# Technology-based Schools Monitoring and Learning Outcomes: Evidence from Public Schools in Pakistan*

**Inayat ULLAH**
National University of Sciences & Technology (NUST), Pakistan (inayat@s3h.nust.edu.pk)

**Booyuel KIM**
Seoul National University (SNU), Republic of Korea

**Hee-Seung YANG**
Yonsei University, Republic of Korea

## ABSTRACT

We examine the effect of a technology-based monitoring of public schools on teachers' attendance and learning outcomes. Our identification is based on a large-scale monitoring program implemented in over 28,000 primary and secondary public schools in the Khyber Pakhtunkhwa (KP) province, Pakistan. We find ideal conditions for a natural experiment and apply *difference-in-difference*, *event study* and *instrumental variable* approaches to causally attribute changes in the learning outcomes of enrolled children to the monitoring program. We utilize seven rounds of a nationally representative annually conducted independent and systematically random survey called ASER-Pakistan.

Our findings suggest that technology-based monitoring has increased teachers' attendance by nearly 8 percentage points in the first year after the program. Despite a slight decrease in the second year after the intervention, the long-run effect of the program strongly persists with significant impact on learning outcomes of enrolled children. We find that enrolled children's standardized Reading, Math and English test score in the monitored schools has improved significantly by 0.08, 0.09 and 0.10 standard deviations points respectively at the lower (0~5) grades. Using exogenous program effect (obtained through *diff-in-diff*) as an instrument, we use observed variation in teachers' attendance to predict changes in the standardized test score of children. The 2SLS results are strikingly (nearly two-times) larger than fixed-effect OLS estimates. More specifically, on average, standardized Reading, Math and English test score in the monitored region has improved significantly by .33, .22 and .62 SD points respectively at the lower (0~5) grades. We also utilize the post-merger (of two regions) data to test the difference in observed outcomes when the monitoring program was extended to the comparison region. We also examine the program's effect on enrollment and school participation. Our results are robust on a number of alternate specifications, sub-samples, and falsification tests.

**Key Words:** schools monitoring, learning outcomes, governance.
**JEL Codes:** C55, D02, I21, I28, O38

---

* We thank the KDI school of Public Policy and Management for providing funding for this research project. This work heavily relies on the survey data annually published by the ASER-Pakistan for which we are grateful. We also thank Andrew Zeitlin and other participants at the SREE Conference 2019, Washington DC, Prashant Bharadwaj and other experts of 2018 KDI School – 3ie Conference on Impact Evaluation conference and participants of Asian and Australian Society of Labor Economics (AASLE) Conference, 2018 and other colleagues of the KDI School for their helpful comments and feedback.

## 1. INTRODUCTION

The populous countries in South Asia such as India and Pakistan are suffering from multi-dimensional poverty despite some countries such as Bangladesh, Nepal and Sri Lanka have shown significant improvements in human development index (Alkire et al., 2019, 2020; Asad et al., 2020). Albeit 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.). Recent national surveys on tracking progress in educational attainment in India finds over 55 percent of the children aged 7–12 cannot read a basic paragraph, and 40 percent cannot do simple subtraction despite increased school enrollment (ASER-India, 2021). According to Annual Status of Education Report (ASER-Pakistan) which reveals important trends each year covering over 255,000 children from 144 districts, Pakistan has suffered substantial learning loss during the year 2020 in the post-Covid-19 scenario. This is evident from its recent report showing that 55% children in grade 5 could read at story level in 2021 dipping from 59% in 2019. Similarly, 51% of children in grade 5 could do simple division as compared to 57% in 2019, while only 50% of grade 5 children could read at least English word (ASER, 2022). In similar circumstances, as Banerjee et al. (2013) suggest, 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).

The Sustainable Development Goals (SDGs) emphasize the need for more rigorous efforts through empirical findings that suggest feasible courses of actions to improve teaching quality and learning achievement in developing countries (UN SDGs, 2016). 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's learning outcomes (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 major factors to account for differences in 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. the principal) 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 records about schools and teachers' presence.

A number of reforms initiatives have been proposed for developing countries that can overcome learning inequality among 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 (2018) suggested 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. However, contextual evidence on community-based monitoring indicate less effectiveness of such programs particularly in rural areas (Banerjee and Duflo, 2005; Kremer & Vermeersch, 2005). One important for this is the awareness of the 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, and hence monitoring through them may not be effective. While much has been researched about significance of teacher's availability and school facilities, less is known about how to increase teachers' attendance especially in rural and remote areas in an effective and cost-efficient way and to make it conditional on learning performance.

This paper takes advantage of data that is panel at the district level and rotating panel at the village level collected by the Annual Status of Education Report (ASER-Pakistan), to attempt a natural experiment on a recently introduced public schools monitoring project by the KP government in Pakistan. We attempt to find a comparable administrative unit called FATA (which has recently been merged with the KP province) that was not affected by the policy yet shared similar socio-economic and demographic characteristics across the border with the treated administrative unit. We first apply a different-in-difference method and find causal effect of the program on the teachers' attendance ratio in public schools. We also apply event-study mechanism to show the causal effect of the program on teachers' attendance. Drawing on the first stage strong effect of the program, we then adopt an instrumental variable approach to causally attribute changes in the standardized test scores of children enrolled in lower grades 0~5 to the monitoring program.

Our diff-in-diff findings suggest that the technology-based monitoring has increased teachers' attendance by nearly 8 percentage points in the first year after the program. The program effect slightly decreases in the 2nd year of the intervention which coincides with the terrorists' attacks on public schools after the government launched a large-scale offensive against terrorism in the region. However, despite unstable law-and-order in the region, the program effect persists and rather bounces back after four years of the program implementation. Using exogenous program effect (obtained through diff-in-diff) as an instrument, we use variation in teachers' attendance to predict changes in the standardized test score of children enrolled in public schools. We find that enrolled children's standardized Reading, Math and English ability in the monitored region has improved significantly by 0.33, 0.22 and 0.62 standard deviations points respectively at the lower (0~5) grades. We also conduct a sub-set analysis of the program effect using gender, grades, and districts along the border and away from border of control region. Our results are robust on these alternate specifications, sub-samples, and falsification tests on the private school's data.

The results discussed in this research suggest a number of practical and methodological insights. First, school performance in terms of teachers' attendance and school facilities can be increased by increasing monitoring of schools using professionally trained monitors and adaptation of latest technology. Secondly, conditional on improvements in teachers' attendance, learning performance can be significantly enhanced by technology-based monitoring programs. Also, we find suggestive evidence to support the idea that improving schools' performance affects parents' response behavior in terms of sending children to schools. 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 learning outcomes in developing countries. Third, given the poor public education system in developing countries, monitoring of schools and teachers should be coupled with appropriate incentive and penalty mechanism in order to have a lasting impact on children performance. Finally, we argue that there is scope for the use of nationally representative large-scale surveys in conducting natural experiments for assessing the impact of education policy reforms introduced by sub-national governments in developing countries particularly in South Asia.

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. Section 3 provides a detailed description of the monitoring program and its implementation procedures. Experimental design, data and empirical approach is discussed in section 4 followed by results and discussion in section 5 and 6 respectively. The last section concludes.

## 2. LEARNING INEQUALITIES AND 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 latest projections, those less than 18 years old will account for about 50% of total population in 2030 (UNICEF, 2020; Burki, 2005). 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 the public sector, education departments in provinces were responsible for administering primary and secondary schools and colleges with public sector teachers' training schools and colleges. For several decades, the number of private schools was not much within the system of education. However, after the denationalization in 1990s, 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 (0~5 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 of 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 was 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 ( Ganimian et al., 2016 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). The poor education quality, both at primary and secondary level, is at the centre of many problems that the country faces in almost all its 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 grades, the net enrollment is even worst. For example, for middle schools, the net enrollment was hardly 40%

reflecting a significant dropout or non-enrollment during the middle school age group (11~15 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 indicators. 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 the education department with a significant number of teachers. For example, teachers make up around 75% of the 180,000 employees overall in elementary and secondary education departments. 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 Khyber Pakhtunkhwa Education Monitoring Authority (previously known as Independent Monitoring Unit (IMU)) is one such initiative taken by the provincial government of Khyber Pakhtunkhwa (KP) that aims at monitoring teachers and schools' performance through professionally trained monitors equipped with smart-phone/tablet facility (section 3 provide more details on IMU). According to an analysis on the IMU school level data in 2014, there was significant variation in teachers' attendance and student attendance rates at the primary and secondary level (CDPR, 2014). Also, large variations 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 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 gave additional allowances for female education supervisors to increase their inspections to schools. Similarly, to attract girls' enrollment, the KP government offered stipend program for secondary school 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 teachers' 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 teachers' attendance measurements or student performance in exams, suggesting the need for a more objective criteria for measuring teacher's performance.

[1] In 2014, 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)

[2] These low enrollment districts of KP included Hangu, Peshawar, Bannu, Lakki, D.I Khan, Shangla and Nowshera

## 3. PROGRAM DESCRIPTION

In struggle for educational quality enhancement and meeting one of the key the sustainable development goals, universal primary education, in 2014, the Khyber Pakhtunkhwa provincial government in collaboration with U.K's Department of International Development (DfID), took an important large-scale initiative called, Independent Monitoring Unit (IMU[3]) as part of the Khyber Pakhtunkhwa Education Sector Program (KESP) to credibly monitor and report all public-school indicators. Lunched formally in April 2014, the IMU's mandate was to monitor over 28000 schools with over 121,618 public schools' teachers across the province besides collection and compilation of data on basic schools' facilities such as electricity, boundary wall, toilets, and furniture etc., (Khan, 2019). Before IMU, school related data used to be collected through the head teachers or principals in the form of annual school census, a process that used to take several months to compile all information at the provincial level. Figure 1 shows the map of the Khyber Pakhtunkhwa province and erstwhile FATA by district.

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 its neighbor 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 (section 7.3 sheds light on this issue). 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[4] (at least one time per month to each school). The assignment of clusters rotates clock-wise on a monthly basis to minimize the possibility of relationship bias. For instance, the monitor who inspected cluster-A in January, will inspect cluster-B in February and so on. Each DCMA is 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 or tablets. The performance of DCMAs is in turn supervised by the District Monitoring Officers (DMOs) appointed one for each district across the province (H. Altaf[5], 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

[3] In 2019, the IMU was renamed as Education Monitoring Authority (AMA) after making it a regular government authority.
[4] Generally, a district is divided into 10 to 30 clusters(depending on the population of schools and gender)
[5] 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 protocols.

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 the penalty of low performing teachers, there is no such record of punishment, or any decision whatsoever published by the KP elementary and secondary education department.

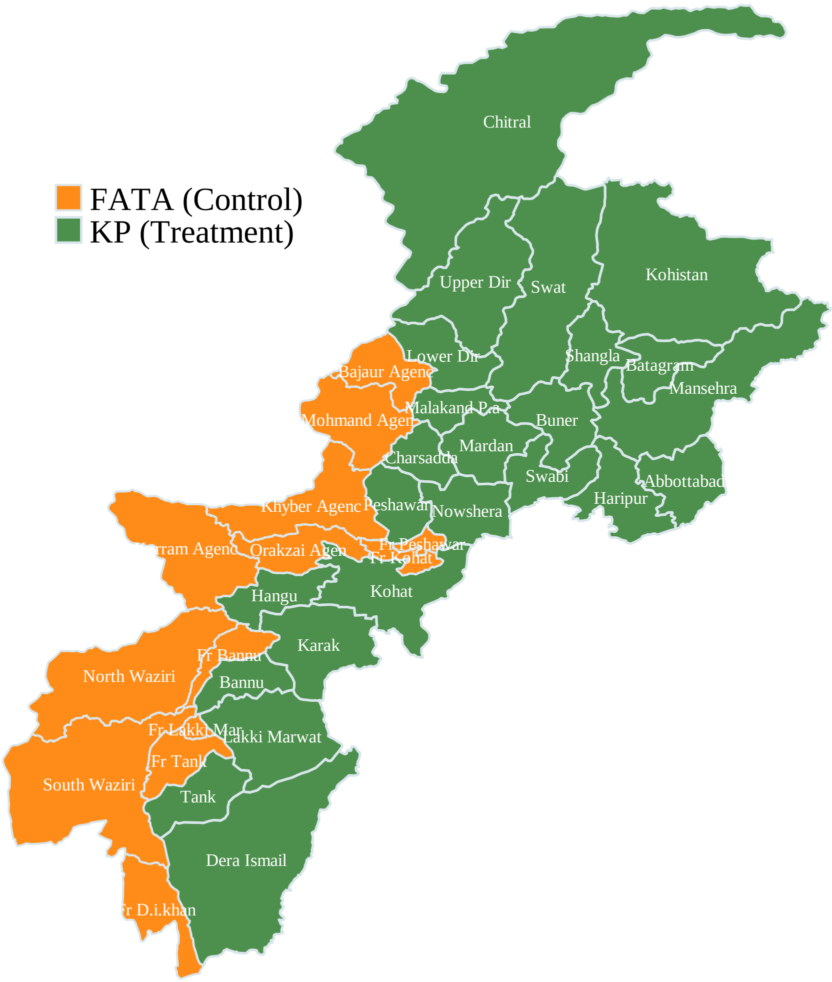



Fig 1. District wise map of KP and FATA

So far, the government reports suggest 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' learning outcomes. This research therefore examines whether the effect is causally attributable to the program.

## 4. EXPERIMENTAL DESIGN, DATA AND METHODS

In order to examine the effect of the IMU program on the school quality measured in the form of teachers' attendance, and learning outcomes of enrolled children, it is important to use a model that causally identify changes in the outcomes to the monitoring program. Literature on impact evaluation methodologies suggests several approaches 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 of interventions particularly in 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) found that reduced teachers' attendance significantly improved test score. An older example but relevant is the randomized controlled trial in Nicaragua, where radio instructions had significant impacts on pupils' math score (Jamison et al., 1981). In Kenya, randomized experiment of provision of school meals 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 were observed at a reasonably low cost (Banerjee et al., 2004). Also in India, a computer-assisted learning program found 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 estimating 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 quasi-experimental designs is the availability of good counterfactual- a control (untreated) group that satisfies all conditions for an ideal comparison. For example, in the context of school' monitoring programs, 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 differ before the interventions, studies attempt to mitigate this challenge by visualizing parallel trends. Recently, the two stage least square (2SLS) or instrumental variables (IV) have been adopted as an alternative approach to estimating the impact of education policy interventions. According to this approach, the exogenous variation caused by the program can be utilized as an instrument which must be strongly correlated with the endogenous variable and uncorrelated with the unobserved factors that might affect the outcome variable. 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 trails are considered the most effective means for assessing improvements in education quality caused by certain policies/reforms (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 collected 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 outcome variables used in our estimations.

### 4.1. DATA

Out main data source is the 6 rounds of country wide Annual Status of Education Reports (ASER-Pakistan) survey from 2012 to 2019. The ASER[6] is frequently cited in reference to teachers attendance, children enrollment and attendance, learning ability tests, private school enrollment, and other key education indicators by education researchers (Jones et al., 2014; Banerji et al., 2013; Zaka & Maheen, 2010; French, Kingdon, & others, 2010). ASER-Pakistan 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

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

national dataset of 81600 households and around 286,000 children per year. Recently the ASER-Pakistan has introduced tablet-based data collection which has improved the accuracy and effectiveness of the data. Table 1 provides year wise coverage of ASER data for KP province and FATA (the focus of our study). The ASER household survey includes learning 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 instance, 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 households are selected from each village. The ASER adopts two stage sampling designs. In the first stage 30 villages are selected using probability proportional to size (PPS) method. In the second stage, 20 household are selected[7] 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 the previous year, 10 new villages are added, and 10 villages are dropped from the previous year. In this way the ASER survey gives us a "rotating panel" of villages over the years. With regard to school selection, ASER choose at least one government school which is mandatory (could be more than one) and one private school from each selected village. The later ASER surveys also include urban regions in Pakistan.

TABLE 1: ASER SURVEY COVERAGE (2012 TO 2019) FOR KP AND FATA

| | **2012** | | **2013** | | **2014** | | **2015** | | **2016** | | **2018** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **KP** | **FATA** | **KP** | **FATA** | **KP** | **FATA** | **KP** | **FATA** | **KP** | **FATA** | **KP** | **FATA** |
| Districts | 23 | 9 | 25 | 9 | 27 | 9 | 26 | 11 | 24 | 9 | 25 | 9 |
| Villages | 688 | 270 | 763 | 265 | 789 | 270 | 769 | 330 | 704 | 270 | 688 | 253 |
| Households | 13,702 | 5,375 | 15,144 | 5,271 | 15,663 | 5,369 | 15,032 | 6,544 | 13,807 | 5,390 | 14229 | 7271 |
| Children | 41,003 | 18,529 | 46,877 | 18,722 | 49,473 | 18,743 | 46,045 | 22,890 | 41,804 | 17,753 | 41466 | 22078 |

*Notes:* The number of districts covered each year in KP, and FATA are not equal over the years because of two reasons. First, coverage in districts which were affected by military operation against terrorist such as Mohmand Agency, North Waziristan Agency and South Waziristan Agency was skipped by ASER. 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 districts each year ranges between 1 and 4. In 2019 however, the ASER has covered all districts in KP and FATA. There was no survey conducted in the year 2017.

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

7 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).

variables that might affect learning ability such as age, gender, enrollment status, school status (government or private), current grade, tuition facility, house-condition and ownership and parents' education etc. Table 2 (a) and (b) show the summary statistics of the 6 years ASER surveys annual data pooled from 2012 to 2019. 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 tests on all relevant factors affecting school-based and children-related outcome.

TABLE 2 (a)- SUMMARY OF ASER PUBLIC SCHOOLS SURVEY

| | 2012-18 (Pooled) | |
|---|---|---|
| Variables | KP (T) | FATA (C) |
| Primary school (1 to 5) | 0.67 | 0.78 |
| Middle schools (1 to 8) | 0.10 | 0.10 |
| High schools (1 to 10) | 0.22 | 0.12 |
| All other school types | 0.01 | 0.00 |
| Enrollment | 234.6 | 160.3 |
| Children presence on the day of visit | 202.3 | 135.0 |
| Appointed teachers | 7.3 | 4.7 |
| Present teachers on the day of visit | 6.4 | 4.1 |
| Student teacher ratio | 40.1 | 39.3 |
| Teachers-attendance ratio | 0.88 | 0.87 |
| Children attendance ratio | 0.84 | 0.81 |
| Laboratory available (yes=1) | 0.21 | 0.10 |
| Compute lab available (yes=1) | 0.07 | 0.04 |
| Internet availability (yes=1) | 0.05 | 0.02 |
| Observations (no. of schools surveyed) | 4306 | 1745 |

Notes: Data from ASER public schools' surveys from 2012-2019 pooled excluding 2017 for Khyber Pakhtunkhwa (KP) and FATA is summarized. Teachers' attendance ratio is calculated as number of teachers present on the day of visit / number of appointed teachers. Similarly, the student teacher ratio is calculated as Enrollment/Appointed Teachers in the surveyed school. Children attendance ratio is calculated as number of children present on the day of visit / total enrollment in the surveyed school.

TABLE 2 (b) - SUMMARY OF CHILDREN ENROLLED IN PUBLIC SCHOOLS (GRADE 0~5)

| Variables | KP | FATA |
|---|---|---|
| Demographic characteristics | | |
| Age (years) | 8.2 | 8.0 |
| Female | 0.45 | 0.45 |
| Test score (min 0 ~ max 5) | | |
| Reading | 2.97 | 2.88 |
| Math | 3.09 | 3.05 |
| English | 3.07 | 2.92 |
| Household socio-economic conditions | | |

| | | |
|---|---|---|
| Private tutoring | 0.02 | 0.01 |
| House ownership | 0.88 | 0.91 |
| Household size (total number of dependent children) | 6.0 | 5.70 |
| House construction strong | 0.25 | 0.18 |
| Electricity connection available | 0.88 | 0.84 |
| Mobile service available | 0.80 | 0.63 |
| TV available | 0.45 | 0.35 |
| Parents information | | |
| Father age (years) | 40.2 | 39.0 |
| Father ever attended the school | 0.54 | 0.49 |
| Father years of education | 5.70 | 4.60 |
| Mother age (years) | 35.0 | 34.8 |
| Mother ever attended the school | 0.25 | 0.12 |
| Mother years of education | 1.93 | 0.75 |
| Observations(#of Children surveyed children | 79313 | 38170 |

Notes: ASER- household survey (2012 to 2018) of children aged 3-to-16 years enrolled in grade 0-to-5 grades in public schools who performed the ASER learning test at their homes is summarized.

### 4.2. Rationale of the Comparison Group

We have several reasons to believe that FATA can be a valid comparison group. Firstly, districts located in FATA were completely not exposed to the monitoring program because this region was governed by the federal government directly until late 2018 when the government decided to merge this region with the KP government. Secondly, despite being governed by the federal government, the education system in the FATA region is similar to the KP province in terms of curriculum, examination system and school infrastructure including teachers appointing criteria etc. Third, FATA region has similar socio-economic characteristics to most of the KP province such as language, culture, traditions, and economic opportunities etc. FATA did have a different judicial system called Frontier Crime Regulations (FCR), a special set of laws applied since British India, however, these laws have directly nothing to do with the schooling particularly public schools located in FATA. Fourth, FATA shared the same educational boards with KP province which means that curriculum and examination schedules are the same as KP province. In section 4.4 we provide evidence of the common trend between FATA and KP before the monitoring program was launched in 2014. Hence, all schools located in FATA region are considered as control groups with no exposure to the program up until 2019. Regarding the pre- and post-intervention periods, 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 during the period from September-to-October each year. In this context, 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 in the same year of its implementation. 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. 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 consists of four years (2015 till 2018) in districts of the KP province. It might be argued that whether the anticipation effect (of the program) might have driven higher attendance of teachers[8]. To overcome this doubt and for robustness, we check both 2012-to-2013 and 2012-to-2014 as pre-program period (considering the year 2014 as an implementation or transition period) and find no significant change in the results. Figure 2 shows the timeline and ASER data collection from 2012 to 2018.

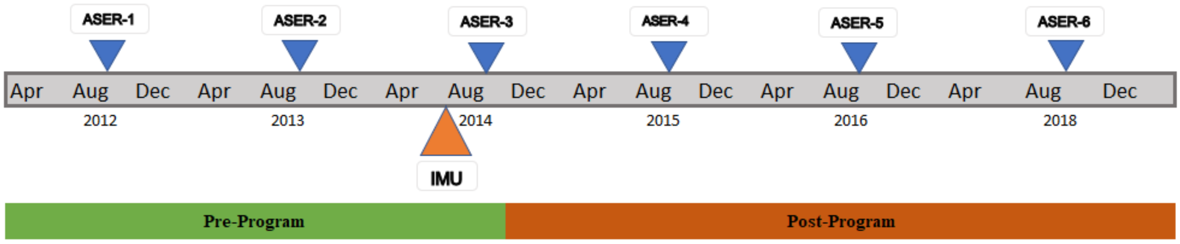



Fig 2. ASER survey timeline and implementation of IMU

This unique setting in which IMU was implemented in KP in 2014 and not in FATA until 2019, and the availability of ASER survey gives us an opportunity to apply differencing-in-differencing method to estimate the causal effect of intervention on outcomes of interests.

### 4.3. The Difference-in-Difference Method

Our main interest in the first stage is whether the intervention has increased teachers' attendance in the public schools in the KP province. We hold the following assumptions to carry out diff-in-diff analysis:

- The primary, and secondary education system in FATA is same as the KP due to the exam systems conducted by designated education boards[9].
- There was no significant difference in teachers' 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 etc.
- Our treatment period consists of four years (2015 till 2019) 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:

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

[8] Anticipation effect in this case could be the possible fear among teachers about the new system of monitoring and hence, it is possible that some teachers even before the monitoring had actually started, might have increased their attendance.

[9] Education boards are regulating 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 have jurisdiction both in KP and FATA districts. In total, there are 8 education boards in KP province.

where $TA_{idt}$ represents outcome on surveyed government school *i* in district $d$ in time *t*; $Monitoring_{idt}$ is a *differencing-in-differencing* interaction of districts in KP and post year *t* (e.g. $Monitoring_{idt}$=1 if school *i* belongs to district *d* of KP province & $t$ = 2015 or 2016 or 2018, $Monitoring_{idt}$=0 otherwise; $X_{idt}$ $is$ vector of school level controls; $\propto_d$ is the district fixed effect; $T_t$ is year fixed effect; and $\varepsilon_{idt}$ is an error term clustered at village (=school) level.

Similarly, we examine the direct effect of the program on normalized test score through 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 a *differencing-in-differencing* interaction of districts in KP 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 $\varepsilon_{igdt}$ is an error term clustered at village level. The ASER team conducts basic ability test at home of the surveyed children and record responses of children to each question starting from easy to the difficult question. These questions 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. 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 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 (0~5) to gauge the learning ability of enrolled children. We assign 1 to the easiest question and 5 to the difficult question of the ASER basic test. In this way, there are 5 marks of a child if he/she is able to answer all five questions. Following Banerjee & Duflo (2007), we subsequently normalize the test score by year and by grade using the mean and standard deviation of the control group in the pre-treatment period. The ASER bonus[10] test is conducted with only those children who are able to answer all basic five questions and are used as binary limited dependent variable (LDV). We estimate the LDV by passing the equation (2) through a probit link function as below:

$$\Pr(Y_{igdt} = 1) = \Phi(\beta_0 + \beta_1 Monitoring_{idt} + \beta_2 X_{igdt} + \propto_d + T_t + G_g + \varepsilon_{igdt}) \quad (3)$$

Where Φ is the probit link function with the cumulative distribution function of the standard normal distribution.

### 4.4. Pre-program Trend in KP and FATA

We take advantage of the pre-program data to test the common trend assumption, e.g., the outcome in treatment and control group would follow the same trend in the absence of the

[10] These are relatively difficult questions that are conducted with children who pass the basic test. For instance, if a child can read at least 2 out of the 4 sentences fluently, then he/she is asked to translate the sentence into his/her local language. If the child can translate the sentences correctly, he/she is marked as a "yes".

treatment. The results suggest that teachers' attendance on average did not vary significantly between treatment and control before the IMU was introduced. The same is true for children's test performance. Table 3 (a) & (b) present results on equation (1) and (2) using the pre-program data on our main outcome variables, teachers' 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, and fixed effects of districts and years, the difference between KP and FATA in terms of teachers' attendance ratio is not statistically significantly different in 2013 as well as in 2014.

TABLE 3 (a) -PARALLEL TREND TEST: PUBLIC SCHOOLS' TEACHERS ATTENDANCE

| | 2013 | | 2014 | | 2013+14 | |
|---|---|---|---|---|---|---|
| | *Teachers' Attendance Ratio* | | | | | |
| Diff-in-Diff (Treatment*Year) | 0.031 | 0.023 | -0.017 | -0.016 | 0.018 | 0.016 |
| | (0.022) | (0.022) | (0.017) | (0.017) | (0.018) | (0.019) |
| School's facilities controls | No | YES | NO | YES | NO | YES |
| Fixed effects (district, year) | YES | YES | YES | YES | YES | YES |
| R-squared | 0.063 | 0.074 | 0.051 | 0.059 | 0.051 | 0.059 |
| Observations | 1,981 | 1,982 | 3,014 | 3,015 | 3,014 | 3,015 |

*Notes*. The pre-program difference between KP province (treatment) and the control region FATA in terms of school outcome is reported. Dependent variable is the ratio of teachers present in school to the total appointed teachers estimated by the diff-in-diff interaction of treatment (KP) and year 2013 (column (1) & (2)), Year 2014 (column (3) & (4)), and year 2013+2014 (column (5) & (6). Due to district and year fixed effect applied in each regression, we do not report coefficient for treatment and posts separately. School related controls include variables school size (represented by the number of enrolled children), school teaching and training quality (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 wall, toilet, library, playground, laboratory, computer and internet.

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 for the normalized score at lower grades (0 to 5) is not statistically significant indicating similar performance of KP children with FATA children in learning outcomes. This is in line with previous studies that have documented lower performance of both KP province and FATA compared to the country-average in terms of basic learning tests at lower grades (ASER, 2014). 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.

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 Khyber Pakhtunkhwa Education Sector Programme (KESP) of which IMU is part of, and that mainly focused on teachers attendance, school infrastructure and oversight (Khan, 2019; CDPR, 2014).

TABLE 3 (b)-PARALLEL TREND TEST: LEARNING OUTCOMES

| | 2013 | | | 2014 | | | 2013+14 | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Standardized Test Score* | | | | | | | | |
| | Reading | Math | English | Reading | Math | English | Reading | Math | English |
| Diff-in-Diff (Treatment*Year) | -0.032 | -0.099 | -0.030 | -0.091 | 0.013 | -0.053 | -0.077 | -0.066 | -0.033 |
| | (0.067) | (0.064) | (0.064) | (0.064) | (0.057) | (0.055) | (0.058) | (0.058) | (0.058) |
| Child-related controls | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Fixed effects (district, year and grade) | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations | 38,151 | 38,151 | 38,151 | 57,861 | 57,862 | 57,862 | 57,862 | 57,862 | 57,862 |
| R-squared | 0.070 | 0.067 | 0.070 | 0.071 | 0.056 | 0.063 | 0.061 | 0.056 | 0.063 |

Notes: The pre-program difference in the standardized test scores between treatment(KP) province and control (FATA) region is reported. The dependent variable is the test score standardized by year and grade using the mean and SD of the comparison region (FATA) in the pre-treatment period (e.g., 2012~2014). Fixed effects of individual grade, district and year are included in all regressions. Child-related controls include age, gender, private tuition, mother and father highest education in years, house ownership, house condition, and availability of electricity, mobile and television facilities. The data is from the ASER Household Surveys 2012 to 2014. The unit of observation is surveyed 3 to 16 year's old child enrolled in government school from Grade-0 to grade-5. In all regressions, standard errors shown in parentheses are clustered at village level .

### 4.5. Instrumental Variable Approach

Working with natural experiments often requires us to hold strong assumptions about the absence of potential sources of endogeneity (Huntington-Klein, 2021; Wooldridge, 2013; Angrist, 2009). These include self-selection into treatment, unobserved time-varying confounders, unit heterogeneity, measurement error, reverse causality, and autocorrelation. For our expected effect on teachers' attendance, we do not have any issue with selection bias as the program was exogenously implemented in the KP province and the initial purpose of the program was increased monitoring of schools. Yet, for the direct effect of the program on children's learning score, we need to establish a channel through which the program might have transmitted its effect. In theory, the effect of increased monitoring of teachers can only be translated to learning outcomes if teachers not only come to school but also teach the content they are supposed to (Duflo, 2012; Duflo & Hana, 2005). In other words, if we expect learning outcomes to be improved due to the monitoring, we need to explain the channel through which such a program can affect the cognitive abilities of children with whom the test was conducted at their homes. In this context, we assume that the program has first increased teachers' attendance ratio, and through increased teachers' presence in the school, children have benefited in terms of learning the content. We thus model this indirect relationship through two-stage least square approach in which we utilize the first stage strength of relationship between program and teachers' attendance ratio (estimated through equation (1)), and then use the predicted endogenous variable (teachers' attendance ratio) to estimate effect on the standardized test score. Our second stage specification is therefore identified as below:

$$Y_{igdt} = \beta_0 + +\beta_1 \widehat{TA}_{idt} + \beta_2 X_{igdt} + \propto_d + T_t + G_g + \varepsilon_{igdt} \qquad (4)$$

Where other things are same as equation (2) except the diff-in-diff interaction term (monitoring) is replaced by the predicted teachers' attendance ratio (TA) in school i in district d in year t. District, grade and time fixed effects are applied while the error term is clustered at the village level. For this two-stage specification, we strictly assume that the program affects the children performance only through teachers' performance (instrumental relevance, e.g., COV (Monitoring, TA)$\neq$ 0) , and that there is no other channel through which the program might have affected learning outcomes other than teachers' attendance (exclusion restriction). In theory, the effect of increased monitoring of teachers can only be translated to learning outcomes if teachers not only come to school but also teach the content they are supposed to (Duflo, 2012; Duflo \& Hana, 2005). We assume that the program has first increased teachers' attendance ratio, and through increased teachers' presence in the school, children have benefited in terms of learning the content.

## 5. RESULTS

### 5.1. Overall Impact of the Monitoring Program

In table 4, we present the overall program effect represented by the coefficient of the interaction of treatment province with post period pooled from 2015 to 2018. The overall effect of the program on the ratio of present teachers to the total appointed teachers is 0.081 percentage points and is statistically significant after controlling for covariates and district and year fixed effects. 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 .081 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, teachers' attendance in public schools is likely to increase by nearly 9 % in the first four years of program implementation. 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 proximity to district administration offices etc., might affect the outcome variable. To overcome any such 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, school 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 a better teaching environment is well documented in education literature (Abhijit Banerjee & Duflo, 2006; Robert, 2005). We control all school-related facilities surveyed by ASER (e.g., availability of water, boundary wall, toilet, library, playground, laboratory, computer, and internet).

We turn to our second outcome of interest, children test performance, to examine the direct effect of the monitoring program on the test performed by enrolled children at home. Table 4 report the direct program effect on the normalized test score for lower grades (0~ 5) in column (2) to (4), for upper grades (6~8) in column (5) to (7). For simplicity purpose, we only report coefficients of the diff-in-diff interaction to show the differential effect of the treatment after the program. The direct impact of the program on the standardized test score is significant and positive. On average, a child enrolled in a public school under the IMU program is likely to score 0.084, 0.093 and 0.98 SD points higher as compared to the child enrolled in the comparison school. This effect is largely significant suggesting a causal difference in learning outcome between two regions. The program significantly increases the probability of upper-grade children to answer relatively difficult questions (.23, .37 & .41 respectively for reading, math and english). In the next section, we return to this impact more in detail.

We also produce event-study plots following Anderson et al. (2016) to examine the program effect using different post years and to visualize any difference during post-program periods. We construct interaction terms of KP with each year and use them as explanatory dummies in estimating the effect on outcome variables after controlling for district, year, school level and household level characteristics. Our event-study plot (Fig 3) shows the yearly effect of the program using the intervention year (e.g., 2014) as a reference year. The program effect decreases to nearly 0.025 percentage points after two years of the program but bounces upward in the year 2018 and 2019. As we discuss later, the 2016 data shows that some districts in the KP province had limited coverage by ASER in 2016 due to the ongoing military offensive against terrorists in different parts of the region. For instance, the average school coverage

TABLE 4- PROGRAM EFFECT ON SCHOOLS AND LEARNING OUTCOMES

| | School Outcome | Basic Learning Test (Grade 0~5) Standardized | | | Difficult Question ( Grade 6~8) (LDV)-Probit | | |
|---|---|---|---|---|---|---|---|
| | TA_ratio | Reading | Math | English | Reading | Math | English |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Diff-in-Diff (Treatment*Post) | 0.081*** | 0.084** | 0.093** | 0.098*** | 0.235** | 0.370*** | 0.413*** |
| | (0.013) | (0.040) | (0.038) | (0.037) | (0.102) | (0.120) | (0.102) |
| Child level and HH-related controls | - | YES | YES | YES | YES | YES | YES |
| School level controls | YES | - | - | - | - | - | - |
| District fixed effect | YES | YES | YES | YES | YES | YES | YES |
| Year fixed effect | YES | YES | YES | YES | YES | YES | YES |
| Grade fixed effect | - | YES | YES | YES | YES | YES | YES |
| Mean of Dep. Var (in FATA in pre-program period) | 0.89 | | | | | | |
| Observations | 6,022 | 115,696 | 115,696 | 115,696 | 22,199 | 22,199 | 22,199 |
| R-squared | 0.053 | 0.054 | 0.046 | 0.058 | | | |

Note: For teachers attendance ratio in column (1), ASER public schools' dataset is used while for children standardized test score, ASER HH children dataset is used. The diff-in-diff estimator is an interaction of the treatment (KP) and post-period (2015, 2016 & 2018 pooled). The dependent variable in column 1 is the teachers' attendance ratio in public schools. The dependent variable in column (1) is the teachers' attendance ratio in public schools. Dependent variable in column (2) to (4) is the basic test score normalized using the mean and standard deviation of the comparison group FATA in the pre-program period (2012-2014). Dependent variable in column (5) to (7) is the binary outcome (1,0) of a difficult question attempted by enrolled children. This test was not performed during survey year 2012 and 2013. For difficult questions in column (5) to (7), we run a probit regression of a binary dependent variable that shows whether a child successfully answers the difficult questions after performing the basic questions. School related controls include variables school size (represented by the number of enrolled children), school teaching and training quality (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 wall, toilet, library, playground, laboratory, computer and internet. Child and HH controls include, age, gender, parents' education, private tuition, house ownership, household size, house condition, and availability of electricity, mobile and television facilities. District and year fixed effect are applied in all regressions. Standard errors are corrected for clustering at the village level. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

per district is 30, while for some districts in the KP province the ASER reported data for as few as 15 schools per district. Using data from Global Terrorism Database (GTD), we provide a detail analysis of the terrorism wave during the period between 2012 and 2018 in section 5.3 and conclude that due to large number of terrorist attacks on educational institutions in the KP province, our results might be suffering from potential downward bias. The impact of the program increases back to 0.08 percentage points in 2019. This effect is larger given the mean value of the dependent variable (.87 or 87% teachers attendance). Table 4 includes 2014 years part of the pre-program period. Although albeit low in 2016, the average program effect is still significant and persistent across all post years.

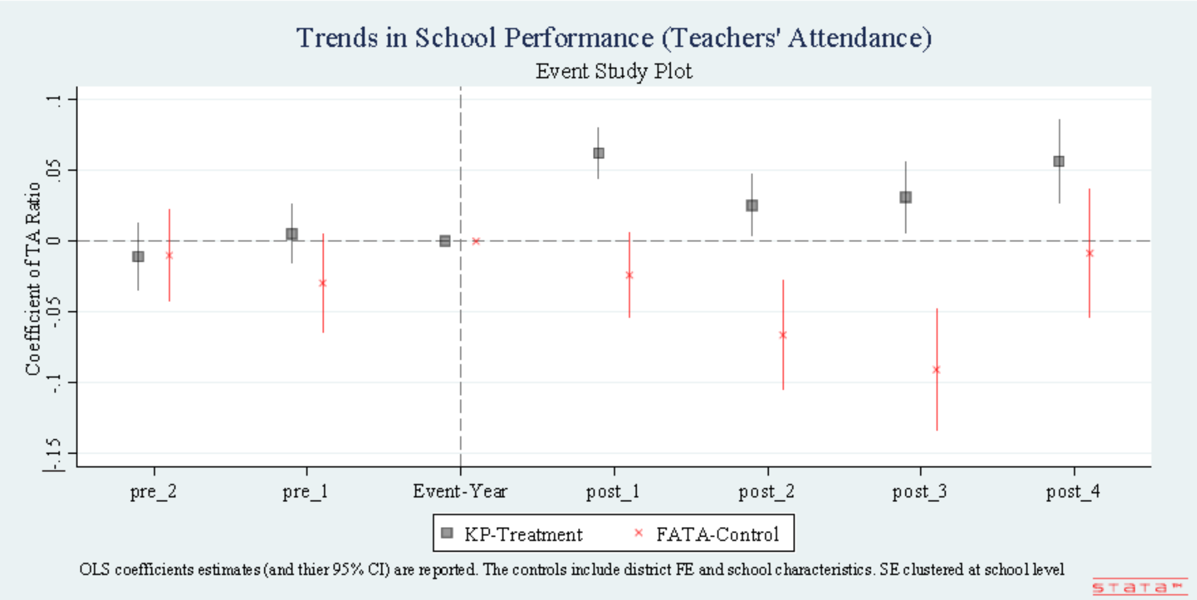



Figure 3. Program effect on teachers' attendance (*Event-study Design*)

### 5.2. Program Effect on Learning Outcomes

Even if monitoring increased teacher's presence in schools, we still need to examine whether increased teachers' presence has impacted learning outcomes on average. In other words, whether teachers teach once they decide to be in the schools is the question of our interest in this section. Despite our observed direct effect on learning outcome, it is important to provide a causal link through which one can explain enhancement in learning outcomes attributable to teachers' presence in schools. 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 et al., 2007; Croke, 2014; Jackson, 2009; Raikes, 2016). We therefore control individual child-specific characteristics, parents' education, and household characteristics along with district fixed effect and year fixed effect.

Table 5 presents the grade wise direct effect of the monitoring program on lower and upper grade children using basic test score and difficult test questions respectively. We can observe that the combined effect of the program is positive and significant at a 1 % level for grade 0 to grade 3. The magnitude of the program is even bigger. Overall children belonging to the treatment provinces on average score 0.14, 0.12 and 0.11 SD points higher in basic reading, math and english test questions respectively at grade 0 to grade 1, while .14 SD points higher at grade 2 and grade 3 compared to those belonging to the control region. We observe a significantly positive effect of the IMU program on the predicted probability of answering a difficult question by children enrolled in upper grades in the KP province compared to the children in the control region. These effects are significant after all observed child-related, parents-related, household-related variables are controlled for while district, year and grade-fixed effect applied in each regression. Please note that the ASER basic test questions are generally designed for lower grade children and hence we do not observe any significant effect on the grade 4 and 5 children. Similarly, the difficult question designed by ASER is meant to test only those who are able to perform in the basic test question. We also present impact of the program on learning outcomes using the event-study design shown in Appendix II for basic reading, math and english. The event study results are in line with our expectation as the program was focused on teachers' attendance and not directly on learning outcomes. A slightly better performance of the KP children compared to the FATA region in the year 2016 through 2018 indicates the existence of indirect effect of the program on children performance. The parallel trend before the program implementation year (e.g., 2014) in terms teachers' attendance ratio and children learning outcomes is also evident from the event study plots. Though statistically not significant for highest grades, the program effect is positive and significant for grade 6 and grade 7 enrolled children. 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[11] test questions. In estimating results in table 4, table 5 and table 6, we only include children that are currently enrolled in government schools and belonged to the same village in which the government school was surveyed.

We now turn to our 2SLS effect of program on the learning outcomes through teachers 'attendance as endogenous variable in equation (3). Table 6 reports the first and second stage (2SLS) program effect on the standardized test score. The results in the 2nd stage column for children outcome show the Local Average Treated Effect (LATE) on the treated children. Conditional on the first stage impact, if a child belongs to the school that has been positively affected by the monitoring program, then his/her reading, math and english score is likely to be increased by .33, .22 and .62 standard deviation point respectively. We use the standardized form of the teachers' attendance ratio and test performance. On a standardized scale, with mean 0 and standard deviation 1, this effect is substantial and statistically significant with 1% level. Compared with the reduced form results, our 2SLS results are larger and provide us with more causal interpretation of the program effect on learning outcomes.

Our 2SLS results are bigger than the OLS diff-in-diff estimates. There might be a few reasons behind this larger effect. Firstly, an omitted variable that could be negatively correlated with teachers' attendance ratio. For instance, the ongoing military operations against the militant organizations and resulting retaliation by terrorists is likely to decrease the teachers' presence in school specially when terrorists in the region announced attacks on public schools. We verify this possibility using the GTD data on attacks on education institutions in KP and FATA.

[11] ASER-Pakistan follows a similar procedure of conducting basic test as ASER-India.

TABLE 5-PROGRAM EFFECT ON LEARNING OUTCOMES BY GRADE

| | *Basic Test Score* | | | *Difficult Questions(LDV)* | |
|---|---|---|---|---|---|
| | grade-0~1 | grade 2~3 | grade 4~5 | grade 6~8 | grade-9~10 |
| | (1) | (2) | (3) | (4) | (5) |
| OV: *Reading* | | | | | |
| Diff-in-Diff [KP]×Post [2015-18] | 0.142*** | 0.141*** | 0.036 | 0.235** | 0.148 |
| | (0.051) | (0.053) | (0.057) | (0.102) | (0.134) |
| OV: *Math* | | | | | |
| Diff-in-Diff [KP]×Post [2015-18] | 0.118** | 0.140*** | 0.090 | 0.370*** | 0.263 |
| | (0.051) | (0.049) | (0.057) | (0.120) | (0.165) |
| OV: *English* | | | | | |
| Diff-in-Diff [KP]×Post [2015-18] | 0.113** | 0.142*** | 0.048 | 0.413*** | 0.178 |
| | (0.050) | (0.049) | (0.053) | (0.102) | (0.138) |
| Child and HH Controls | Yes | Yes | Yes | Yes | Yes |
| Fixed Effects (District, Year and Grade) | Yes | Yes | Yes | Yes | Yes |
| Observations | 35,648 | 43,858 | 36,190 | 22,199 | 10,269 |

Notes: Grade-wise program effect on the learning outcomes for treatment (KP) province and control (FATA) using diff-in-diff specification is reported. The pre-program years are 2012, 2013 & 2014 while the post-program period is 2015-to-2018. Dependent variable of the basic test score in column (1) to (3) is the basic test score standardized using the mean and standard deviation of the control region in pre-program period. Child and HH related controls include age, gender, private tuition, mother and father highest education in years, house ownership, house size and condition, and availability of electricity, mobile and television facilities. For difficult questions in column (4) & (5), we run a probit regression of a binary dependent variable that shows whether a child successfully answers the difficult questions after performing the basic questions. The coefficients for probit regression represent the predicted probabilities of passing the difficult question test. Since data on difficult question is not available for 2012 & 2013, therefore we use 2014 as a pre-program period for our probit regressions. Fixed Effect of individual grade, district and year is applied in each all regressions. 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. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

TABLE 6-PROGRAM EFFECT: INSTRUMENTAL VARIABLE APPROACH

| | 1st Stage | (2nd Stage) | | |
|---|---|---|---|---|
| | TA Ratio | Reading | Math | English |
| | (1) | (2) | (3) | (4) |
| Diff-in-Diff (Treatment*Post) | 0.241*** | | | |
| | ( 0.016) | | | |
| Teachers' attendance ratio (endogenous var) | | 0.335*** | 0.224*** | 0.620*** |
| | | (0.074) | (0.071) | (0.080) |
| School level controls | | YES | YES | YES |
| Child related and HH level controls | | YES | YES | YES |
| Grade fixed effect | | YES | YES | YES |
| Fixed effect (district, year) | YES | YES | YES | YES |
| F-stat value | 204.9 | | | |
| Observations | 99,443 | 99,443 | 99,443 | 99,443 |

Note: Two-stage least square regression results after merging the public-school survey data with household survey is reported. The program effect represented by the Diff-in-Diff is used as an instrument to exogenously affect the TA-ratio (endogenous). For ease of interpretation, the TA ratio is also normalized using the mean and SD of the comparison group in the pre-treatment period. The outcome variable in the 2nd stage is the test score normalized by the year and grade using the mean and SD of the control group FATA in the pre-treatment period. Controls include all those mentioned in table 4. Robust standard errors are shown in parentheses.

Secondly, possible measurement error in TA ratio during the survey. The survey officials ask the head of school about the total number of appointed teachers in the school and the number of those present on the day of visit. Since the IV estimate is unaffected by the measurement error, they tend to be larger than the OLS estimates. Thirdly, it is possible that the IV estimate are larger than the OLS estimate because IV is estimating the Local Average Treatment Effect (LATE) while OLS is estimating the ATE over the entire population. The program effect in our case is likely to have shifted the behavior of a subgroup of children for whom the teachers' attendance is larger than average. In other words, the IV estimate is the effect of increasing teachers' attendance for schools where teachers' attendance ratio was lower (note that average attendance ratio is 87) on learning outcomes, while the OLS estimate describes the average difference in learning outcomes for those schools whose teachers' attendance ratio differs by treatment region only.

### 5.3. Terrorists Attacks on Educational Institutions in KP and FATA

Education has been a casualty throughout the conflict between government and militant groups in Pakistan (Javeid et al., 2022). Notably, in KP province where the TTP controlled Swat Valley 2009, and in erstwhile FATA, non-state actors violently targeted the state, women's rights, and girls' education. According to GTD data, between 2012 and 2018, there were 332 attacks on educational institutions[12] that included attacks on schools, teachers, education buildings and other facilities. From 2012 to 2016, there was a gradual decline in these attacks due to the ongoing military operations, however, in some sensitive districts13, there was a surge in attacks on educational institutions in years 2016 and 2017 specially in the KP province.

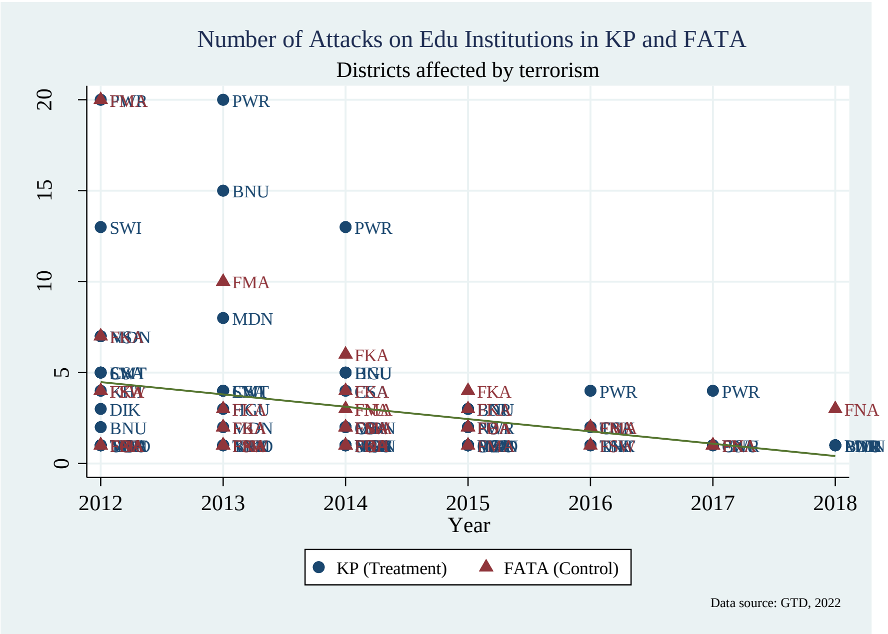



**Figure 4:** Terrorist Attacks on Educational Institutions in KP and FATA (Source: GTD, 2022)

Due to its huge population, vicinity to Afghanistan Border, and vulnerability to militants' attacks, district Peshawar experienced the highest number of attacks on educational institutions in the year 2016 and 2017. The district level average of the number of attacks on educational presented in figure 5 shows that KP province suffered more attacks than FATA in years 2016 and 2017. This surge in attacks might have driven the program effect downward in these two years. Further analysis of this surge in attacks on educational institutions shows that school, universities, and educational building were the major targets of terrorists between the period 2012 to 2018. Also, over 95% of attacks targeted schools including girls' and boys' school. Data form the Global Coalition to Protect Education from Attack (GCPEA) shows that approximately one-third of these attacks were reported to have affected girls' schools. In these reports however, it should be noted that not all schools can be differentiated whether they were boys' schools, girls' school, or both. In many cases, these attacks damaged or destroyed schooling infrastructure and killed students and/or teachers.

[12] Includes attacks on schools, universities, educational infrastructure, teachers, professor, instructors, and other personnel of educational institutions.
[13] These districts include Peshawar, Swat, North Waziristan Agency , Mohmand Agency, and Khyber Agency. The timing of military operations varies depending on the activity of terrorist groups and political consensus.

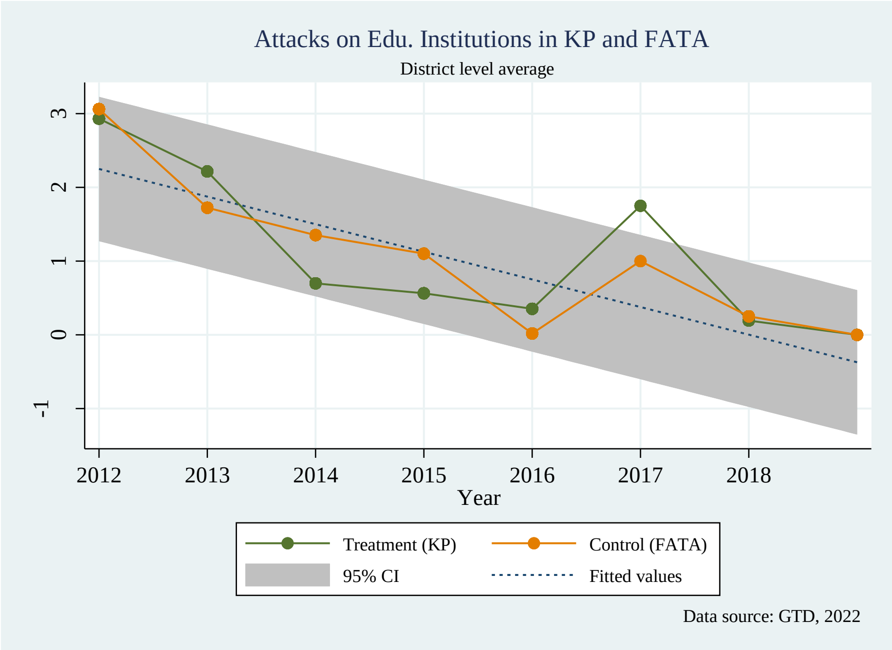



**Figure 5**. Average terrorist attacks on educational institutions per district in KP and FATA

### 5.4. Falsification Tests on Post-Merger and Private Schools Data

FATA was officially merged with the KP province following 25th Amendment the national constitutions of the Islamic Republic of Pakistan on 31st May 2018. Although, transition of merger takes time and not every institution can immediately start functioning in the post-merger era, the IMU program however was considered a priority by the KP government to be expanded to the FATA region in the year 2019. Utilizing the ASER 2019 survey, we conduct a falsification test on the program effect using 2019 as a post-treatment period and 2012 to 2014 as pre-program period. The results presented in appendix table A1 show no difference between the KP and FATA region in terms of teachers' attendance while FATA performs better than KP in terms of standardized test score. The reason for higher learning outcomes in FATA in the post-merger period may be the newness of the program as this was the first year of implementation.

The ASER also conducts a private schools survey alongside public school's survey. In each surveyed village (if there exists) at least one private school is surveyed and same information including appointed teachers, present teachers, and other school indicators. We run the same diff-in-diff specification on the private school's survey data to examine the placebo effect. The results presented in table 7 column (1) show no statistically significant difference between the KP and FATA region in terms of teachers' attendance in private schools. Similarly, using the HH survey of children enrolled in private schools, we run the same diff-in-diff specification to estimate program effect on their test score. Table 7 column (2) to (4) shows no statistically significant difference in the standardized test score of children enrolled in private schools between KP and FATA suggesting the validity of the program effect on public schools keeping in view the fact that the program focused only on public schools and not on private schools.

TABLE 7-FALSIFICATION ANALYSIS (SURVEY OF PRIVATE SCHOOLS AND CHILDREN ENROLLED IN PRIVATE SCHOOLS)

| | School Outcome | Test Score(Grade 0 to 5) | | |
|---|---|---|---|---|
| | TA Ratio | Reading | Math | English |
| | (1) | (2) | (3) | (4) |
| Diff-in-Diff (Treatment*Post) | -0.0272 | 0.103 | 0.089 | 0.156 |
| | (0.0248) | (0.090) | (0.091) | (0.103) |
| Other HH characteristics | - | YES | YES | YES |
| School level controls | YES | - | - | - |
| District fixed effect | YES | YES | YES | YES |
| Year fixed effect | YES | YES | YES | YES |
| Grade fixed effect | - | YES | YES | YES |
| Mean of the Dep. Var | .86 | | | |
| Observations | 2,089 | 37,491 | 37,491 | 37,491 |
| R-squared | 0.051 | 0.076 | 0.052 | 0.073 |

Note: Diff-in-diff estimates of our main specification using the ASER private school's data (column (1)) and ASER data on children enrolled in private schools (last three columns) is reported. Dependent variable in column (1) is teachers attendance ratio measured as a ratio of present teachers to the total appointed teachers in a surveyed private school. Dependent variable in the last three columns is the test score normalized by year and by grade using the mean and SD of the control group FATA in the pre-treatment period. School level controls include school teaching quality and school training quality, school facilities including dummies on availability of water, boundary, toilet, library, playground, laboratory, computer and internet and school Size. Child related and HH controls include age, gender private tuition, house ownership, house' condition, and availability of electricity, mobile and television facilities. In all regressions, district and year fixed effect is applied while for learning outcomes, grade fixed effect is also applied. Standard errors are corrected for clustering at the village level.

### 5.5. Impact on 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 from low enrollment and high dropout rate at primary and middle level schooling (Gouleta, 2015, ASER, 2018). A review by the International growth Centre (ICG) shows in KP province 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 grades, the net enrollment is even worst. For example, for middle schools, the net enrollment was hardly 40% reflecting a significant dropout or non-enrollment during the middle school age group (11 to 15 years).

To investigate the overall direct effect of the monitoring program on the enrollment status of children surveyed at home, we analyze ASER household survey data from 2012 to 2018. The ASER household survey includes a variable on the status of children of age 5 to 16 asking whether they are enrolled in schools or not with further identification of enrollment in public or private school. We drop all those children enrolled in private school, madrassas, or any other school to obtain reduced sample of children either enrolled in public schools or not enrolled.

We attempt our diff-in-diff model for post-program year as 2015, 2016 and 2018 to see the four years post program effect. Results reported in table 8 are suggestive of the positive direct effect of monitoring program on gross enrollment of up to 6 years old children in year 2015 and 2018 while there is no significant effect in 2016. The overall difference in enrollment status between children in KP and FATA suggests that being in the KP province increase the probability of six years old child to be enrolled in public schools by 0.15. The lower number of observations in 2016 indicates the limited coverage of ASER Survey in 2016 and is suggestive of the possible effect of terrorism wave on enrollment. Gross enrollment mainly depends on supply side factors such as school density and demand-side factors such as awareness campaigns run by either government or non-government organizations. For instance, if the government schools (e.g., per village) increase, it might increase the gross enrollment per village. 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). 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-8: PROGRAM EFFECT ON CHILDREN ENROLLMENT STATUS

| Dep. Var: Enrollment Status [0,1] | Post=2015 | Post=2016 | Post=2018 | Post=2015~18 |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| KP *Year | 0.193*** | -0.114 | 0.165** | 0.158*** |
| | (0.072) | (0.107) | (0.075) | (0.058) |
| Child related and HH controls | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Observations | 52,561 | 41,418 | 53,427 | 71,200 |

Notes: Probit coefficients of the post-program difference between the KP and FATA regions is reported. The pre period in all columns is 2012 to 2014 pooled. The dependent variable is a binary indicating whether a child is enrolled in public school and zero otherwise. The sample does not include children that are enrolled in private or other schools. District and year fixed effect, controls for child and HH related characteristics included in all regressions. Standard errors clustered at village level are shown in parentheses. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively

### 5.6. Robustness Check

TABLE 9-PROGRAM EFFECT BY THE TYPE OF SCHOOL AND BY GENDER

| | School Outcome | Basic Test Score (Grade 0~5) | | | Difficult Test (Grade 6~8) Probit | | |
|---|---|---|---|---|---|---|---|
| | TA Ratio | Reading | Math | English | Reading | Math | English |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A. Girls* | | | | | | | |
| Diff-in-Diff (Treatment*Post) | 0.085*** | 0.115** | 0.076* | 0.104** | 0.528*** | 0.626*** | 0.729*** |
| | (0.015) | (0.047) | (0.044) | (0.044) | (0.170) | (0.205) | (0.170) |
| Fixed Effects (District, Year) | YES | YES | YES | YES | YES | YES | YES |
| Grade Fixed Effect | - | YES | YES | YES | YES | YES | YES |
| Controls | YES | YES | YES | YES | YES | YES | YES |
| Observations | 4,052 | 51,000 | 51,000 | 51,000 | 6,196 | 6,196 | 6,196 |
| *Panel B. Boys* | | | | | | | |
| Diff-in-Diff (Treatment*Post) | 0.092** | 0.104** | 0.143*** | 0.123*** | 0.153 | 0.292** | 0.349*** |
| | (0.040) | (0.047) | (0.045) | (0.045) | (0.108) | (0.125) | (0.108) |
| Fixed Effects (District, Year) | YES | YES | YES | YES | YES | YES | YES |
| Grade Fixed Effect | - | YES | YES | YES | YES | YES | YES |
| Controls | YES | YES | YES | YES | YES | YES | YES |
| Observations | 1,275 | 64,696 | 64,696 | 64,696 | 16,003 | 16,003 | 16,003 |
| *Panel B. Boys and Girls (both)* | | | | | | | |
| Diff-in-Diff (Treatment*Post) | 0.076* | 0.084** | 0.093** | 0.098*** | 0.235** | 0.370*** | 0.413*** |
| | -0.045 | (0.040) | (0.038) | (0.037) | (0.102) | (0.120) | (0.102) |
| Fixed Effects (District, Year) | YES | YES | YES | YES | YES | YES | YES |
| Grade Fixed Effect | - | YES | YES | YES | YES | YES | YES |
| Controls | YES | YES | YES | YES | YES | YES | YES |
| Observations | 695 | 115,696 | 115,696 | 115,696 | 22,199 | 22,199 | 22,199 |

*Note*: For teachers attendance ratio in column (1), ASER public schools' dataset is used while for children standardized test score, ASER Children dataset is used. For school outcomes in panel A, panel B and panel C column (1), we use data on girls' public schools , boys' public schools and mixed public schools respectively. Column (2) to (4), panel A, panel B and panel C show sample of enrolled girls, boys and combined sample (boys and girls) respectively. The dependent variable in column (1) is the teachers' attendance ratio in public schools. Dependent variable in column (2) to (4) is the basic test score normalized using the mean and standard deviation of the comparison group FATA in the pre-program period (2012-2014). Dependent variable in column (5) to (7) is the binary outcome (1,0) of a difficult question asked.

TABLE 10- PROGRAM EFFECT ON SCHOOL AND CHILDREN'S LEARNING OUTCOME USING BORDERING[14] DISTRICTS

| | School Outcome | Basic Learning Test (Grade 0~5) Standardized | | | Difficult Question ( Grade 6~8) (LDV) | | |
|---|---|---|---|---|---|---|---|
| | TA_ratio | Reading | Math | English | Reading | Math | English |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Diff-in-Diff (Treatment*Post) | 0.089*** | 0.153*** | 0.125*** | 0.105** | 0.245** | 0.286** | 0.284** |
| | (0.014) | (0.048) | (0.046) | (0.045) | (0.113) | (0.132) | (0.111) |
| Child related and HH Controls | - | YES | YES | YES | YES | YES | YES |
| School Level Controls | YES | - | - | - | - | - | - |
| District Fixed Effect | YES | YES | YES | YES | YES | YES | YES |
| Year Fixed Effect | YES | YES | YES | YES | YES | YES | YES |
| Grade Fixed Effect | - | YES | YES | YES | YES | YES | YES |
| Observations | 4,776 | 82,869 | 82,869 | 82,869 | 14,503 | 14,503 | 14,503 |
| R-squared | 0.049 | 0.048 | 0.047 | 0.059 | | | |

Note: For teachers attendance ratio in column (1), ASER public schools' dataset is used while for children standardized test score, ASER children dataset is used. The diff-in-diff estimator is an interaction of the treatment (districts in KP sharing geographical borders with control region) and post-period (2015, 2016 & 2018 pooled). Dependent variable in column 1 is the teachers' attendance ratio in public schools. Dependent variable in column (2) to (4) is the basic test score normalized using the mean and standard deviation of the comparison group FATA in the pre-program period (2012-2014). Dependent variable in column (5) to (7) is the binary outcome (1,0) of a difficult question attempted by enrolled children. This test was not performed during survey in year 2012 and 2013. For difficult questions in column (5) to (7), we run a probit regression of a binary dependent variable that shows whether a child successfully answers the difficult questions after performing the basic questions. Child related and HH controls include age, gender, parent education, private tuition, house ownership, household size, house condition, and availability of electricity, mobile and television facilities. District and year fixed effect are applied in all regressions. Standard errors are corrected for clustering at the village level. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

[14] Districts that do not share geographical border with FATA are excluded from the sample. These excluded districts include, Abbottabad, Battagram, Chitral, Haripur, Kohistan, Manshera, Buner, Shangla and Torghar.

## 6. DISCUSSION

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 higher grades such as math, english and science courses of 9th or 10th grade. Previous studies support the idea that developing countries such as 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 $9^{th}$ and $10^{th}$, 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 teachers' 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 table A2 suggest a slight increase (overall 0.015, with 10 % significance level) in children attendance ratio from 2016 to 2018. The program effect on children attendance is not significant in the year 2016. In either of our specifications, children attendance appears to be less affected (or unaffected) in the years immediately after the program. This is surprising as previous 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.

There could be several reasons for the decreasing effect of the program in the year 2016. Firstly, the surge in terrorists' attacks on educational institutions in years 2016 and 2017. Pakistan has been battling terrorism over the last two decades with various militant groups carrying out attacks across the country targeting civilians, security forces, and government officials and educational institutions (Malik et al., 2019; Javeid et al., 2022; Khan & Seltzer, 2016; Muhammad, 2018). Between 2001 and 2013, the total number of terrorist incidents took place in Pakistan were 13,721 while during the 2001–2005 period, only 523 terrorist incidents occurred (Khan, 2016). This wave of terrorism peaked during 2009-10 when different terrorist groups including the "Tehrik-i-Taliban Pakistan (TTP)" started large-scale attacks on military and civilian population.

The KP and erstwhile FATA have been the most affected regions due to the presence of TTP, the most notorious militant group and its shared border with Afghanistan. On December 16, 2014, unknown terrorists penetrated a public school (APS) run by Pakistan Army in a heavily fortified military area in the capital of KP province, Peshawar. In this tragic incident, 140 persons including a large number of young students were brutally killed. This massacre served as an eye-opener for Pakistani society to raise questions about the ability of terrorist groups and their affiliates to strike soft targets especially educational institutions and also about the failure of the military to protect their own schools. The entire political spectrum reached a consensus on a new National Action Plan (NAP), an ambitious list of objectives to curb terrorism in the country (Faiz et al., 2017). Over the following years, several targeted military operations were conducted in KP and FATA in districts where TTP and their affiliates were active. These operations also sparked retaliations from militant groups who also targeted civilian population and soft targets such as educational institutions (Malik et al., 2019; Naseem et al., 2019; Khan, 2012). After selective military operations conducted by Pakistan Army in different parts of the country, there has been a continuous decline in the number of terrorist incidents and resulting casualties. According to South Asian Terrorism Portal Index (SATP), terrorism in Pakistan has declined by 89% in 2017 since its peak years in 2009. As in shown in figure 5 , attacks on schools and other education related facilities were larger in KP as per the GTD data specially when we take into account the number of districts in each group. Secondly, the expected penalty (or reward) based on the IMU data was not strictly observed specially in the first year of its implementation despite absenteeism reported by IMU. Also, as other studies observe, there could be a learning effect (Banerjee & Duflo, 2006), from the perspective of teachers such that teachers might have learnt tactics of shirking by establishing contacts with local people who might have collaborated and intimated teachers once they see 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 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 slightly, the overall impact of IMU program appears to bring immediate improvement in the teachers' attendance over a large area. We check the robustness of our model on various sub-samples of school levels such school types, gender of children (table 9), and a reduced sample of districts bordering with FATA (table 10) as well district not bordering with FATA (appendix table A3). There are twelve districts in KP province which share border with FATA region. The program effect for bother teachers' attendance and learning outcomes are consistent and statistically significant across these subsamples. Our results on the children's test score provide evidence in support of the idea that absence of teachers at lower grades schools causes low learning achievements in developing countries and hence is a critical issue. Thus, addressing teachers' 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 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.

How costly is the monitoring program in the KP province? We calculate the cost effectiveness of the program following the standard J-PAL costing guidelines after obtaining expenditure related information from the project completion reports published on the website of International Aid Transparency Initiative (IATI) accessed at http://d-portal.org. Table 11 shows the cost per additional standard deviation achieved in teachers' attendance and learning outcomes. According to these calculations, cost associated with achieving additional SD for reading, math and english is $2.65, $3.96 and $7.63 respectively.

| COST-EFFECTIVENESS | STANDARD | PPP |
|---|---|---|
| **Total Cost** | | |
| Total Cost to Govt & Donors (DFID and DFAT) as of 2021 | $25,301,850.00 | $98,833,330.00 |
| Total Budget Spent on IMU/EMA until 2018 | $13,743,313.00 | $53,683,719.00 |
| **School Level Outcomes** | | |
| Cost per additional SD improvement, to Govt & Donors | $414.07 | $2,977.76 |
| **Basic Learning Test** | | |
| Cost per Additional SD, Reading | $2.65 | $10.35 |
| Cost per Additional SD, Math | $3.96 | $15.48 |
| Cost per Additional SD, English | $1.43 | $5.59 |
| **Bonus Question Test** | | |
| Cost per Additional SD, Reading | $13.41 | $52.39 |
| Cost per Additional SD, Math | $8.52 | $33.28 |
| Cost per Additional SD, English | $7.63 | $29.81 |
| **Demographics** | **Number** | **Unit** |
| Program duration (from 2014-2023) Continued as a Regular EMA | 8 | year |
| Number of public schools in treatment group in 2018 | 27544 | Schools |
| No of teachers in the treatment group in 2018 | 170355 | Teachers |
| No of monitors hired under the IMU in 2014 | 550 | Monitors |
| Total number of children enrolled under the treatment group | 4445000 | pupils |

## 7. 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 innovative technology-based monitoring initiatives maintain their effectiveness and how much they affect children learning outcomes is rarely understood. In this paper, we examined the effect of a large-scale public schools monitoring program featured by the use of smart-phone and tablet aided facility through professionally trained monitors in the KP province, Pakistan. We use seven years data from a country wide nationally representative annual survey to compare program region with a neighboring untreated region that share similar characteristics in all aspects except the program. Our data consists of a rich set of variables that allow causal 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/tablet-aided biometric facility improved teachers' attendance by nearly 10% in the year immediately following the program.

We 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 0.13, 0.10 and 0.17 standard deviations points respectively at the lower (0-5) grades. There is also significant improvement in the probability of answering a relatively difficult question by upper grade children. Using an alternative 2SLS specification, we examine the causal effect of the technology-based monitoring program on the learning outcomes using teachers-attendance as endogenous variable. The 2SLS results are strikingly (nearly two-times) larger than fixed-effect OLS estimates and consistent over time providing causal evidence in support of the idea that at the lower grades, learning outcomes may be significantly enhanced using stringent monitoring of teachers in rural contexts. 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's 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 incentive 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 range effects than the targeted outcomes. Secondly, how long such effects sustain, depends on complementary measures that links teachers' performance with children performance. This study also contributes to the efforts of utilizing large-scale survey data in developing countries to produce causal evidence in impact evaluation of programs that are aimed at achieving sustainable development goals in a cost-effective way.

## 8. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Our instrumental variable approach has enabled us to overcome several concerns related to the natural experiment, especially when we observed decreasing effect after two years of program. However, despite having obtained causal effects and a clear identification strategy, we still want to point to some limitations. First, we use survey data that is collected on annual basis, and only captures the yearly inspections of schools. Using monthly data on teachers' 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 teachers' 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 lower grade children as shown in Appendix V. Although we utilize the bonus questions to examine the effect on grade 6 to 8, a more standardized design of test taken at home for higher grade children would be more useful in gauging children performance. We nevertheless are confident about consistency, internal and external validity, and the economic significance of our findings.

## 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.

Alkire, S., Dirksen, J., Nogales, R., & Oldiges, C. (2020). Multidimensional poverty and COVID-19 risk factors: a rapid overview of interlinked deprivations across 5.8 billion people.

Alkire, S., Ul Haq, R., & Alim, A. (2019). The state of multidimensional child poverty in South Asia: a contextual and gendered view.

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Annual Status of Education Reports (ASER) Pakistan (2015; 2016;2018; 2022), March 10, 2022 available www.aserpakistan.org

Annual Status of Education Report, India (Rural) (2021) Available on the ASER Centre website (www.asercentre.org)

Asadullah, M. N., Savoia, A., & Sen, K. (2020). Will South Asia achieve the sustainable development goals by 2030? Learning from the MDGs experience. Social Indicators Research, 152(1), 165-189.

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

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

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., Glewwe, P., Powers, S., & Wasserman, M. (2013). Expanding access and increasing student learning in post-primary education in developing countries: A review of the evidence. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab.

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

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

*Education Under Attack 2018, Country Profile-Pakistan*. (2018). http://protectingeducation.org/wp-content/uploads/eua2018_pakistan.pdf

Faiz, U. R., Muhammad, N., & Muhammad, S. (2017). What have we learned? Assessing the effectiveness of counterterrorism strategies in Pakistan. *Economic Modelling*, *64*, 487–495. https://doi.org/10.1016/J.ECONMOD.2017.02.028

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).

Ganimian, A. J., South, J.-P., Richard, A., & Murnane, J. (2016). Improving Education in Developing Countries: Lessons From Rigorous Impact Evaluations. *Review of Educational Research* , *86*(3), 719–755. https://doi.org/10.3102/0034654315627499

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.

Huntington-Klein, N. (2021). The effect: An introduction to research design and causality. Chapman and Hall/CRC.

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.

Javeid, U., Pratt, S., Li, H., & Zhao, G. (2022). The effect of terrorism on continuing education: evidence from Pakistan. *Education Economics*. https://doi.org/10.1080/09645292.2022.2073584

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.

Khan, H. (2019). *Establishing Data Regimes to Increase Accountability: Transforming Education in Khyber Pakhtunkhwa*, Global Development Initiative. https://effectivecooperation.org/system/files/2021-06/GDI%20Case%20Study%20Education%20Khyber%20Pakhtunkhwa.pdf

Khan, R. (2012). "Pakistan rebuilds its education network after Taliban are driven out of Swat". The Guardian's weekly outlook on international development (26 June 2012). https://www.theguardian.com/global-development/2012/jun/26/pakistan-education-swat-valley-taliban [Accessed 27/03/2023]

Khan, S., & Seltzer, A. J. (2016). *The Impact of Fundamentalist Terrorism on School Enrolment: Evidence from North-Western Pakistan, 2004-09*.

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

Malik, Z. U. A., Zhilong, H., & Ashraf, D. (2019). Terrorism: The biggest security challenge to the integrity of Pakistan. Orient Research Journal of Social Sciences, 4(1), 96-106.

Muhammad, N. (2018). *Annual Security Report 2013-2018 Special Edition Rule of Law-Security-Governance*.

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

Naseem, S., Murtaza, G., & Aslam, M. (2019). *Statistical Significance of Factors Influencing Terrorism in Pakistan. Economic and Social Development, Conference*, 1280–1288. https://www.webofscience.com/wos/woscc/full-record/WOS:000471844300159

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.

START (National Consortium for the Study of Terrorism and Responses to Terrorism). (2022). Global Terrorism Database 1970 - 2020 [data file]. https://www.start.umd.edu/gtd

UNICEF Pakistan, Investing in Pakistan's Young People (2020), Islamabad available online at https://www.unicef.org/pakistan

Wooldridge, J. M. (2015). Introductory econometrics: A modern approach. Cengage learning.

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 USING POST-MERGER SURVEY DATA

| | School Outcome | Test Score [Grade 0~5] | | |
|---|---|---|---|---|
| | TA Ratio | Reading | Math | English |
| | (1) | (2) | (3) | (4) |
| Diff-in-Diff (Treatment*Post) | 0.039 | -0.151** | -0.214*** | -0.084 |
| | (0.027) | (0.063) | (0.067) | (0.055) |
| Child related and HH controls | - | YES | YES | YES |
| School level Controls | YES | - | - | - |
| District fixed effect | YES | YES | YES | YES |
| Year fixed effect | YES | YES | YES | YES |
| Grade fixed effect | - | YES | YES | YES |
| Observations | 3,703 | 68,597 | 68,597 | 68,579 |
| R-squared | 0.057 | 0.057 | 0.124 | 0.051 |

Note: For teachers attendance ratio in column (1), ASER public schools survey dataset is used while for children standardized test score, ASER Children survey dataset is used. The diff-in-diff estimator is an interaction of the treatment (KP) and 2019 post period while the pre-period is 2012-to-2014. The dependent variable in column 1 is the teachers' attendance ratio in public schools. Test score is normalized using the mean and standard deviation of the comparison group FATA in the pre-program period (2012-2014). School level observed characteristics included in our main specification are controlled. Child related and HH controls include age, gender, parent education, private tuition, house ownership, household size, house condition, and availability of electricity, mobile and television facilities. Standard errors are corrected for clustering at the village level. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively.

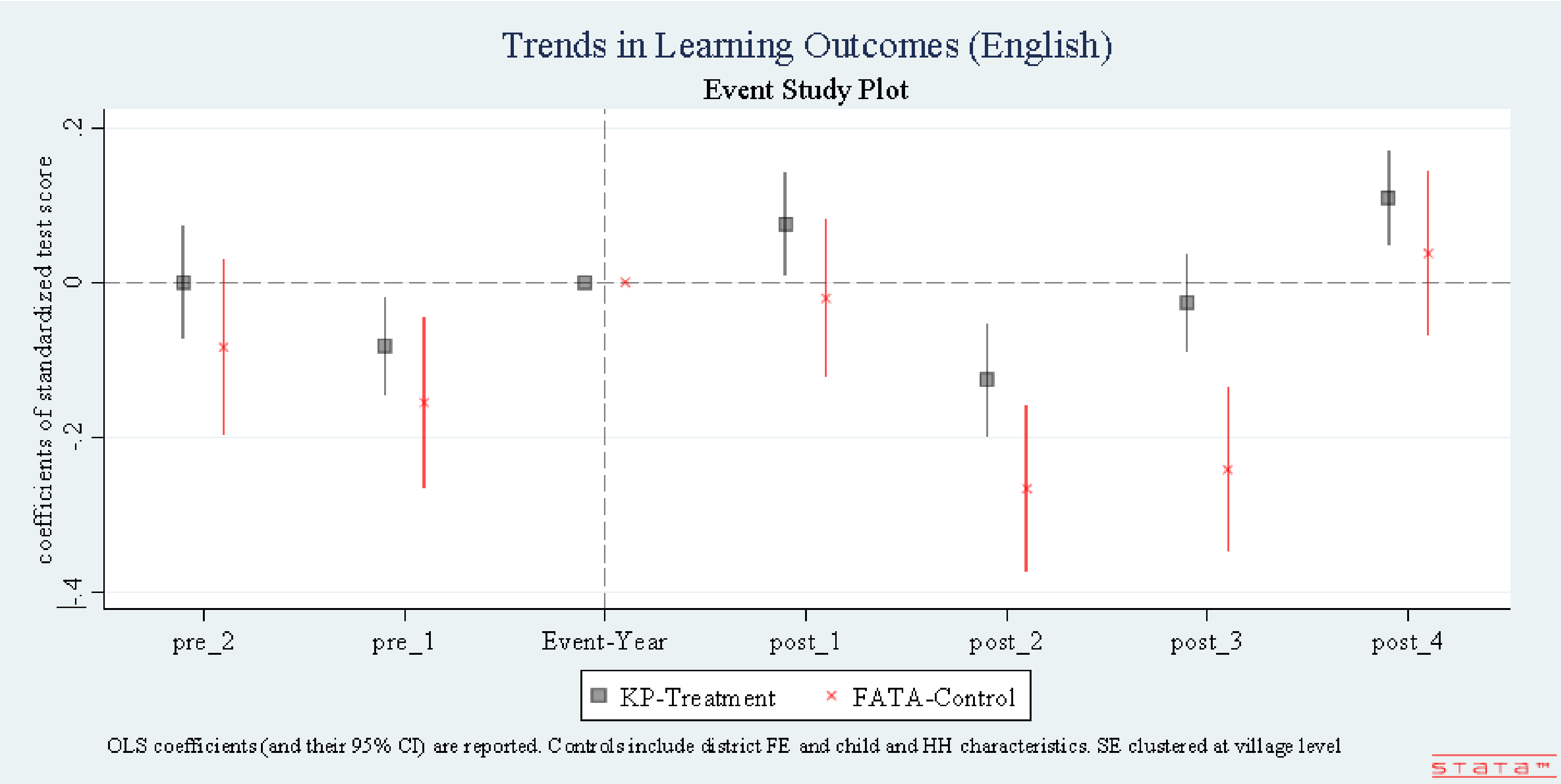
Table-A2: PROGRAM EFFECT ON SCHOOL PARTICIPATION

| Dep. Var: Children Attendance Ratio | Post=2015 | Post=2016 | Post=2018 | Post=2015~18 |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| KP *Year | 0.018** | -0.000 | 0.024* | 0.015* |
| | -0.009 | (0.017) | (0.012) | (0.009) |
| Child related and HH controls | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Mean of the Dep. Var | 0.82 | 0.82 | 0.82 | 0.82 |
| Constant | 0.802*** | 0.806*** | 0.695*** | 0.729*** |
| | (0.017) | (0.019) | (0.034) | (0.024) |
| Observations | 4,072 | 3,889 | 4,033 | 6,022 |
| R-squared | 0.105 | 0.093 | 0.126 | 0.108 |

Notes: OLS coefficients of the post-program difference between the KP and FATA regions is reported. The pre period in all columns is 2012 to 2014 pooled. The dependent variable is ratio if present children in schools to the total number of enrollments in the public school on the day of survey. District and year fixed effect, controls for school related characteristics included in all regressions. Standard errors clustered at village level are shown in parentheses. Statistical significance at the 1, 5, 10% levels are indicated by ***, **, and *, respectively

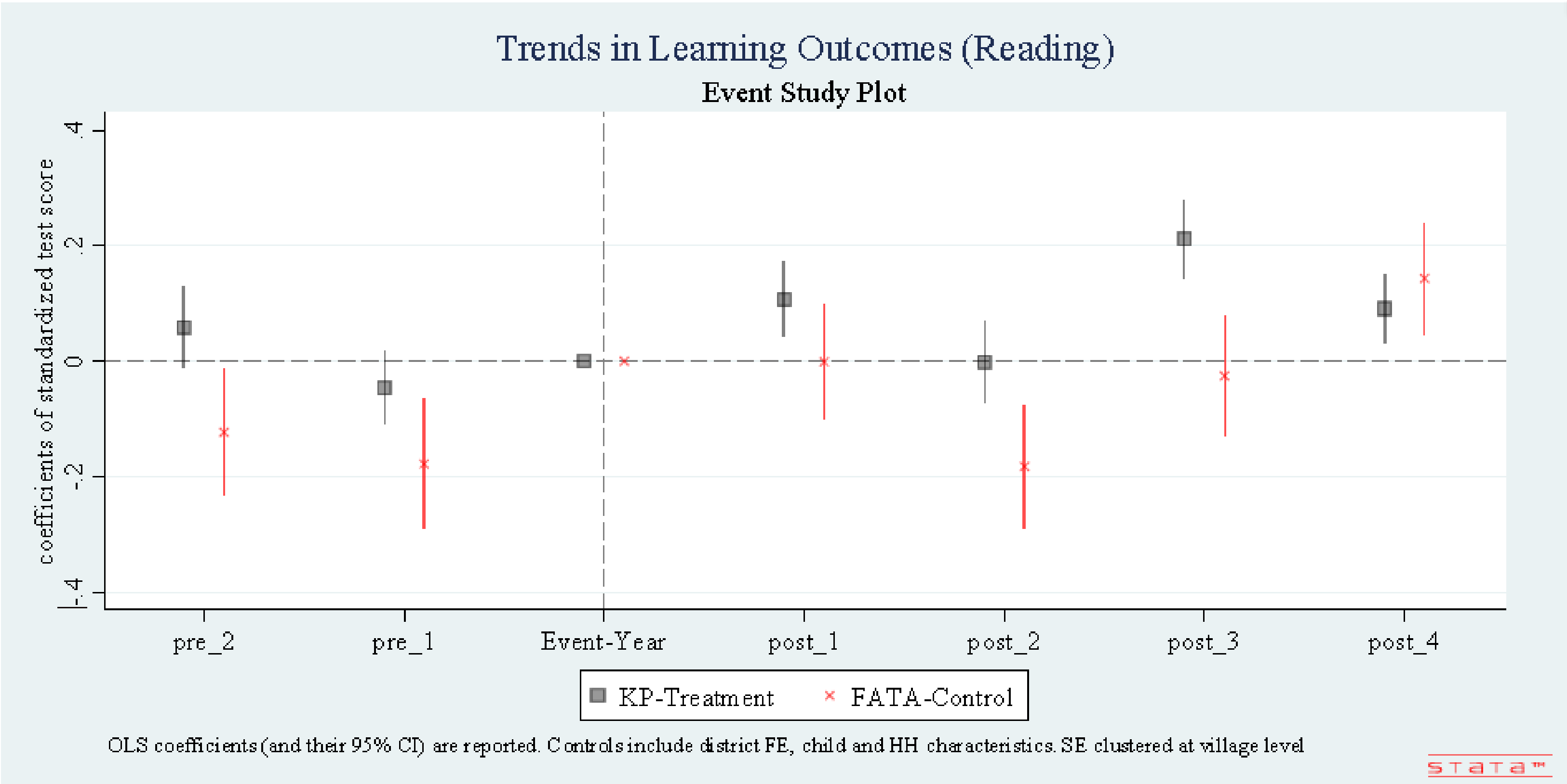
Appendix II

a. Event Study Plots (English)



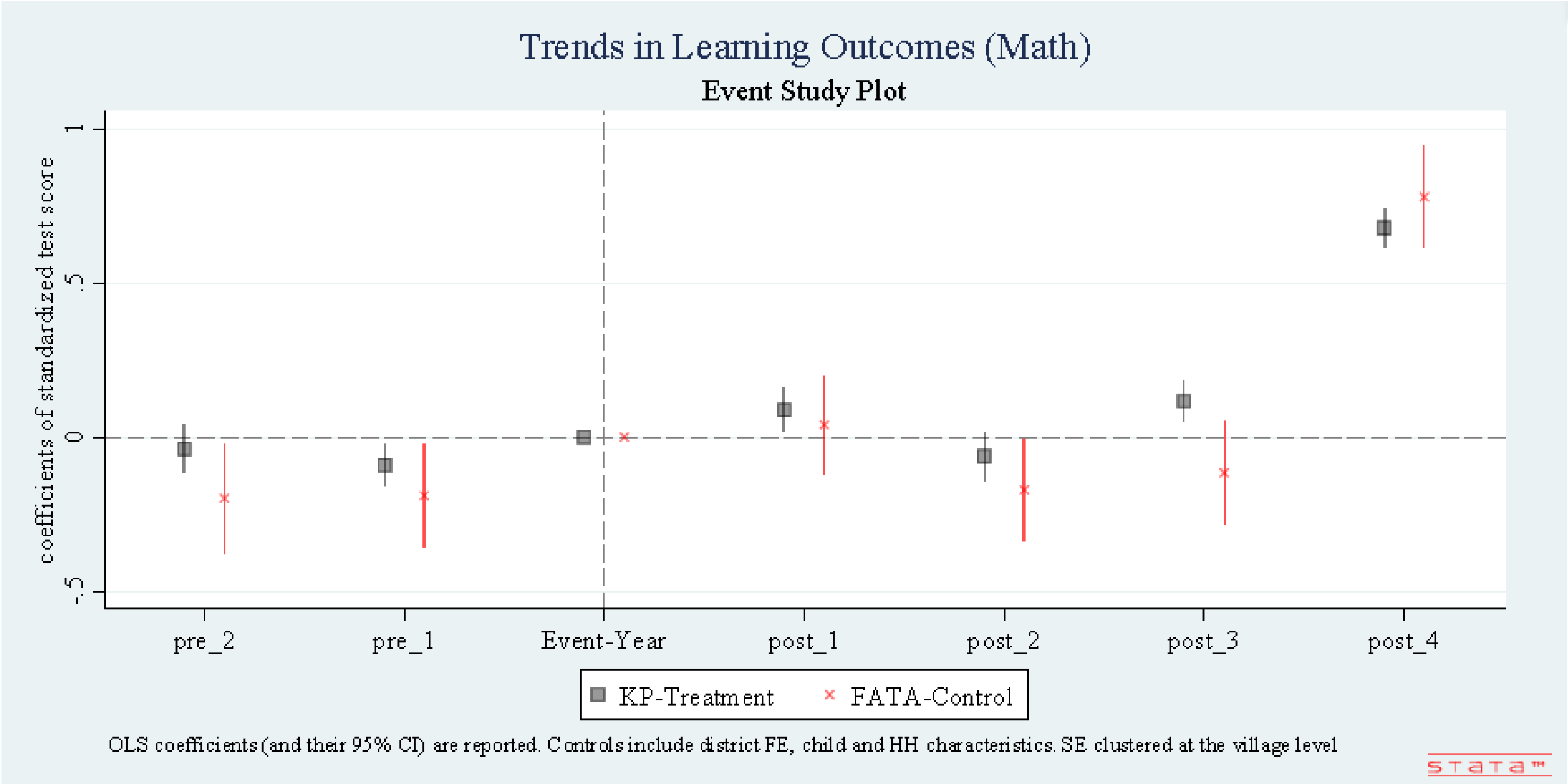
Appendix II

b. Event Study Plots (Reading)



Appendix II

c. Event Study Plots (Math)



Appendix IV

**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 V

a. 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 (11-99) task.

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

### Number Recognition (11-99)

- 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 11-99.

If s/he cannot recognize 4 out of 5 numbers from 11-99, then give her/him the number recognition 1-9 task.

### Division (2 digit by 1 digit)

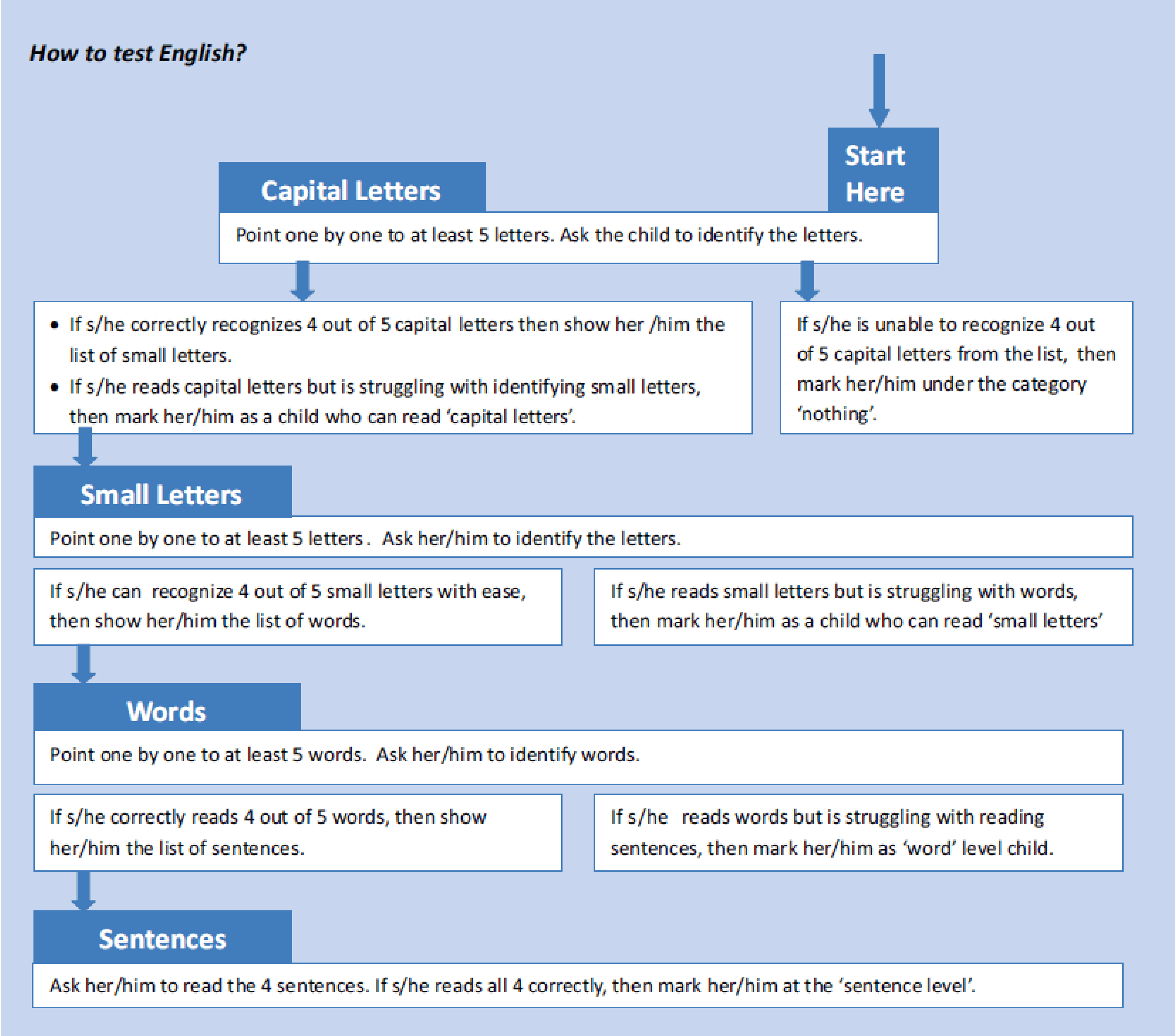
- Show the child the division problems. S/he can choose one out of the rest.
- Ask her/him to write and solve the problem.
- Observe and see if s/he is able to correctly solve the problem, and then mark her/him as a child who can do 'division'.
- If s/he is unable to solve a division problem correctly,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

b. ASER-Pakistan Children Test Procedure (English)




Source: ASER-Pakistan (http://aserpakistan.org/)

c. ASER-Pakistan Children Test Procedure (English)

***HOW TO TEST READING?***

**Start Here**

**Sentences**

- Ask the child to read any paragraph. Listen carefully as to how s/he reads.
- S/he may read slowly.
- However, as long as the child reads the text like a sentence and not like a string of words, mark her/him as a 'sentence' level child.

If the child stops very often while reading the sentence or has difficulty with more than 4 words in the sentence or reads it as a string of words than show her/him the list of words.

If the child reads the sentences fluently and with ease, then ask her/him to read the story.

**Words**

- Ask the child to read any 5 words from the word list. Let the child choose the words themselves. If s/he does not choose, then point out words to her/him.
- If s /he can correctly read at least 4 out of 5 words with ease, then ask her/him to try to read the paragraph again.
- S/he will be marked at the 'words' level if s/he can correctly read words but is still struggling with the paragraph.

**Story**

- Show the child the story. If s/he can read fluently and with ease, then mark her/him as a child who can read a story.
- If she is unable to read the story fluently and stops a lot, mark her/him as a child who is at the paragraph level.

If s/he cannot correctly read at least 4 out of 5 words she chooses, then show her/him the list of letters.

**Letters**

- Ask the child to read any 5 letters from the list. Let her /him choose the letters. If s/he does not choose then point out letters to her/him.
- If s/he can correctly recognize at least 4 out of 5 letters with ease, then show her/him the list of words again.
- If s/he can read 4 out of 5 letters but cannot read words, then mark her /him as a child who 'can read letters' .
- If s/he cannot read 4 out of 5 letters correctly, then mark her as a child as a 'beginner' .

Source: ASER-Pakistan (http://aserpakistan.org/)