.00865
	(0.00534)	(0.00620)	(0.00528)	(0.00606)
Schools Facilities Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Constant	0.802***	0.800***	0.792***	0.790***
	(0.0186)	(0.0214)	(0.0170)	(0.0190)
Observations	4,053	3,019	4,953	3,919
R-squared	0.095	0.125	0.092	0.112

Notes: Table-A2 shows main effect of the monitoring program on children attendance. Column (1) & (2) represent Post=2015 while Pre=2012-2014(1) & Pre=2012-2013 respectively. Similarly column (3) & (4) represent Post=2015-2016 while Pre=2012-14(1) & Pre=2012-13(2) respectively. The outcome variable is the ratio of children present in school to the total enrollment. Variable Monitoring is an interaction of treatment (KP) and Post (for corresponding year). Due to District and year fixed effect applied in each regression, we do not include variables for treatment and posts. Variables *School Teaching Quality* and *School Training Quality* are continuous variables which show the ratio of teachers with master's degree and specific training level to the total appointed teachers in each school. *School Facilities controls* include availability of water, boundary, toilet, library, playground, laboratory, computer and internet. The data is taken from the ASER-Pakistan School Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is the surveyed government school. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

Appendix III

Table A3 (b): Pre-Program Difference, Normalized Test Score [Equation (8)]

	Normalized Test Score					
	Lower Grades-(0-to -5)			Upper Grade (6 -10)		
	Reading	Math	English	Reading	Math	English
Pre-Program Difference (KP*Post)	-0.150**	-0.0112	-0.0657	0.0218	0.0783	-0.334***
	(0.0633)	(0.0571)	(0.0601)	(0.0622)	(0.0590)	(0.0917)
Child -Related Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parents Education Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Characteristics Dummies	Yes	Yes	Yes			
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.443***	-0.00221	-0.0470	1.211***	0.685***	0.257
	(0.0927)	(0.0846)	(0.0906)	(0.134)	(0.161)	(0.186)
Observations	38,923	38,818	38,762	11,054	11,054	10,942
R-squared	0.068	0.062	0.069	0.115	0.111	0.116

Notes: Table 3(b) reports the pre-program difference using diff-in-diff estimates on the *children test performance* for Post=2014 vs Pre=2012 & 2013 using the ASER Household Survey data. Standard errors clustered at village level are shown in parentheses. The unit of observation is surveyed 3 to 16 year's old child enrolled in government school from Grade-0 to grade-5(first three columns) and grade-6 to 10(last three columns). The dependent variable is the *test score normalized by grade*. The pre-program difference is a typical diff-in-diff interaction of *to-be-treated province* (KP) and Post (which is equal to 1 if year==2014 and 0 if year=2012 or 2013). Fixed Effect on individual grade, District and year applied in each regression. Child-related controls include age, private tuition; parent's education controls include, mother and father highest education in years; household characteristics include ownership, house condition, and availability of electricity, mobile and television facilities. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

Appendix IV

Table -A4 (a) : Teachers Attendance Ratio[Equation(7)]

Dep. Var: Teachers Attendance	(1)	(2)	(3)	(4)
Monitoring (Treatment*Post)	0.0756***	0.0665***	0.0344**	0.0256
	(0.0151)	(0.0172)	(0.0140)	(0.0162)
School Teaching Quality	0.0344***	0.0375**	0.0278**	0.0301**
	(0.0125)	(0.0150)	(0.0111)	(0.0127)
School Training Quality	0.0129	-0.00375	0.0167	0.00607
	(0.0143)	(0.0182)	(0.0125)	(0.0147)
urban	-0.0303	0.0620	0.00645	0.0159
	(0.0469)	(0.0408)	(0.0310)	(0.0346)
old-school	-0.00379	0.000548	-0.00650	-0.00469
	(0.00785)	(0.00919)	(0.00751)	(0.00863)
School Size(enrollment)	0.0945***	0.0460	0.0789**	0.0368
	(0.0357)	(0.0448)	(0.0351)	(0.0433)
Schools Facilities Controls	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Constant	0.884***	0.880***	0.846***	0.839***
	(0.0295)	(0.0302)	(0.0344)	(0.0350)
Observations	4,053	3,019	4,953	3,919
R-squared	0.066	0.075	0.054	0.055
Mean of the dep. Var:	.883	.886	.881	.883

Notes: Table-4(a) shows the main effect of the monitoring program on teacher's attendance. Column (1) & (2) represent Post=2015 while Pre=2012-2014 & Pre=2012-2013 respectively. Similarly Column (3) & (4) represent Post=2015-2016 while Pre=2012-14(1) & Pre=2012-13 (2) respectively. The outcome variable is the ratio of teachers present in school to the total appointed teachers. Variable Monitoring is a typical diff-in-diff interaction of treatment (KP) and Post (for corresponding year). Due to District and year fixed effect applied in each regression, we do not include variables for treatment and posts. Variables *School Teaching Quality* and *School Training Quality* are continuous variables showing the ratio of teachers with master's degree and specific training level to the total appointed teachers in each school. *School Facilities dummies* include availability of water, boundary, toilet, library, playground, laboratory, computer and internet. *School Size* is a continuous variable representing the ratio of children enrolled in surveyed school to the school with highest number of enrolled children. The data is taken from the ASER-Pakistan School Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is the surveyed government school. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

Appendix V

Table A4 (b): Program Effect, Normalized Test Score [Equation (8)]

	Normalized Test Score					
	Lower Grades-(0-to -5)			Upper Grade (6 -10)		
	Reading	Math	English	Reading	Math	English
Monitoring (KP*Post)	0.130**	0.140***	0.150***	0.100*	0.0307	0.121*
	(0.0524)	(0.0478)	(0.0508)	(0.0594)	(0.0561)	(0.0707)
Child -Related Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parents Education Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Characteristics Dummies	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.431***	-0.0736	-0.0709	1.439***	1.017***	0.710***
	-0.0779	-0.0703	-0.0751	(-0.124)	(-0.13)	(-0.149)
Observations	60,308	60,082	60,076	17156	17156	17059
R-squared	0.067	0.067	0.070	0.147	0.160	0.143

Notes: Table 4(b) reports the Post-program difference using diff-in-diff estimates on the *children test performance* for Post=2015 vs Pre=12-to-2014(pooled). The data is from the ASER Household Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is surveyed 3 to 16 year's old child enrolled in government school from Grade-0 to grade-5(first three columns) and grade-6 to 10(last three columns). The dependent variable is the *test score normalized by grade*. Variable Monitoring is an interaction of *treated province* (KP) and Post (which is equal to 1 if year==2015 and 0 if year=2012 or 2014). Fixed Effect on individual grade, District and year applied in each regression. Child-related controls include age, private tuition; parents education controls include, mother and father highest education in years; household characteristics include ownership, house condition, and availability if electricity, mobile and television facilities. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively. (See Appendix table A4 (b) for complete regression results.)

Appendix VI

Table A4 (c): Program Effect, Normalized Test Score [Equation (8)]

	Normalized Test Score					
	Lower Grades-(0-to -5)			Upper Grade (6 -10)		
	Reading	Math	English	Reading	Math	English
Monitoring (KP*Post)	0.0730	0.0221	0.0657	0.127**	0.0198	0.136**
	(0.0474)	(0.0440)	(0.0461)	(0.0506)	(0.0490)	(0.0618)
Child -Related Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parents Education Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Characteristics Dummies	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.355***	-0.0691	-0.0226	1.435***	0.963***	0.814***
	-0.0724	-0.0672	-0.071	(0.107)	(0.114)	(0.129)
Observations	77724	77515	77509	21,744	21,744	21,373
R-squared	0.053	0.064	0.067	0.128	0.147	0.113

Notes: Table 4(b) reports the Post-program difference using diff-in-diff estimates on the *children test performance* for Post=2015 & 2016 vs Pre=12-to-2014 (pooled). The data is from the ASER Household Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is surveyed 3 to 16 year's old child enrolled in government school from Grade-0 to grade-5(first three columns) and grade-6 to 10(last three columns). The dependent variable is the *test score normalized by grade*. Variable Monitoring is an interaction of *treated province* (KP) and Post (which is equal to 1 if year==2015 & 2016 and 0 if year=2012 or 2014). Fixed Effect on individual grade, District and year applied in each regression. Child-related controls include age, private tuition; parents education controls include, mother and father highest education in years; household characteristics include ownership, size, house condition, and dummies for availability if electricity, mobile and television facilities. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

Appendix VII

Program Effect on Children Test Performance

Table-A4(b): Normalized Test Score [Post=2015 vs Pre==2012-14]

	Grade-0 to Grade-5			Grade-5 to Grade-10		
	Reading	Math	English	Reading	Math	English
DiD(treatment*Post)	0.130**	0.140***	0.150***	0.100*	0.0307	0.121*
	(0.0524)	(0.0478)	(0.0508)	(0.0594)	(0.0561)	(0.0707)
Post(=2015, Pre=2012-14)	0.0310	-0.00182	0.00452	-0.246***	-0.119**	-0.222***
	(0.0478)	(0.0418)	(0.0450)	(0.0496)	(0.0494)	(0.0630)
Treatment(KP)	-0.169**	-0.395***	-0.275***	-0.543***	-0.404***	-0.844***
	(0.0850)	(0.0787)	(0.0819)	(0.130)	(0.117)	(0.132)
Child Age	0.0847***	0.0722***	0.0674***	-0.00800	-0.0109	0.0157
	(0.00814)	(0.00749)	(0.00737)	(0.00828)	(0.00854)	(0.00969)
Mother Highest Education	-0.00139	-0.00102	0.000449	-0.00220	0.000379	0.00317
	(0.00238)	(0.00214)	(0.00229)	(0.00296)	(0.00258)	(0.00311)
Father Highest Education	0.00103	0.00150	0.00161	0.00512***	0.00410**	-0.000725
	(0.00157)	(0.00155)	(0.00158)	(0.00177)	(0.00176)	(0.00221)
House-ownership	0.0737***	0.0367	0.0219	0.0295	0.0228	0.0492
	(0.0256)	(0.0242)	(0.0261)	(0.0321)	(0.0304)	(0.0385)
Private Tutoring	0.196***	0.154***	0.160***	-0.00228	-0.0585	0.160***
	(0.0407)	(0.0380)	(0.0426)	(0.0428)	(0.0451)	(0.0435)
Electricity Availability	0.0136	-0.0412	-0.0542	-0.0325	0.0168	0.0892*
	(0.0454)	(0.0405)	(0.0430)	(0.0419)	(0.0427)	(0.0514)
Mobile service Availability	0.0848***	0.0543**	0.0889***	0.0975***	0.0510	0.0840**
	(0.0255)	(0.0233)	(0.0254)	(0.0363)	(0.0340)	(0.0390)
TV availability	0.0241	0.0158	0.00277	-0.0299	-0.0432**	0.0207
	(0.0191)	(0.0184)	(0.0186)	(0.0221)	(0.0210)	(0.0240)
House condition	0.0375	0.0436	0.0497	0.0297	-0.0135	0.0581
	(0.0331)	(0.0320)	(0.0321)	(0.0412)	(0.0381)	(0.0456)
HH- Size	-0.000153	0.00184	0.000509	9.57e-05	0.00170	0.00136
	(0.00145)	(0.00138)	(0.00148)	(0.00135)	(0.00135)	(0.00152)
District FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Grade FE	YES	YES	YES	YES	YES	YES
Constant	-0.431***	-0.0736	-0.0709	1.439***	1.017***	0.710***
	-0.0779	-0.0703	-0.0751	(-0.124)	(-0.13)	(-0.149)
Observations	60,308	60,082	60,076	17156	17156	17059
R-squared	0.067	0.067	0.070	0.147	0.160	0.143

Notes: Table A4 (b) reports the Post-program difference using diff-in-diff estimates on the *children test performance* for Post=2015 vs Pre=12-to-2014(pooled). The data is from the ASER Household Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is surveyed 3 to 16 year's old child enrolled in government school. The dependent variable is the *test score normalized by grade*. Variable Monitoring is an interaction of *treated province* (KP) and Post (which is equal to 1 if year==2015 and 0 if year=2012 or 2014). Fixed Effect on individual grade, District and year applied in each regression. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively

Appendix VIII

Program Effect on Children Test Performance

Table-A4(3): Normalized Test Score [Post=2015 & 16 vs Pre==2012-14]

	Grade-0 to Grade-5			Grade-5 to Grade-10		
	Reading	Math	English	Reading	Math	English
DiD(treatment*Post)	0.0730	0.0221	0.0657	0.127**	0.0198	0.136**
	(0.0474)	(0.0440)	(0.0461)	(0.0506)	(0.0490)	(0.0618)
Post(=2015, Pre=2012-14)	-0.108**	-0.256***	-0.235***	-0.119**	-0.0562	0.126**
	(0.0477)	(0.0424)	(0.0450)	(0.0488)	(0.0488)	(0.0596)
Treatment(KP)	-0.0447	-0.263***	-0.115	-0.418***	-0.215**	-0.591***
	(0.0755)	(0.0735)	(0.0750)	(0.101)	(0.0923)	(0.102)
Child Age	0.0802***	0.0673***	0.0623***	-0.00715	-0.0115	0.00688
	(0.00685)	(0.00660)	(0.00618)	(0.00719)	(0.00742)	(0.00825)
Mother Highest Education	0.00250	0.000588	0.00310	-0.00231	-0.00128	0.00413
	(0.00216)	(0.00210)	(0.00215)	(0.00252)	(0.00232)	(0.00269)
Father Highest Education	0.00378***	0.00419***	0.00340**	0.00410***	0.00342**	-0.00337*
	(0.00140)	(0.00145)	(0.00142)	(0.00154)	(0.00154)	(0.00191)
House-ownership	0.0468**	0.0381*	0.00586	0.0230	0.0454	0.0613*
	(0.0226)	(0.0223)	(0.0243)	(0.0281)	(0.0301)	(0.0331)
Private Tutoring	0.277***	0.249***	0.266***	0.0151	-0.0589	0.112***
	(0.0371)	(0.0397)	(0.0399)	(0.0345)	(0.0402)	(0.0382)
Electricity Availability	0.0166	-0.00693	-0.0377	-0.0308	0.0510	0.0587
	(0.0374)	(0.0340)	(0.0359)	(0.0380)	(0.0422)	(0.0439)
Mobile service Availability	0.0333	0.00955	0.0560**	0.0712***	0.0277	0.0395
	(0.0217)	(0.0204)	(0.0222)	(0.0272)	(0.0261)	(0.0301)
TV availability	0.0205	0.0180	0.0189	-0.0140	-0.0524***	0.0408*
	(0.0169)	(0.0169)	(0.0169)	(0.0195)	(0.0186)	(0.0209)
house_condition	0.0282	0.0302	0.0302	0.0453	0.0550	0.0655*
	(0.0299)	(0.0304)	(0.0299)	(0.0361)	(0.0348)	(0.0391)
HH- Size	-0.00129	0.00193	0.000142	-0.000398	5.15e-05	-0.000119
	(0.00146)	(0.00140)	(0.00148)			
District FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Grade FE	YES	YES	YES	YES	YES	YES
Constant	-0.355***	-0.0691	-0.0226	1.435***	0.963***	0.814***
	-0.0724	-0.0672	-0.071	(0.107)	(0.114)	(0.129)
Observations	77724	77515	77509	21,744	21,744	21,373
R-squared	0.053	0.064	0.067	0.128	0.147	0.113

Notes: Table A4 (c) reports the Post-program difference using diff-in-diff estimates on the *children test performance* for Post=2015 & 2016(pooled) vs Pre=12-to-2014(pooled). The data is from the ASER Household Survey. Standard errors clustered at village level are shown in parentheses. The unit of observation is surveyed 3 to 16 year's old child enrolled in government school. The dependent variable is the *test score normalized by grade*. Variable Monitoring is an interaction of *treated province* (KP) and Post (which is equal to 1 if year==2015 or 2016 and 0 if year=2012 to 2014). Fixed Effect on individual grade, District and year applied in each regression. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively

Appendix IX

List of Districts in Khyber Pakhtunkwha and FATA

Federally Administered Tribal Areas(FATA)	*Khyber Pakhtunkhwa (KP)*	*Bordering*
FATA-Bannu	Abbottabad	No
FATA-Lakki Marwat	Bannu	YES
FATA-Peshawar	Battagram	No
FATA-Tank	Buner	No
Khyber Agency	Charsadda	YES
Mohmand Agency	Chitral	No
Orakzai Agency	D.I.Khan	YES
Bajaur Agency	Hangu	YES
FATA-Kohat	Haripur	No
Kurram Agency	Karak	YES
FATA-DIKhan	Kohat	YES
	Kohistan	No
	Lakki Marwat	YES
	Lower Dir	YES
	Malakand	YES
	Mansehra	No
	Mardan	YES
	Mardan-Urban	YES
	Nowshera	YES
	Peshawar	YES
	Peshawar - Urban	YES
	Shangla	No
	Swabi	No
	Swat	No
	Swat-Urban	No
	Tank	YES
	Tor Ghar	No
	Upper Dir	YES

Appendix X

ASER-Pakistan Children Test Procedure(Math Test)

How to test Arithmetic?

Start Here

Subtraction

- ☐ Show the child the subtraction problems. S/he can choose, if not you can point.
- ☐ Ask her/him to write and solve the problems. Observe to see if s/he does it in the correct written numerical form.
- ☐ Ask her/him to do a second one.

If s/he cannot do both subtraction problems, then give her/him the number recognition (10-99) task.

If s/he does both the subtraction problems correctly, ask her/him to do a division problem.

Number Recognition (10-99)

- ☐ Point one by one to at least 5 numbers. Child can also choose.
- ☐ Ask her/him to identify the numbers.
- ☐ If s/he can correctly identify at least 4 out of 5 numbers then mark her/him as a child who can 'recognize numbers from 10-99.

If s/he cannot recognize 4 out of 5 numbers from 10-99,then give her/him the number recognition 1-9 task.

Division (2 digit by 1 digit)

- ☐ Show the child the division problems. S/he can choose one out of the rest.
- ☐ Ask her/him to write and solve the problem.
- ☐ Observe and see if s/he is able to correctly solve the problem, and then mark her/him as a child who can do 'division'.
- ☐ If s/he is unable to solve a division problem correctly, mark her/him as a child who can do 'subtraction'.

Number Recognition (1-9)

- ☐ Point one by one to at least 5 numbers. Child can also choose.
- ☐ Ask her/him to identify numbers.
- ☐ If s/he can correctly identify at least 4 out of 5 numbers then mark her/him as a child who can 'recognize numbers from 1-9'
- ☐ If not then mark her/him at the level 'nothing'.

Source: ASER Pakistan, 2016