

Saher Asad, Masooma Habib, Naureen Karachiwalla, Katrina Kosec, Clare Leaver, and Attique ur Rehman

# Reforming educator accountability systems

## Evidence from Pakistan



### In brief:

- Pakistan suffers from extremely low learning levels, and past research has shown that a lack of effective incentives for educators is a substantial contributor to this problem.
- Recognising this, the Khyber Pakhtunkhwa (KP) government wanted to reform the educator accountability system. Under a proposed reform, educator promotions would be linked to *performance measures* collected by government district officials, rather than, for example, seniority.
- The research team worked with the KP government to conduct a randomised controlled trial (RCT) in 240 rural primary schools during the 2017-18 school year. The RCT evaluated two variants: linking *teacher promotions* to their and their students' performance; and linking *head teacher promotions* to their and their teachers' performance.
- The study found no significant impacts on outcomes for students or educators. This may be due to 'system incoherence'— the accountability relationships created for educators and district officials were not sufficiently aligned around improving learning because of design constraints but also serious implementation challenges.
- Notably, the study provides *documentary evidence* showing that many government district officials did not follow the measurement protocols, possibly because of previous standard practices, suggesting that 'mental models' maybe be difficult to change.

*This project was funded by IGC Pakistan*

## Policy motivation for research

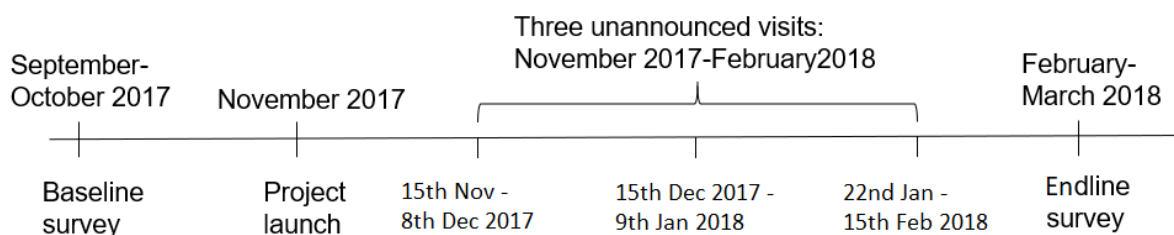
Pakistan is a prime example of low learning levels for primary school children (Andrabi et al., 2007). A recent study found that only 50 percent of children in Grade 5 could fill in a missing letter in a four-letter word accompanied by a picture, and only 54 percent could perform simple division (Andrabi et al., 2012). Learning levels are especially low in Khyber Pakhtunkhwa (KP) province. While in neighbouring Punjab province, 97 percent of Grade 4 math students in rural primary schools can correctly answer a first-grade question regarding which of 2 boxes has more objects in it, in study districts in KP, that share was just 43 percent.

Career incentives—specifically, the promotions and dismissals that educators face—are the primary source of accountability for teachers and head teachers in low-income countries. Educators are held accountable to the government because they care about their future career. However, in KP, these career incentives were weak. Although the province’s Elementary and Secondary Education (E&SE) Department had teacher evaluation and school inspection processes in place, neither system collected relevant and well-measured data that could be used to construct informative and objective performance measures. Compounding this, promotion and dismissal decisions were not made transparently on the basis of performance, but instead hinged on seniority, political connections, and other factors unassociated with student learning, which is the main outcome that the Department wants to improve.

## What we do

This project, in close partnership with the E&SE Department, sought to strengthen teacher and head teacher career incentives by reforming the process for evaluating educators. The design of the evaluations was based on a successful example of career incentives for teachers in China (Karachiwalla and Park, 2015). A pilot was introduced in 240 public, rural primary schools in the districts of Charsadda, Mardan, and Nowshera during the 2017-2018 school year, using a randomised controlled trial (RCT) design (see timeline in Figure 1). Within each of 20 school districts, 12 schools were randomly assigned to one of three equally-sized study arms: (1) a *control group* in which business as usual inspections were to occur, (2) a *teacher treatment* in which a government inspector would collect four performance measures relating to the Grade 4 math teacher (the teacher’s attendance and use of appropriate pedagogy during a classroom observation, and the attendance and performance of his/her students in Grade 4 math orally administered test) which would then determine the Grade 4 math teacher’s chances of promotion; and (3) a *head teacher treatment* in which a government inspector would collect four performance measures relating to the school (the head teacher’s attendance, the attendance of all teachers and all students in the school, and teachers’ use of appropriate pedagogy across two classroom observations) which would then determine the head teacher’s chances of promotion. The same inspectors were used across all study arms, but they were carefully trained on the distinct, new role they were to play in the treatment arms: one of passive observation using a tablet to take photos and videos to collect data for promotions rather than active engagement with educators and children, which was standard practice in business as usual school visits. Each treatment arm was to receive three surprise visits from an inspector during the school year.

**Figure 1: Timeline of events**



## Results

A number of problems arose during implementation. First, the E&SE Department agreed to monitor inspectors' performance using independent auditors but chose relatively limited rewards: top performers would receive public recognition at an awards ceremony and a note would be added to their promotion dossier. As a result, even at the design stage, this incentive system was low stakes for inspectors. Cilliers et al (2019) show that it is important to 'monitor the monitors' in order to motivate inspectors such as district officials to perform well and as intended.

Second, the government's support for the interventions waned just before implementation began. During the design phase, the E&SE Department agreed to send official letters of notification to study participants stressing the importance of the research and encouraging them to cooperate with the researchers and to strictly adhere to study protocols. Unfortunately, the Department did not send out the usual official directive, possibly due to changes in key personnel between the design and implementation phases of the project. When this became apparent, the study team stepped in and sent out notifications to teachers, head teachers, inspectors who would conduct the inspections, and higher up district officials just before the interventions began. Thus, the interventions may have had less "bite" than intended (c.f. Muralidharan and Singh, 2019).

Third, inspections occurred more slowly than anticipated, and only a subset of schools received three visits. On average, schools received 2.5 inspections due to scheduling constraints.

Finally, and likely as a result of the first two factors, inspectors did not always collect data passively and follow the study protocol. Many engaged actively during the classroom observations of teacher pedagogy (e.g., they suggested a good pedagogical practice to the teacher) and the administration of oral student math tests (e.g., they provided students with hints to the correct answer).

On the basis of data collected at endline, there is no evidence of an impact on any student outcome (attendance and dropout rates measured in the endline survey or learning in written math tests administered at endline) in either the teacher or head teacher treatment. Coefficients on treatment indicators are very small in magnitude and statistically insignificant. There is also no evidence of an impact on any school-level outcome (teacher and head teacher attendance or student enrollment as reported in the endline survey).

Why did both accountability interventions fail to impact outcomes, including behaviours like teacher or head teacher attendance that could feasibly change during the short project timeframe? We identify as

an explanation the ‘system incoherence’ between accountability relationships (Pritchett, 2015). The E&SE Department chose to design a high stakes incentive system for agents at the bottom tier (teachers/ head teachers) but not for agents at the top tier (district inspectors), and then compounded this by not sending a credible commitment of its support for the intervention to participants. As a result, both incentive systems were likely perceived to be close to zero stakes.

The video recordings collected during the study provide a valuable insight into how the inspectors responded to this system incoherence. Students hired by our team watched all of the available videos and observed that, rather than conducting passive performance monitoring as the study design had intended, many inspectors continued to use the same active practices during their visits as they had under ‘business as usual’ prior to the intervention. For instance, in a quarter of (treatment) schools there is documentary video evidence that inspectors intervened during what should have been a passive classroom observation to correct a teacher’s actions, rather than simply noting a low score. Even more strikingly, in 43 percent of (teacher treatment) schools there is documentary video evidence that inspectors either hinted at, or provided, the answer to students during the oral Grade 4 math tests. Clearly, a deviation from protocol may have occurred off camera (or was hard to assess and code), and so these already large numbers are likely underestimates. Given the lack of implementation fidelity, it was felt that the data collected by the inspectors did not represent a balanced and fair assessment of teacher and head teacher performance and promotions were not ultimately tied to these data.

One interpretation of these findings is that some inspectors deviated from the intervention protocols because it was difficult for them to change their “mental model”; i.e. some inspectors continued to feel that it was their responsibility to coach and provide advice to educators. To start to gain insight into this explanation, we tested for correlations between the incidence of deviations and the observed characteristics of inspectors. The most consistent finding relates to experience: across both the classroom observation and Grade 4 math test protocols, inspectors with below median experience in the profession were more likely to deviate compared to their peers with more experience. Other characteristics were correlated with one but not both outcomes. Female inspectors (inspecting girls’ schools) were more likely to deviate from the classroom pedagogy observation protocol compared to male inspectors (inspecting boys’ schools), whereas inspectors with local family ties were less likely to deviate from the classroom observation protocol compared to their peers who were not connected. Intriguingly, inspectors with an above median score for civic-mindedness were more likely to deviate from the protocol for the oral tests compared to their peers with a below median score. We plan to probe these correlations further in future qualitative work.

It is obviously disappointing that by endline in March 2018, there was no evidence that the pilot interventions had improved student learning outcomes, or indeed shifted any aspect of student or educator behaviour. However, the videos recorded during the inspectors’ school visits provide a rich source of information and, in particular, documentary evidence of a lack of implementation fidelity. In many trials with null results, researchers must conjecture about underlying mechanisms. Here, thanks to the videos, it is unambiguous that the intervention failed to change the behaviour of many inspectors. This evidence makes it easier to draw out lessons and to make recommendations for the E&SE Department moving forward. Given that the E&SE department requested evidence-based research to inform policy, lessons may well be taken up in the future.

## Policy recommendations

Despite the study's findings of null impacts on students, teachers, and schools, the documentary evidence it collected on the implementation process provides several lessons and policy recommendations.

### 1. Changing district officials' mental model is difficult.

Using existing government personnel (government inspectors) to undertake performance monitoring clearly has benefits, as accountability can be embedded within the system. However, staff who previously visited schools to provide coaching and feedback appear to need encouragement and direction to change their mental model, and to switch from this active role to that of a passive performance monitor. Without this, objective data cannot be collected and used for educator promotions.

### 2. Increasing the stakes for district officials could be important.

The incentive system for inspectors was low stakes. While some inspectors did make the switch from active feedback to passive performance monitoring, many did not. This was particularly true of less experienced inspectors. Stronger incentives---i.e. more salient rewards for good monitoring, and perhaps sanctions for poor monitoring, backed up by a stronger commitment from the E&SE Department to follow through---may be necessary to encourage all inspectors in all schools to switch to the role of a passive performance monitor.

### 3. There are unlikely to be improvements without system coherence.

The incentive system for teachers and head teachers was intended to be high stakes, with meaningful career prospects on the line. Ultimately, this design feature appears to have had little bite because of incoherencies elsewhere in the system. The E&SE Department did not make a demonstrable, strong commitment to the interventions. And many inspectors did not objectively monitor teacher/head teacher performance. The interventions strengthened one element of the accountability system but likely had no impact on educator and student behaviour, at least in part, because other elements were not similarly strengthened at the same time.

The natural policy question is whether the KP government should work towards a more coherent accountability system with strengthened, credible incentives for both teachers/head teachers *and* performance monitors. Or should it adopt a different approach, for instance, by building on the active role played by inspectors and introducing greater in-service training and coaching for educators in schools? In fact, these policies are not mutually exclusive, and the government could consider both. The key point is that the roles of an active coach and passive performance monitor are strategically different and, for incentive reasons, may need to be performed by different personnel.

## References

- Andrabi, T., J. Das, and N. Karachiwalla. (2012). Scope of Work, Education Fund for Sindh voucher evaluation. Technical report.
- Andrabi, T., J. Das, A. Khwaja, T. Vishwanath, and T. Zajonc (2007). The Learning and Educational Attainment in Punjab Schools (LEAPS) Report. World Bank, Washington, DC.
- Cilliers, J., Kasirye, I., Leaver, C., Serneels, P., & Zeitlin, A. (2019). Improving Teacher Attendance Using a Locally Managed Monitoring Scheme: Evidence from Ugandan Primary Schools.
- Habib, M. (2015). Teacher and School Administrator Incentives for Improved Education Delivery in Khyber Pakhtunkhwa Province, Pakistan. The International Growth Center, Working paper, May 2015.
- Karachiwalla, N. and A. Park, (2015). Promotion Incentives in the Public Sector: Evidence from Chinese Schools, HKUST IEMS Working Paper No 2015-09.
- Muralidharan, K., & Singh, A. (2019). Improving Public Sector Management at Scale: Experimental Evidence on School Governance in India, mimeo.
- Pritchett, L. (2015). Creating education systems coherent for learning outcomes: Making the transition from schooling to learning. Research on Improving Systems of Education (RISE) Working Paper, Preliminary draft.