Improving Public Sector Management at Scale? Experimental Evidence on School Governance in India*

Karthik Muralidharan† Abhijeet Singh‡

December 1, 2022

Abstract

We present results from a large-scale experimental evaluation of an ambitious attempt to improve management quality in Indian schools (implemented in 1,774 randomly-selected schools). The intervention featured several global "best practices" including comprehensive assessments, detailed school ratings, and customized school improvement plans. It did not, however, change accountability or incentives. We find that the assessments were near-universally completed, and that the ratings were informative, but the intervention had no impact on either school functioning or student outcomes. Yet, the program was scaled up to cover over 600,000 schools nationally. We find using a matched-pair design that the scaled-up program continued to be ineffective at improving student learning in the state we study. We also conduct detailed qualitative interviews with frontline officials and find that the main impact of the program on the ground was to increase required reporting and paperwork. Our results illustrate how ostensibly well-designed programs, that *appear* effective based on administrative measures of compliance, may be ineffective in practice.

JEL Codes: H75, I28, O15

Keywords: bureaucracy, management, state capacity, foreign aid, RCT, education

*We thank Felipe Barrera-Osorio, Eli Berman, Nick Bloom, Martina Björkman-Nyqvist, Gordon Dahl, Esther Duflo, Tore Ellingsen, Alejandro Ganimian, Bob Gibbons, Joshua Goodman, Marc Gurgand, Caroline Hoxby, Hyuncheol Bryant Kim, David McKenzie, Paul Niehaus, Lant Pritchett, James Rauch, Imran Rasul, John van Reenen, Sujata Visaria, Martin Williams, Kaspar Wuthrich, and several seminar participants for comments. We are grateful to senior leadership in the School Education Department, Government of Madhya Pradesh – in particular, Mr. S.R. Mohanty, Ms. Rashmi Shami, Ms. Deepti Gaur Mukherjee, Mr. Lokesh Jatav and Mr. K.P.S. Tomar – for their continuous support in evaluating this program. We are also grateful to staff at Ark, especially Kruti Bharucha, Susannah Hares, Kanupriya Misra and Sachin Ashapure, for their collaboration in this project. Aditi Bhowmick, Akankshita Dey, Ghazal Gulati, Krishanu Chakraborty, Nawar El Abadi, Ramamurthy Sripada, Sabareesh Ramachandran and Urmi Bhattacharya provided outstanding field management and research assistance. This evaluation was supported by funding from the International Growth Centre, the J-PAL Post Primary Education Initiative, ESRC and the DFID RISE program. This trial was registered on the AEA RCT Registry (AEARCTR-0000851).

†Department of Economics, UC San Diego; JPAL; NBER. Email: kamurali@ucsd.edu

‡Department of Economics, Stockholm School of Economics; JPAL. Email: abhijeet.singh@hhs.se

1 Introduction

Improving state capacity for public-service delivery in low- and middle-income countries remains among the biggest development challenges in the world (World Bank, 2003; Andrews, Pritchett, and Woolcock, 2017). A growing body of evidence has found positive impacts of improving incentives for front-line government staff in many settings (see e.g. Mbiti et al. (2019), Leaver et al. (2021), Khan, Khwaja, and Olken (2016, 2019)). Yet, political and measurement constraints often make scaling and sustaining such high-powered incentives difficult in large bureaucracies. Hence, policymakers seeking to improve bureaucratic performance at scale often look for options that do not involve introducing high-powered incentive schemes.

One popular approach has been to try to improve management quality within bureaucracies, motivated in part by the finding that better management practices are correlated with more effective public service delivery across sectors (Rasul and Rogger (2018), Rasul, Rogger, and Williams (2018)). Interventions to improve public sector management have been especially popular in education, where prior work has found that management quality is positively correlated with both test-score levels and value-added (Bloom et al., 2015; Lemos, Muralidharan, and Scur, 2021). For instance, the World Bank alone has funded school management reform programs in 84 countries in a 20-year period from 1998-2017 (listed in Appendix A). However, despite the popularity of such programs, and the associated fiscal and personnel costs, there is very little evidence on whether they are able to change actual behaviors and outcomes at scale.

In this paper, we present experimental evidence on the impact of a comprehensive school management program, which was a more intensive precursor of a variant that has since been rolled out to over 600,000 schools in India and is expected to cover 1.6 million schools eventually. We worked with the government of the state of Madhya Pradesh (MP) to conduct an experimental evaluation of the first phase of this program (in 2014-16) that was implemented across 1,774 elementary schools (randomly selected from a universe of 11,235 schools). The program was developed by the Government of MP (GoMP) and Ark, a leading international education services provider, and aimed to incorporate several global "best practices" in school management.

The program's theory of change reflected insights from management theory, which argue that poor management practices may persist in organizations for a combination of reasons including managers (a) not knowing that they are performing poorly, (b) not knowing what they need to do to improve, (c) not being motivated or held accountable for improvements or (d) not succeeding in improving performance due to coordination and agency problems (Gibbons and Henderson, 2012). The program aimed to address each of these issues by: (a) conducting independent customized assessments of school quality, to identify strengths and weaknesses, (b) creating school-specific improvement plans with concrete action steps, (c) stipulating regular follow-up by supervisory staff to monitor progress and provide guidance and support, and (d) aiming to

involve school inspectors, all school staff and parent representatives in the assessments and the creation of improvement plans. The program also created online tools to make school assessment results and progress reports easily accessible to senior officials in the education department.

The program aimed to foster cooperation across all actors in the education system, and to avoid an adversarial framing between administrators and teachers. It was therefore framed as a collaborative effort to support continuous improvement of schools, but did not formally change teacher incentives or accountability for improved outcomes. While school-level assessments were an integral part of the intervention, the program aimed to use them as a starting point for customized school-specific planning for improvement, and not as a tool for accountability. The idea of continuous improvement is also similar to the highly-regarded Japanese *kaizen* management system (Morgan and Liker, 2006). Overall, the program had a coherent theory of change, and reflected perceived global "best practices". It was the flagship education reform in MP, and international funding agencies as well as senior GoMP officials were highly optimistic that it would improve school processes and learning outcomes.

The program was designed to be scaled up across MP in a staggered way, starting with an initial pilot phase of 100 schools, and expanding to ∼2,000, ∼25,000, and ∼100,000 schools in three phases. Our experimental evaluation was conducted in the expansion to ∼2,000 schools after the program design had been adapted to the local context and stabilized during a 100-school pilot. Working closely with GoMP, we identified a population of 11,235 elementary schools across five districts, and randomly-assigned 1,774 to receive the program. We randomized units into treatment at the level of an academic cluster, the lowest unit of school administration in the state. Clusters have around 40 schools on average, with dedicated cluster resource coordinators (CRCs) to provide mentoring and academic support to schools. Our study design, combined with at-scale program implementation by the government, is likely to yield estimates of program impact at a policy-relevant scale (Al-Ubaydli, List, and Suskind, 2017; Muralidharan and Niehaus, 2017).

Our primary outcome of interest is student learning. We measure this using three data sources: (a) independently designed and administered tests of student learning in a representative sample of 100 treatment and 202 control schools, (b) individual student-level scores on official assessments in these schools and (c) school-level aggregate scores on official assessments across all treatment and control schools. We also collected data on teacher and student absence, conducted classroom observations, and surveyed principals, teachers and students.

We report four main results. First, the school assessments were completed in 93% of schools assigned to treatment and were of high quality. School improvement plans were also made and uploaded on the program website. In the overall ratings, 91% of schools were classified as not meeting standards, suggesting low collusion between assessors and school staff. The assessments

also contained meaningful variation and information on school quality, seen by their ability to predict *future* student achievement and teacher absence, even conditional on past achievement.

Second, though the initial assessment was implemented in treated schools and was informative, there was no sustained improvement in support or oversight in treated schools. Supervising officials did not increase their monitoring of treated schools, whether in frequency of visits or the content of inspections. School Management Committees, which could have exerted community-level accountability, and were explicitly targeted by the intervention, also did not play a more active role in treatment schools.

Third, while there was no change in oversight, the program could have still improved teacher effort and classroom processes through the information in the assessments and the school-improvement plans. However, we find no evidence of improved pedagogy or effort within schools. Teacher absence was high (33%) and did not differ across treatment and control schools. We also find no impact on instructional time, use of textbooks and workbooks, or the likelihood of checking student homework books (a measure of teacher effort). Student absence rates were also high (47%) and unaffected by treatment. Thus, the program was ineffective in changing *any* observed measure of teacher effort or student engagement.

Finally, consistent with the lack of impact on any school-level processes, we find no impact on student learning outcomes either in the short run (3-4 months after the intervention) or over a longer horizon (15-18 months after program rollout). This is true in both school-administered tests, and independently-administered tests conducted by the research team. These null results are precisely estimated and we are able to reject modestly-sized effects between 0.1 and 0.15σ.

At the time that we presented these results to GoMP officials, they had already planned for the expansion of the program to the next phase of ∼25,000 schools in late 2016. The impetus for continuing the expansion (despite evidence of non-impact) was magnified by the creation of a national program along similar lines, for which MP was designated as a "forerunner" state. This national program has now been implemented in over 600,000 schools, is expected to reach 1.6 million schools, and is the national flagship program for improving school management. The ∼25,000 school expansion in MP was part of this nationwide rollout.

We complement our main experimental evaluation of the ∼2,000 school roll-out with a non-experimental evaluation of the ∼25,000 school expansion using a matched-pair treatment control design (we also test for and verify parallel trends in test scores in prior years). We again find no impact on student learning outcomes. Thus, even after over five years of iterating on the design of the program and expanding its scale, it had no discernible effect on student learning.

We conducted extensive qualitative interviews with teachers, principals, and field-level supervisory staff to obtain insights into the reasons for the program's ineffectiveness. We

document that, for these officials, the program was reduced to an exercise in administrative compliance, i.e. ensuring that the required paperwork was submitted on time. Both teachers and supervisors perceived the program primarily as a data collection effort: program delivery effectively ceased after filing the school improvement plans and, de facto, the reform was very far from the reflective exercise in self-evaluation and improvement envisaged in the program design.

Our first contribution is to the literature on improving public sector effectiveness in developing countries. While observational evidence suggests that better management practices are correlated with productivity in the public sector (e.g. Rasul and Rogger (2018); Lemos, Muralidharan, and Scur (2021)), there is much less evidence on how these practices can be improved. The program we study copied several global "best practices" that are ubiquitous in education management reforms around the world (see Appendix A). Yet, in both the ∼2,000 and ∼25,000 school implementations, we find no evidence of impact on either school functioning or learning outcomes. As shown by Abadie (2020), a well-identified and well-powered null result is especially valuable in settings where the prior beliefs of program effectiveness are high (as stated in policy documents here, and implied by the widespread adoption of similar reforms globally).

Our results contrast with recent evidence of large and persistent productivity gains from providing management consulting to private firms in developing countries (see e.g. Bloom et al. (2013); Bruhn, Karlan, and Schoar (2018); Iacovone, Maloney, and McKenzie (2022)). One possible reason for this difference is that unlike private firms who have incentives for improving productivity, officials and teachers in most public-sector contexts (including ours), had few incentives to improve outcomes, and the intervention did not change these. Thus our results are consistent with growing evidence of complementarities between inputs (including knowledge) and incentives in improving outcomes in developing countries across education, health, and even in the private sector.[1] It may also explain the contrast between our findings and evidence from the US and the UK, where school ratings have been found to improve student performance when accompanied by the threat of sanctions for low-performing schools and principals (Figlio and Loeb, 2011; Figlio and Rouse, 2006; Rockoff and Turner, 2010; Hussain, 2015).

Another possibility is that the program was too "light touch" to be effective. Even in the private sector, successful management interventions (such as those cited above) have featured intensive follow-up efforts after the initial diagnostic assessment. Thus, program effects in our setting may have been higher if the government had been able to follow up on implementing the school

[1] In education, Mbiti et al. (2019) find no impacts of providing school grants alone, modest impacts of teacher incentives, and strong evidence of complementarities between the two. Muralidharan and Sundararaman (2011) show that trained teachers are not more effective than untrained ones in public schools, but are significantly more effective in the presence of performance-linked pay. In health, Das et al. (2016) show that the *same* doctors provide significantly better care on the same case in their private practice compared to in public clinics. Finally, Atkin et al. (2017) show how misaligned incentives between workers and owners may limit the effectiveness of providing workers with more productive technologies even in private firms.

improvement plans. However, both of these potential reasons for non-impact — weak incentives and the lack of capacity to deliver intensive treatments at scale — are ubiquitous in bureaucracies. Thus, our results should *not* be interpreted as implying that management interventions do not work in the public sector. Rather, they demonstrate that improving management quality at scale in public bureaucracies is very challenging, and that we need much more research on how to do so.[2] We examine promising approaches further in the conclusion, where we discuss insights from comparisons with other public-sector interventions that *have* been successful at scale.

What is perhaps even more striking is that the program was scaled up to over 600,000 schools even though it did not improve any ultimate outcome of interest.[3] Our second contribution, is to shed light on the bureaucratic incentives that pushed the program to be scaled up despite lack of any impact. Specifically, our detailed qualitative interviews illustrate how paperwork and the *appearance* of activity may be an end in itself even when ultimate goals are unaffected (Gupta, 2012; Aiyar, Dongre, and Davis, 2015; Levy et al., 2018). Thus, despite lack of any impact on outcomes, the program was deemed a success by administrative metrics since there was a paper trail of assessments done and reports uploaded to prove it.

Our combination of experimental program evaluation and qualitative interviews also provide a clear illustration of "institutional isomorphism", whereby bureaucracies focus on copying the practices of counterparts that are perceived to be more successful, *regardless* of their actual effectiveness (DiMaggio and Powell, 1983; Pritchett, 2013). As noted by DiMaggio and Powell (1983): "as an innovation spreads, a threshold is reached beyond which adoption provides legitimacy rather than improves performance." This phenomenon may explain both the initial enthusiasm for the program and the subsequent scale up, despite a lack of any impact.

Finally, our results also contribute to the literature on the effectiveness of aid-funded technical assistance programs. Critics of such programs have noted that these are often designed by global experts located in high-income countries who aim to translate "best practices" across settings, without recognizing that the binding constraint to improvement is not the knowledge of the practices *per se*, as much as the weak local institutions within which these practices have to be embedded (Easterly, 2002, 2007). However, since these programs are rarely evaluated, aid agencies (supported by an ecosystem of consultants and experts) continue to produce similar programs without being held accountable for outcomes (Martens et al., 2002). Our results suggest that these criticisms may have considerable merit. We discuss policy implications for the billions of dollars spent every year on similar technical assistance programs in the conclusion.

[2]They are also consistent with recent evidence from Gambia and Mexico showing that interventions to improve school management have had no impact on student learning (Blimpo et al., 2015; Romero et al., 2022)

[3]Our setting and results, thus, differ importantly from prior studies on ineffective governance reforms where the interventions were poorly implemented (and later abandoned) due to subversion and resistance by lower-level officials (see, e.g., Banerjee, Duflo, and Glennerster (2008) and Dhaliwal and Hanna (2017)).

2 Background and Intervention

2.1 Setting

Our study is set in the state of Madhya Pradesh (MP), which had a population of around 72.6 million (72% rural) in 2011. It is one of India's more disadvantaged states, with a lower literacy rate and a higher poverty rate than the national average.

The MP public education system comprises four hierarchical levels of administration. At the apex is the state level, where policy, program, and budget decisions are taken for the full population of over 110,000 public schools.[4] Next are the district and block levels, which coordinate policy and program implementation at a scale of ∼2,000 and ∼300-400 schools. The lowest level is the academic cluster, which typically caters to around 40 schools, and is staffed with two Cluster Resource Coordinators (CRC) who represent the frontline interaction of the education bureaucracy with schools. CRCs and block-level officials are meant to oversee school functioning, monitor school compliance with official norms, exert accountability pressure, and provide administrative support as needed. In addition, all schools are expected to have a School Management Committee comprising representatives of parents, the school staff, and local functionaries to provide "bottom up" community-based monitoring of schools.

Yet, despite this well-defined *formal structure* for school accountability, the performance of the public education system in MP is weak. Student learning levels are low. In 2016, only 31% of Grade 5 students in government schools were able to read a text at Grade 2 level, and only 15.3% of Grade 5 students could do division (Pratham, 2017). There is also evidence of deterioration of public school quality in MP in recent years. For instance, student attendance (measured by unannounced school visits) in MP public schools fell from 68% in 2010 to 55% in 2016 (Pratham, 2017). Teacher absence was also high, and rising, with 2010 levels of teacher absence in rural MP estimated at 26.2%, compared to 18.2% in 2003 (Muralidharan et al., 2017).[5]

2.2 Intervention

Concerned about the low quality of public schooling in the state, the Government of MP (GoMP) made school improvement a high priority and requested technical support from the UK Department of International Development (DFID). DFID contracted Ark, a leading education charity headquartered in London, with a track record of leading school improvements in public systems internationally and in Academies (Charter schools) in the UK. The program of school

[4]The large number of schools reflects a policy priority on providing primary schools in every habitation in the state to facilitate universal school enrollment. As a result, there are a large number of small schools, with 40% of rural primary schools having enrollment below 60 students (across grades 1-5) in 2016.

[5]The trend line of deteriorating school quality highlights the importance of a credible contemporaneous control group for evaluating the impact of policy interventions.

inspections and feedback for improvements was informed by the school accountability regime run by the Office for Standards in Education, Children's Services and Skills (OfStEd) in the UK but extensively adapted to the context in MP.[6] This eventual intervention, officially called the MP *Shaala Gunvatta* program (translated as the "MP School Quality Assurance" program), consisted of three main components, summarized below and described in greater detail in Appendix B:

1. ***Developing a school rating scorecard:*** These were developed after extensive consultation and piloting. The school rating scorecard was based on structured indicators in seven domains:

   (a) **Mentoring**, which looks at whether there is a School Development Plan in place and whether there is a vision for school development;

   (b) **Management**, which assesses the use of financial and human resources in the school, morale and coordination of staff, and the effectiveness of administrative routines and procedures;

   (c) **Teacher practice and pedagogy**, which is based on detailed classroom observations in each grade within the school and focuses on teaching practices, teacher-student interactions and teacher subject knowledge;

   (d) **Student support**, focusing especially on availability of remedial education, support for students with special needs and on the absence of discrimination among students;

   (e) **School Management Committee and interaction with parents**, which focuses on the involvement of parents of students in planning of school activities as well as in sharing information about the academic progress of individual students;

   (f) **Academic outcomes**, based both on the distribution of scores based on the state-wide assessments (Pratibha Parv) as well as in-school testing;

   (g) **Personal and Social Outcomes** which assesses (based on class and school observations as well as interactions with students and staff), the attitudes, values, and relationships between the students and with teachers.

   Since the school rating was the foundation of the program, the CRC was teamed with an external evaluator (such as a retired headteacher) to ensure high-quality assessments. Schools were rated overall, and in each of the seven domains, which provided the basis for school-specific recommendations for improvement.[7]

[6] For more details on OfStEd inspections and their impact on schools and students, see Hussain (2015). In addition to obtaining inputs from global experts, Ark took substantial effort to make the program context-specific and spent 2 academic years to pilot and refine program details in 100 schools prior to the larger-scale implementation that we evaluate. See Appendix B for details.

[7] The detailed assessment rubric intended to alleviate an important constraint for CRC effectiveness, namely the lack of sufficient structure and training in how to conduct school inspections and how to address deficiencies in school practice.

2. ***School Improvement Plans:*** The school assessments and ratings were then used to develop customized School Improvement Plans (SIPs). These were expected to lay out in detail the proposed improvements in school functioning, the concrete steps that would need to be taken to achieve the improvements and a deadline for the actions to be undertaken. These also mentioned the official(s) responsible for executing each specific task (e.g. the headteacher or other teachers) and for verifying that the task has been completed (e.g. the CRC or the School Management Committee (SMC)). Overall, the goal of the SIPs was to set schools with manageable improvement targets and goals, that they could aim to achieve in incremental steps over three-month periods.

3. ***Quarterly follow up visits by CRCs:*** These were meant to review the improvement made on each component of the School Improvement Plan (SIPs) The follow-up visits aimed to provide an external impetus to make progress towards the goals in the SIP, and set progressively more ambitious goals for subsequent quarters. This was an important component of the "theory of change" that aimed to motivate schools to deliver *continuous* improvement.

All assessments and School Improvement Plans (SIPs) were uploaded to a dedicated online portal. These reports, with user-friendly visual aggregates, were available to view for any user with requisite administrative access including head-teachers, assessors and higher officials at cluster/block/state levels. This was intended to present relevant information for prioritization in decision-making, and administrative follow-up.

Overall, the MPSQA program did not focus on one single component of governance (such as inspections or incentives) because it recognized that the binding constraint for quality would be likely to differ across schools. Given the considerable heterogeneity in school quality in the state (which we present evidence of below), the premise of the intervention was to motivate head teachers and teachers to focus on actions that are in their control that could alleviate constraints to quality that they saw as being relevant to their school.

3 Study Design

3.1 Sampling and experiment design

GoMP conducted the first phase of program roll-out in 5 contiguous districts (out of a total of 51 in the state) that included the state capital (Bhopal) and 4 adjacent districts (Figure 1). These districts had a combined population of $\sim$ 8 million in 2011 and $\sim$12,000 government schools, of which GoMP aimed to treat $\sim$2,000 schools.

We worked with GoMP to conduct an experimental evaluation of this program by randomizing the assignment of schools to the program using a clustered design. Since the unit of project

implementation was the cluster, we randomized entire academic clusters to treatment and control groups. We did this to minimize the possibility of spillovers between control and treatment schools assigned to the same CRC.[8] Randomization proceeded in two steps. First, we first drew a representative sample of 153 clusters out of a total of 308, across 29 blocks in 5 districts as our "study universe".[9] Second, we randomized 51 of the 153 clusters into treatment status and 102 into control (stratified within administrative blocks). This resulted in a total of 1774 elementary schools which were randomly assigned to treatment, and 3661 elementary schools (Grades 1–8) assigned to the control group.

We did not collect independent baseline data before randomization, relying instead on detailed administrative data (which includes test scores from an annual state-level standardized test). Table 1 shows that pre-program characteristics observed in the administrative data are balanced between the treatment and control groups. We also cannot reject equality of distributions of school-level test scores from March 2012-13 administered in all elementary schools (see appendix Figure C.1).[10] The randomization was completed in August 2014. The government notified treatment schools on 9th September, followed by the training of the school inspectors. School inspections and evaluations were conducted primarily in late-September 2014.

Our study design aimed to improve the external validity and policy relevance of our results by combining random assignment in a sample that is representative of a large population, implementation at scale by government, and randomizing in large units (Al-Ubaydli et al., 2017; Muralidharan and Niehaus, 2017). Conducting experimental evaluations in near-representative samples reduces the risk of site-selection bias (Allcott, 2015).[11] Evaluating large-scale implementation is relevant for policy because effect sizes have been shown to decline with size of implementation (Vivalt, 2020). Finally, randomizing large units into treatment and control

[8]Spillovers could, in principle, be positive (if CRCs applied the training under the intervention to control schools) or negative (if CRCs diverted time and effort away from control to treatment schools) with the potential to bias results from school-level randomization in an unknown direction.

[9]We did this because the government was considering extending the program in the second year while the study was underway. This design allowed them to do so in the non-study universe (the other 155 clusters) without affecting the experiment validity in the study population and at no cost to study objectives or government targets. While we use the English term "cluster" throughout the paper, these clusters are locally referred to as *Jan Shiksha Kendras* or JSKs.

[10]The existence of baseline administrative test-score data allowed us to conduct the experimental evaluation without an independent baseline. As discussed in Muralidharan (2017), such an approach also ensured a prudent use of research funds given the risk of either non-implementation or non-compliance with the RCT protocol by the government. Given the scale of the implementation (which was much larger than the evaluation sample), statistical power was higher in a design that used these resources to increase the sample size of schools that would be tested at the end-line. This is what we did here, resulting in a large study sample of over 300 schools.

[11]In the context of experimental evaluations of management interventions, only 25% of firms approached by Bloom et al. (2013) chose to participate in the trial, even though the (expensive) consulting services were being provided for free. Similarly the experiment conducted by Bruhn, Karlan, and Schoar (2018) was also in a sample of firms that had expressed interest in the program. Thus, results in this universe of motivated firms may be stronger than in the full population of firms.

status helps produce estimates that are inclusive of spillovers, which have been shown to be salient for policy in several studies (see, e.g., Cunha, De Giorgi, and Jayachandran (2019), and Muralidharan, Niehaus, and Sukhtankar (2020)).

This phase of program implementation also included secondary schools (N=116) in the clusters that were randomized into treatment; secondary schools in control clusters were untreated (N=273). Secondary schools were not, however, the focus of the program: they were not part of the initial piloting, nor were they included in later scale-ups in the state or the national policy. We also did not collect independent data on learning outcomes in secondary schools. We therefore focus on results from primary and middle schools. We did, however, collect one round of process monitoring data on secondary schools and matched schools to administrative data on secondary school (Grade 10) exams. For completeness, all results from secondary school are presented in Appendix C.2. We find no significant impact on either monitoring, teaching practices, or student test scores in secondary schools as well.

3.2 Data

We use both administrative data and extensive primary data on both learning and school processes. We collected primary data in a subset of 302 elementary schools, across the 153 study clusters. Schools were selected using simple random sampling in each cluster.[12] Specifically, we conducted three rounds of data collection on intermediate outcomes related to teacher absence and school functioning between September 2015 and February 2016 and conducted independent measurement of student achievement in March 2016. The main data sources are summarized below, with greater detail presented in Appendix D.

3.2.1 Student achievement

We use both administrative and independently-collected data on student achievement to evaluate program impacts. Administrative test score data in primary/middle schools come from the Pratibha Parv annual assessments which are administered to all students in Grades 1-8 in the public schooling system. These are available to us as school-level aggregates for all schools. Student-level data from these tests are not routinely digitized and only recorded in physical registers maintained in schools. We transcribed student-level test-score data from the assessments in 2014-15 and 2015-16 in the sub-sample of schools where we collected independent data.

In primary and middle schools, we conducted independent tests in mathematics and Hindi (the primary language of instruction) in February 2016, 18 months after the program was launched in treated clusters. These tests were designed to capture a wide range of variation

[12] We selected two schools each in 149 clusters. In the remaining 4 clusters, all in one district, we could only sample one school each due to misclassification in the administrative dataset being used as the sampling frame.

in student achievement and will serve as our primary outcome of interest. We privileged our independently-administered tests as core outcome measures to avoid the risk of test score manipulation and to ensure that our outcomes captured a broad range of skills.[13]

3.2.2 Digital records of school assessments

We obtained all the detailed scores (for each domain, and overall) from the school assessments by external inspectors and the CRC (for the 93% of schools in the treatment group where external assessments were conducted). We use this data to summarize the findings of these assessments, and to assess their quality.

3.2.3 School-level accountability and governance

Our primary metric for school level governance is teacher absence. This was collected over a set of three visits in each elementary school in our primary data collection subsample. Teacher absence is defined as being physically not present in school at the time of the visit. School visits were unannounced and staggered through the workday and all teachers are deemed to be absent if a school is found closed within working hours.

We also measured the extent and quality of school supervision. In interviews with headteachers and teachers, enumerators collected details of the extent of monitoring by CRCs and block-level officials, of the functioning of School Management Committees (SMCs) and whether a range of potential monitors (Block/cluster level officials, headmasters, parents/SMC members) had visited classrooms. In addition to establishing frequency of the visits, we also collected data on the content of these visits both through teacher reports and by noting down the comments made by cluster/block-level officials on physical inspection registers maintained in schools.

3.2.4 Classroom pedagogy

In two grades per elementary school in our subsample, enumerators observed one classroom period of teaching in Hindi and Mathematics. They collected information on the time-use of the teacher, whether they were using various instruction materials and/or changing other elements of classroom practice. For a (randomly-selected) subset of students present on the day, they also examined their homework notebooks to see if it had been checked by teachers and in what detail. We collected this as a cumulative measure of past teacher effort, which is less subject to Hawthorne effects than direct observation of pedagogy by surveyors.

[13] Both concerns are important in our setting. Manipulation of official test scores in this setting is high (Singh, 2020) and, since the program encourages student achievement, such manipulation could plausibly differ across treatment and control groups. A broad measure of learning is important because, in settings with substantial mismatch between curriculum and student preparation, even large program effects may not be evident in school exams which only focus on grade-specific syllabi (Muralidharan, Singh, and Ganimian, 2019).

3.3 Estimating equation

We estimate Intent-to-treat (ITT) effects using the following specification:

$$Y_{is} = \alpha + \beta_1.Treatment_s + \beta_2.X_s + \epsilon_{is} \quad (1)$$

where Y_{is} is the outcome for unit i (at the student/teacher/school level) in school s and $Treatment_s$ is a dummy variable for being assigned to treatment. X_s refers to pre-treatment controls, measured at the school level: we control for stratification (block) fixed effects and, when estimating treatment effects on student achievement, we additionally control for school-level lagged achievement in state-level standardized tests.[14] Standard errors are clustered at the level of randomization, i.e. the academic cluster.

4 Results

4.1 Implementation and informativeness of school assessments

We focus first on the diagnostic assessments of school quality. These make up the essential first step for this intervention but may be compromised by non-implementation, poor quality and uninformativeness, or collusion between school staff and inspectors. The diagnostic assessments rate schools in each of seven domains as being in one of four categories: "Above Standards", "Meets standards", "Close to Standards", and "Below standards". The assessment also provides a summary evaluation of each school on the same scale.

Table 2 presents the distribution of these ratings for all schools in our sample where the program was implemented. We note three key points. First, ratings are available for 1643 schools (∼93%) out of 1776 elementary schools assigned to treatment, indicating widespread implementation. Second, there is substantial variation across schools and across indicators in the ratings in whether schools are reported as meeting standards or not. Third, only a small minority of schools (∼9%) was assessed as performing at expected standards overall. 74% of schools were judged to be in the lowest possible category ("Below standards") in teaching and learning. The low absolute value of ratings suggests a lack of collusion between raters and school staff.

We evaluate the quality of these assessments by testing whether the variation in school ratings is able to meaningfully predict *future* outcomes, after controlling for prior test scores.[15] We consider

[14] These baseline scores are balanced across groups, both on average and across the distribution, and are used here for improving precision. Singh (2020) shows, using an independent audit, that administrative test scores in this setting retain ordinal information that is positive correlated over time, even though levels are overstated.

[15] Test scores serve as the default summary statistic for categorizing school performance and the additional effort in conducting comprehensive school assessments would only be justified if they provide useful additional information. Note that by construction, this exercise is limited to the treatment schools who had an assessment and is only meant to assess the quality of the external assessments and not measure treatment effects.

two measures of future performance (i) test scores on administrative exams in the 2014-15 school year (which were held after the assessments were conducted) and (ii) teacher absence, which we collect independently in unannounced visits but is not otherwise visible in the system overall.

Table 3 presents these results, and we see that even conditional on past test scores, schools with higher assessed grades ("Close to Standards" or "Meets standards") have higher future test scores than schools in the base category ("Below standards"). The difference between the bottom group and the top group ("Meets standards") is large at about one-quarter of a standard deviation and statistically significant. In our study sample, we further see that the rating that treatment schools receive is informative of teacher absence measured in the *next* academic year after the assessment. These results suggest that the design of the school ratings, done by GoMP and Ark together, was sound and captured useful variation in school quality (measured by expected future performance rather than just past performance). The variation in ratings across schools also suggests that the program approach of seeking to provide customised feedback (rather than a uniform recommendation for all schools) had conceptual merit.

These assessments were followed by the creation of the School Improvement Plans and we verified that these had been uploaded to the online portal for all schools. The recommendations in the School Improvement Plans differed (naturally) across schools but were comprehensive in covering different aspects of the school assessments.[16] The typical length of a SIP was ∼3-4 pages of bullet points, with each action attached to a person responsible within the school for implementation (usually the head-teacher, but occasionally other teachers or the SMC) and the person responsible for follow-up verification that the action had been implemented (usually the CRC).

4.2 Changes in governance and pedagogy

We next assess the extent to which the school assessments, and the resulting school improvement plans also helped to improve measures of school governance and classroom pedagogy.

4.2.1 Monitoring by officials and School Management Committees

We find no evidence of improvements in monitoring, by either higher officials at the cluster and block levels or by School Management Committees (SMCs). This result is seen in both administrative data on the full sample of schools, and in the independent data we collected in the evaluation subsample of 302 schools (Table 4). We find no impact on the frequency of monitoring visits by officials in either administrative or survey data. In the survey data, we also find no evidence that the inspectors spent any more time in the schools, indicating that

[16] Concrete recommendations could be, for instance, to ensure that students rated in the bottom two grades in the state standardized exams are provided supplementary remedial instruction, to organize staff meetings to review school progress or to host meetings with the parents of all students to increase community engagement.

the quality of monitoring also did not change. All schools report having an SMC, but we see little evidence of them having been more active or being seen to be more effective in treatment schools. We also did not see any evidence of a qualitative difference across treatment and control schools in the content of officially-recorded feedback in inspection registers maintained at the school level. Thus, by all available measures, we do not see any evidence of meaningful follow-up of the school ratings that would be expected to lead to a change in governance.

4.2.2 Teacher absence, classroom practice and student attendance

In Table 5, we first document that teacher absence is high (∼33%) in our setting. However, consistent with the null effect on improving accountability, we find that the intervention had no impact on teacher absence.

It is still possible that the intervention eased information constraints faced by teachers by providing diagnostic feedback to teachers directly, and helping them plan concrete actions. This could, in principle, have improved pedagogical practices. However, we find little evidence of change in pedagogical practices. Teachers in treatment schools do not appear to be more likely to use textbooks or workbooks, praise students or check student homework. They also do not seem to spend their time in class, during a direct observation, very differently: while there is some suggestive evidence that they substitute some time from lecturing to small group work, the magnitude of this change is very small (∼2% of a classroom period).

4.3 Changes in learning outcomes

Consistent with the lack of evidence of the intervention having changed any practices within the schools, we find no impact on student test scores in treatment schools (Table 6). This is evident using student-level data, both from our independently-administered tests in Math and Hindi (Cols 1-2) and in the administrative tests across all core subjects (Cols 3-7), and in school-level aggregated test scores for the full study population (Cols 8-9). Although we have only school-level aggregate test scores for the full sample, the size of the study population ensures that we are sufficiently powered to reject even modestly-positive treatment effects for most assessments. Fig 2 further compares the full distribution of student-level test scores across the treatment and control groups: the distributions overlap near-perfectly, and treatment effects are insignificant at almost every quintile (Table C.1).

Overall, our experimental evaluation of Phase I of the MP School Quality Assurance (MPSQA) program finds that the program was successful in conducting external school assessments, which were not subverted by collusion and captured meaningful variation. However, these ratings and the subsequent creation of School Improvement Plans did not translate into any observable change in practice or student learning outcomes.

5 Evaluating a further scale-up

As planned by GoMP, the Phase-I implementation of the MPSQA program was followed by a scale-up to ∼25,000 schools, across the state. Concurrently, the national government also strongly encouraged the development of such a school management program. This latter variant of the program, named *Shaala Siddhi* (roughly translated as "school success"), was designed for the full country by India's apex education policy institute (the National Institute of Education Policy and Administration or NIEPA). The program design was based on inputs from several states, including Madhya Pradesh, whose MPSQA program was considered to be a forerunner for the national program. The *Shaala Siddhi* program is now India's flagship school management program and has been scaled up to over 600,000 schools.[17] The Phase-II expansion of the school management program in MP was conducted as part of this nationwide program.

The national program is very similar in design and ambition to the original MPSQA program, differing in two main details. First, instead of using an external assessment for the initial rating, the national program required schools to complete the rating exercise themselves through a self-evaluation against the school ratings guidelines. These self-evaluations were meant to be verified by external assessors, but the aim was to shift emphasis from an inspection-focused system of school ratings to a reflective process of school improvement, where the schools were directly given responsibility for self-diagnosis and suggesting improvements. Second, the School Improvement Plan (SIP) format was made much more detailed, and required schools to list out an even more granular set of steps that they would take to improve. Appendix E provides an example of an SIP from this phase to illustrate the level of detail in an SIP.

In addition to the two changes above, GoMP also attempted to respond to our presenting findings that MPSQA had no impact on either teacher behaviors or student outcomes. While the Phase 2 expansion to ∼25,000 schools followed the national guidelines, the GoMP trainings on the program emphasized that schools should focus on three areas in executing their SIP : (i) teaching and learning, (ii) student learning outcomes, and (iii) developing a close collaboration with parents and the community. This change in emphasis was meant to communicate a need to move away from administrative compliance towards student outcomes. However, the basic structure of the program was unchanged. In particular, there was neither any independent measurement of learning nor any high-stakes consequences based on learning outcomes.

For the scale up (which commenced in Fall 2016), GoMP aimed to (a) purposively-target the intervention to schools they identified as high-priority and (b) cover all academic clusters (N=3181) in the state, including 8 schools in each cluster (JSK). While our primary focus in this paper is the experimental evaluation of MPSQA, we supplement this with a non-experimental

[17]The official program website (http://shaalasiddhi.niepa.ac.in/) reports, as of Sept 19, 2020, that 604,824 schools had undertaken self-evaluations in 2016-18, 632,034 schools in 2018-19, and 407,152 schools in 2019-20.

evaluation of the Phase-2 expansion. This allows us to present estimates of program impact over a 4-year period of iteration and scale-up, and offer more conclusive evidence on program impact. In addition to the 5 districts in the original evaluation, we augmented the study sample to include tribal blocks from 5 more districts in a different region of the state (Figure 1).[18]

Given the purposive selection of schools for this expansion, our evaluation strategy uses a matched-pair design which proceeds in four steps. First, we randomly selected 100 academic clusters (10 in each district). Second, we randomly sampled two untreated schools, from the population of all schools not already assigned to the program by the government, to serve as comparison schools. Third, we generated a propensity score into treatment for the full set of schools in these clusters based on pre-program administrative data on enrolment, school infrastructure and administrative test scores. Finally, we matched one treatment school in each cluster to one of the two potential comparison schools in the same academic cluster based on the minimum distance in propensity scores.[19]

The strategy above gives us 100 matched pairs, each with a treated and a comparison school within the same cluster. Importantly, these pairs were identified *before* collecting data on student learning outcomes to assess program impact. Overall, schools selected by GoMP for Phase-2 of the program were considerably better off on observable characteristics than non-program schools (Table 7, Panel A, columns 1-3). This reflected a policy of prioritizing larger schools for the program. Thus, program schools were more likely to be middle schools, and had larger enrollment, more teachers, better facilities, and higher baseline test scores. However, in our matched-pair sample, treatment and comparison schools are similar on most observables (columns 5-7).[20]

The key point for the credibility of our identification strategy is that there is no difference in pre-program school-level test scores on the annual state-wide tests (Pratibha Parv) conducted in every school and every grade (Panel B, columns 5-7). In constructing our matched-pair sample, we verified that there was no difference in these test scores in *any* of the five years before the program was implemented, and also find no evidence of differential trends in test scores across the treated and comparison schools over these five years.

After identifying this matched sample and verifying balance on levels and trends in prior test scores, we collected independent data on learning outcomes in February 2018 (18 months after

[18]Since the initial 5 districts were adjacent to the state capital, we aimed to further improve external validity by including locations (tribal areas) where the population was more disadvantaged, and where program implementation may be more challenging.

[19]We had hoped to also conduct an experimental evaluation of the Phase-2 scale up and had also randomly sampled two further schools in each academic cluster that were to be added to the full program roll-out by the government. Unfortunately, GoMP was not able to dedicate resources to treating these additional schools over and above the ∼25,000 school roll-out. This is why our evaluation strategy uses the randomly-selected control schools and then matches them to the most similar treatment school in the same cluster.

[20]The only exception is that treated schools are still more likely to be middle schools. This was unavoidable since the treated universe had a much larger fraction of middle schools.

the program was notified in treated schools, and ∼15 months after the self-assessments and school improvement plans were prepared). We used very similar test instruments and procedures as the first experimental evaluation. In addition to administrative data on school-level lagged test scores, we also collected student-level data on these scores from the previous school year (2016-17), to include as controls for estimating program impact.

As before, we estimate ITT effects, by regressing student test scores in February 2018 on the treatment status of their school. Our estimating equation is:

$$Y_{igsp} = \alpha_g + \phi_p + \beta_1.Treatment_s + \beta_2.X_{is} + \epsilon_{is} \tag{2}$$

where Y_{igsp} is the test score of student i; g, s and p index grade, school and cluster respectively. We include fixed effects for grade (α_g) to account for the imbalance in the number of middle schools, and for the academic cluster (ϕ_p), which is the stratum within which the matching was done. $Treatment$ is an indicator for whether a school was selected for the program, and X_{is} is a control for lagged test scores. We report results from three regression specifications: (a) with no controls for lagged scores, (b) controlling for lagged achievement at the school level from 2015-16, i.e. the year before the program assignment and, (c) controlling for individual-level student test scores in the state standardized tests in 2016-17, which is done soon after program notification and training.[21] Standard errors are clustered at the academic cluster level. We present results separately for mathematics and Hindi, the two subjects assessed in our independent tests in 2018.

We present estimates on program impact in Table 8. Regardless of specification, we find no evidence of the program having improved student outcomes. Point estimates of program impact are close to zero when we incorporate individual-level controls for baseline achievement (cols. 3 and 6). Given the positive (purposive) selection of schools into the program, our preferred specifications are those which control for lagged achievement, since these may better control for any bias that remains even after the matching procedure. Estimates are sufficiently precise to rule out even modest effects regardless of specification.

6 Understanding the reasons for program failure

On paper, the program we evaluate had several promising features that should have made it likely to succeed. It had a coherent program design that reflected both management theory and inputs from several global experts, substantial buy-in from both the political and bureaucratic leadership, and was the flagship education reform in the state. Yet, we find that it had no impact

[21] By the time of the endline data collection, which is when we collected student-level administrative test score data, it was not possible to get registers from the assessment in 2015-16.

on either the functioning of schools or student achievement (in either Phase-I or the scaled up Phase-II). We now turn to understanding reasons for program failure.

To better understand how the program was implemented and perceived at the field level, we collected extensive qualitative data through semi-structured open-ended interviews of school staff and education officials in 6 districts during the Phase-II scale up. In each of the 6 districts, we randomly sampled three schools: one from the universe of *Shaala Siddhi* schools, one more from the list of "champion schools" which were designated by the government as effective implementers of the program, and one control school (from our matched-pair sample) to understand business-as-usual constraints to school effectiveness, and the extent to which the intervention alleviated these. In each district, we also randomly sampled one academic cluster (JSK) office and one Block Education Office, where we interviewed relevant education officials who are responsible for implementing the program. We highlight key insights below.[22]

6.1 How did the program actually get implemented?

Our interviews confirm the implementation of the initial parts of the program. In most cases, teachers knew about the program, could summarize its core objectives and recalled the process of creating self-evaluation reports and uploading school improvement plans.

They do not, however, see it as leading up to action related to greater accountability or better pedagogy. Rather, they primarily recall the program as a source of paperwork (and logistics associated with uploading reports). As summarized by one respondent:

> "There is a lot of documentation work. We have to make a work plan and then upload it, get it printed. There is so much paper work that by the time some teachers figured that out they had forgotten what was *Shaala Siddhi* itself. I do all the documentation work at home because I have no time in the school."
> *(Headteacher, Champion School)*

When asked to assess what the effect of the program was, multiple respondents spoke about any momentum in the program having dissipated after the preparation of work plans. No teacher or headteacher in a *Shaala Siddhi* program school mentioned any change in accountability relationships, including in monitoring or support by the CRC, after the school evaluations. This assessment of ineffectiveness is reported by school staff and even some cluster officials.

[22]To avoid cherry-picking cases, we interviewed a random representative sample of respondents. However, since the aim of this section is to complement our quantitative evaluation as opposed to writing a dedicated qualitative paper, we do not formally code all the responses from the interview transcripts. Rather, we present what we consider to be the key insights from our review of all the interviews.

> For one year we did everything we were asked to. After that nobody asked us to do anything in the program. So we did not follow up on that at all. I don't remember what I did in the program two years ago. Those files aren't kept properly either. Nothing has happened in the last two years at all. They gave a training once and then forgot about it. There should be some follow up every couple of months. (*Teacher, Shaala Siddhi School*)

> The program initially created a lot of excitement and had that energy continued, the program would have been succesful. There were no additional funds to prepare the files either. Yet the schools had prepared the files. Also, in the last four months nothing has happened in *Shaala Siddhi*. There hasn't been much of an effect of *Shaala Siddhi*. Some 3-4 teachers have been involved in this and they sat and made all the files. During this time, the classes were left unattended and the teaching in the classroom suffered. So there has been no effect of the program. (*Cluster Academic Coordinator*)

In terms of regular functioning, schools and teachers repeatedly mentioned weaknesses in pedagogical support and accountability in schools. However, we found no instances of teachers reporting that this changed as a result of *Shaala Siddhi*. As such, even if the quality of the assessment reports was good, the main problem it identified (of weakness in teaching and learning as seen in Table 2) was already well known to teachers, and there was no follow-up to improve either governance or pedagogy.

In practice, the main consequence of *Shaala Siddhi* perceived on the ground was the increased paperwork required to fill in the school assessments and improvement plans. With no impetus for continuous improvement, it is unsurprising that the program did not change either teaching practice or learning outcomes. In this vein, our results reinforce the findings of Berry et al. (2020) and Muralidharan and Sundararaman (2010) who find that using low-stakes assessments aimed at producing continuous improvement in teaching had no impact on student learning outcomes.

6.2 Why did the program fail (and yet be perceived as successful)?

The in-depth interviews confirm our main results that, despite considerable effort having gone into the assessments and the creation of school improvement plans, this did not translate into changes in school practice or student learning outcomes. In particular, the key place where the theory of change appears to have broken down is at the point of conducting follow-ups to support and monitor schools in achieving the goals outlined in their school improvement plans.

The finding that a program may have failed due to implementation weaknesses is not a new one, and has been documented in several developing country settings. However, what is striking

in our setting is that the program was *perceived* as successful by senior officials and continues to be scaled up. The interviews highlight three broad reasons for program failure, and one key (related) reason for the divergence between perception and reality.

The first reason for failure was the disconnect between the objectives of the program and how it was actually perceived by those implementing it. The theory-of-change prioritized not just assessment, but also further support and self-improvement, with a particular focus on making changes that were within the schools' own control. In practice, however, the program was reduced to an exercise in record-keeping and the focus was on completing paperwork and submitting assessment reports and improvement plans as opposed to doing anything about these.

The second, and related, reason was a disconnect between the role that the program attached to education officials and their actual role as perceived by other agents in the system. Whereas the intervention was premised on CRCs playing a role of monitoring, accountability and coaching for schools, in practice they are seen mainly as conduits for communication, especially of paperwork, from schools to the bureaucracy.[23] Similar evidence from other states is presented in Aiyar and Bhattacharya (2016), who characterize this phenomenon as a "post office state" where lower-level officials are primarily used for transmitting "orders downwards" and "data upwards".

The CRCs also lack the administrative ability to hold schools to account effectively. Block-level officials, in contrast, do have more administrative tools to promote accountability at their disposal but are too dispersed to be effective (an average block includes 300-400 schools), especially given competing demands on their limited time. Thus, program failure was also driven in part by the lack of investment in sufficient managerial staff and capacity in the system to perform the additional functions that the program was premised on. Our qualitative and quantitative findings thus provide a direct example of a "flailing state" that can design sophisticated programs on paper, but lacks the capacity to implement them effectively (Pritchett, 2009).

The final reason relates to the field-level consequences of frequent changes in education policy and programmatic priorities (often due to transfers of senior officials). This leads to a (real and perceived) lack of coherence across initiatives, and a corresponding lack of engagement by implementing staff. Respondents perceived, based on experience, that government policies are highly transient, are often designed without considering implementation constraints, and frequently abandoned or altered.[24] As a result, they do not have any lasting effects.

[23] As summarized by one of our respondents: "We get no support from the Jan Shikshak. Their job is mainly delivering letters and some academic work. They don't even come to deliver the letters and call us to the office; forget about academic work. I have never seen any cluster principal come to the school. If they need any teacher from us then they call us and take them. But they never help us with anything."

[24] Although the program was developed and iterated over 5 years, the three phases of program implementation happened mostly in different samples. Nearly all schools in the ∼25,000 school scale-up would have only experienced the program once, as a short burst of activity focused on completing the self-assessments after which there were no follow-ups. Combined with a general proliferation of short-lived reform initiatives in education,

> So firstly, government should run all the policies for a longer time for them to have a real effect. [...] If they work on one project consistently then over a period of time they will achieve that. By the time the program starts achieving something they switch to something else. They think that the children aren't learning enough, education quality is falling and change the program. But if quality is the goal of the program, then it should be looked at in the long run. [. . .] The government is patting us on the back through their policies and we show them achievement on paper by reporting so. On paper, all policies are a success. (*Teacher, Shaala Siddhi School*)

The discussion above also sheds light on the reasons for divergence between the perceived and actual success of the program. In particular, the qualitative interviews highlight that bureaucratic incentives are geared more towards the *appearance of activity* rather than actual impact. Thus, the successful completion of school assessments and uploading of school improvement plans at scale was the main aspect of the program that was being monitored. On these metrics, the program was a success.

These features of government implementation extend broadly beyond education. As summarized in a classic ethnographic study of the Indian state (Gupta, 2012, p. 48, emphasis added):

> What stands out here are higher-level officials in the administrative hierarchy making decisions about programs and targets that bear little relevance to realities on the ground; also present, in turn, are subordinates *faithfully executing programs on paper but caring little for how well they are implemented.* Targets are indeed met, but the ultimate goals of the programs go unfulfilled.

Higher-level officials are likely to be aware of these constraints. Despite this, such reforms continue to be designed and implemented widely, and proclaimed as successes. Why might this be the case? One compelling explanation is the idea of "institutional isomorphism" pioneered by DiMaggio and Powell (1983) and expanded to developing country bureaucracies by Andrews, Pritchett, and Woolcock (2017). Several direct quotes from DiMaggio and Powell (1983) apply well to our setting. For instance, they note that:

> "Organizations tend to model themselves after similar organisations in their field that they perceive to be more legitimate or successful," (p. 152); "these institutional isomorphic processes can be expected to proceed [even] in the absence of evidence that they increase organisational efficiency," (p.153); and that such mimicry has "a ritual aspect; [organizations] adopt these "innovations" to enhance their legitimacy, to demonstrate they are at least trying to improve." (p. 151)

this view is perhaps unsurprising.

The idea of institutional isomorphism provides an organizational sociology explanation for the persistence of ineffective policies. This phenomenon is also consistent with some models of political incentives. For instance, Majumdar and Mukand (2004) provide a formal model where policymakers may be rewarded for initiating new policies, but have incentives to persist with policies that turn out to be ineffective because changing track may signal low competence, and generate reputational costs. These political incentives may also explain the pressures on officials to ensure that programs are *seen* to be effective regardless of their actual impact.

Thus, a key contribution of our well-identified evaluation(s) is to demonstrate the distinction between a program "meeting targets" and appearing successful on paper, and actually improving outcomes of interest. Doing so is especially important since the program has been scaled up to over 600,000 schools in India alone, and the essential features of the program are similar to other such school management initiatives around the world (as shown in Appendix A).

7 Discussion and Conclusion

This paper evaluates a comprehensive and ambitious program to improve school management in the public sector in multiple implementations at increasing orders of scale over a four-year period. The program was (and continues to be) a flagship national reform. It was designed with detailed national and international technical assistance, commanded substantial support in the higher levels of the bureaucracy, and exemplifies global best practice in a number of areas. Yet, we find no evidence that it improved either school functioning or student outcomes. Rather than starting a reflective process of self-improvement, as envisaged by the program design, the program was in practice, reduced to the demonstration of administrative compliance with the process of conducting school assessments and creating school improvement plans.

These results are directly relevant for the continuing national scale-up of this particular program, which at full scale would incur the costs of conducting assessments and improvement plans in 1.6 million schools annually. We conservatively estimate that the *annual* cost of this exercise is over 35,000 teacher-years of time and over USD 235 million.[25] Similar reforms are common in

[25]Since no additional staff were hired, the costs of the intervention depend on the opportunity cost of staff time. Based on our field interviews, we estimate that the program required 5 full working days of teacher time per school just for the training, filling out of the assessments, creating the school improvement plans and uploading them, even if no further action is taken. At the projected scale of 1.6 million schools, this translates into a total time cost of 8 million days, and 36,363 years (at 220 working days/year). Using an average teacher salary cost of INR 40,000 per month (∼545 US dollars), this translates into an estimated 237 million USD per year. This estimate is conservative: it does not account for the cost of pensions or benefits, or adjust for well-documented teacher absence (the cost per working day would be higher after accounting for absence rates). It also does not include the time and attention cost of senior officials, and district and block level staff, which may in fact be much more constrained than teacher time.

education systems around the world, including other settings with weak governance.[26] While our results do not imply that *all* such programs will be ineffective, they do shift the burden of proof. At the very least, they suggest a need for caution in scaling similar interventions without credible evidence of impact at smaller scales. Note that our results do not imply that management does not matter for school quality or that it cannot be improved. Rather, they highlight the challenge of doing so at scale in public systems.

We cannot speak directly to what features *would* have led the reform to succeed. As is well known, complex projects – such as improving the functioning of government departments at scale – require several components to be executed successfully (Kremer, 1993). Thus, a failed reform could reflect a failure of one or more components, which cannot be econometrically identified. Yet, reflecting on our results in conjunction with evidence from other successful attempts at improving the quality of public service delivery in developing countries, we posit that there are three key factors in programs that have successfully improved public service delivery at scale.

The first is better incentives for improving either effort or outcomes. A large body of evidence has documented the positive effects of well-designed interventions to improve incentives of public-sector workers ranging from teachers (Muralidharan and Sundararaman, 2011; Mbiti et al., 2019; Leaver et al., 2021) to tax collectors (Khan, Khwaja, and Olken, 2016, 2019) and policemen (Banerjee et al., 2021).[27] Conversely, interventions that focus primarily on inputs have had much less success, including expensive reforms such as unconditional increases in teacher pay (De Ree et al., 2018) as well as several school grant programs (Glewwe and Muralidharan, 2016). The failure of the program we study adds to the evidence base on what works to improve public service delivery at scale, and reinforces the difficulty of improving outcomes without changing the incentives of front-line staff and supervisors to do so. In particular, it is important to note that implementation quality is endogenous to the incentives of agents to implement the program. Thus, the extent to which the program was only partially implemented may itself reflect the nature of bureaucratic incentives to successfully implement the program components that were monitored, while not doing so for those that were not.

A second, related but distinct, factor that may help is better visibility on outcomes at the beneficiary level. Even without formal incentives, senior officials do monitor program performance. The problem is that they can only do so based on what they observe. It is

[26] See, for instance, the description of the Whole School Evaluation program in South Africa described in Levy et al. (2018): the program was based on the UK's Ofsted evaluations and, as in our setting, was followed by a program focused on a first step of self-evaluation by teachers but without any external testing of students or any formal accountability. See Appendix A for a list of projects with similar components in various countries, all intended to improve school management.

[27] As shown by Khan, Khwaja, and Olken (2019) and Banerjee et al. (2021), incentives need not be restricted to financial rewards, but can also include options that are more feasible in civil-service settings such as providing more attractive job postings to higher-performing personnel.

noteworthy that the program we studied worked till the point where outcomes were visible to senior officials (school assessments were completed, and school improvement plans uploaded), but stopped working at the point where outcomes were no longer easily visible (classroom effort, and learning outcomes). One common pattern in recent interventions that have improved service delivery at scale is that senior officials were able to access *independent* data on program performance at the beneficiary level, and monitor lower-level supervisors on these metrics.[28] Thus, investing in better measurement and integrity of outcome data may yield high public returns by enabling better monitoring of service delivery (Singh, 2020).

A third factor is staffing. This program, like many others in India, added responsibilities to a supervisory bureaucracy that is already over-burdened and understaffed relative to their expected workload (Kapur, 2020; Dasgupta and Kapur, 2020). The importance of dedicated program staff is illustrated by Dunsch et al. (2017) who report that a management intervention with intensive follow-up and support facilitated by staff from an external agency, was able to improve practices in healthcare centres in Nigeria. However, a lighter-touch intervention which provided only information and diagnostic feedback did not. Thus, it is possible that the program we study may have been more effective if it had augmented staff capacity to conduct the follow-up visits to schools to monitor progress against the goals laid out in the School Improvement Plans.[29]

Of course, while the discussion above reflects our views based on evidence to date, these factors are neither necessary nor sufficient for successfully improving service delivery at scale. Improving state capacity for service delivery is an active area of ongoing research and several promising new approaches are being designed and tested with evidence of positive impacts from approaches such as increasing autonomy for front-line staff (Bandiera et al., 2020) and improving training (Azulai et al., 2020; Banerjee et al., 2021). There may also be complementarities across components of bureaucratic effectiveness such as staffing, training, autonomy, measurement and monitoring, and incentives/accountability. These are important areas for future research.

More generally, governments and donors are constantly designing and deploying programs to improve service delivery and development outcomes, and senior officials and staff often exert considerable good-faith efforts in doing so. These programs, which are often funded by donors

[28]For instance Muralidharan, Niehaus, and Sukhtankar (2016) show large improvements in social welfare payments from a reform that digitized payments and made data on payment delays visible to senior officials. These delays could then be monitored on a regular basis. In a related vein, Muralidharan et al. (2021) and Callen et al. (2020) show that technologies like call-centres and customer-focused apps providing real-time data, which allow senior officials to crowd-source information from final beneficiaries, may significantly increase monitoring and improve service delivery even in the absence of formal incentives.

[29]This is consistent with qualitative evidence that successful at-scale reforms in India have featured the creation of program-specific staff capacity for implementation. For example, the success in consistently providing cooked school meals in Indian public schools, even though many other routine functions fail, may be due to the hiring of dedicated staff for cooking and delivering meals (Chakraborty and Jayaraman, 2019; Singh, Park, and Dercon, 2014). Similarly, the successful biometric smartcards payment program studied by Muralidharan, Niehaus, and Sukhtankar (2016) augmented staffing by deploying additional locally-hired staff to make payments in villages.

and designed by experts and consultants, are typically judged based on the extent to which the program design reflects (perceived) international "best practices", whether it was done in partnership with government, and how many beneficiaries it reached. By all of these metrics, the program we study was a resounding success, and the continued scale up of the program to 600,000 schools reflects the official belief that the program was a success.

Yet, as we demonstrate, no eventual outcomes changed. Our results therefore provide experimental evidence to support the view expressed by scholars of foreign aid (using qualitative evidence of how aid-funded projects actually work) that there are severe agency problems in how foreign aid is implemented in practice (Martens et al., 2002). Thus, a broader lesson from our study and results is to highlight the importance of disciplining the foreign aid "industry" (and the consulting firms and experts who are part of it) with credible evaluations of their impacts on ultimate outcomes of interest. Doing so may help improve the effectiveness of billions of dollars of government and donor aid expenditure on service delivery in developing countries.

One practical way of moving towards this goal may be to link the volume of donor assistance to evidence of positive impact at progressively larger scales. This is the approach taken by recent initiatives such as the *Development Innovation Ventures* (DIV) and the *Fund for Innovation and Development* launched by USAID and the French aid agency (AFD) respectively.[30] Unfortunately, these innovations are still the exception, with the default approach being for governments and donor agencies to scale programs without any evidence of effectiveness. The "push to scale" in turn reflects political, bureaucratic, and donor incentives to *appear* to have gotten things done during their limited tenures in office. Our results suggest that rushing to scale without adequate evidence may often not be in the public interest.[31]

References

Abadie, Alberto. 2020. "Statistical Nonsignificance in Empirical Economics." *American Economic Review: Insights* 2 (2):193–208. 1

Aiyar, Yamini and Shrayana Bhattacharya. 2016. "The post office paradox." *Economic & Political Weekly* 51 (11):61. 6.2

Aiyar, Yamini, Ambrish A Dongre, and Vincy Davis. 2015. "Education Reforms, Bureaucracy and the Puzzles of Implementation: A Case Study from Bihar." *Bureaucracy and the Puzzles of Implementation: A Case Study from Bihar (September 1, 2015)* . 1

[30]The creation of these initiatives have in turn been supported by Michael Kremer and Esther Duflo, whose work has over the years highlighted the need for better evidence to inform policy scale up decisions.

[31]Of course, evidence of program efficacy at small scale does not imply effectiveness at larger scales. Recent studies have shown that program effect sizes are smaller when they are implemented at larger scale (Vivalt, 2020). See List (2022) for a discussion of ideas for ensuring that programs continue to be effective at larger scales.

Al-Ubaydli, Omar, John A. List, Danielle LoRe, and Dana Suskind. 2017. "Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature." *Journal of Economic Perspectives* 31 (4):125–44. URL https://www.aeaweb.org/articles?id=10.1257/jep.31.4.125. 3.1

Al-Ubaydli, Omar, John A List, and Dana L Suskind. 2017. "What can we learn from experiments? Understanding the threats to the scalability of experimental results." *American Economic Review* 107 (5):282–86. 1

Allcott, Hunt. 2015. "Site selection bias in program evaluation." *The Quarterly Journal of Economics* 130 (3):1117–1165. 3.1

Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2017. *Building state capability: Evidence, analysis, action*. Oxford University Press. 1, 6.2

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K Khandelwal, and Eric Verhoogen. 2017. "Organizational barriers to technology adoption: Evidence from soccer-ball producers in Pakistan." *The Quarterly Journal of Economics* 132 (3):1101–1164. 1

Azulai, Michel, Imran Rasul, Daniel Rogger, and Martin Williams. 2020. "Can training improve organizational culture? Experimental evidence from Ghana's civil service." Tech. rep., University College London. 7

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat. 2020. "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats." Tech. rep., National Bureau of Economic Research. 7

Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh. 2021. "Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy, and training." *American Economic Journal: Economic Policy* 13 (1):36–66. 7, 27

Banerjee, Abhijit V, Esther Duflo, and Rachel Glennerster. 2008. "Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system." *Journal of the European Economic Association* 6 (2-3):487–500. 3

Berry, James, Harini Kannan, Shobhini Mukherji, and Marc Shotland. 2020. "Failure of frequent assessment: An evaluation of India's continuous and comprehensive evaluation program." *Journal of Development Economics* 143:102406. 6.1

Blimpo, Moussa Pouguinimpo, Moussa Blimpo, David Evans, and Nathalie Lahire. 2015. "Parental human capital and effective school management: evidence from The Gambia." *World Bank Policy Research Working Paper* (7238). 2

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does management matter? Evidence from India." *The Quarterly Journal of Economics* 128 (1):1–51. 1, 11

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2015. "Does management matter in schools?" *The Economic Journal* 125 (584):647–674. 1

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar. 2018. "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in mexico." *Journal of Political Economy* 126 (2):635–687. 1, 11

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee. 2020. "Data and policy decisions: Experimental evidence from Pakistan." *Journal of Development Economics* 146:102523. 28

Chakraborty, Tanika and Rajshri Jayaraman. 2019. "School feeding and learning achievement: Evidence from India's midday meal program." *Journal of Development Economics* 139:249–265. 29

Cunha, Jesse M, Giacomo De Giorgi, and Seema Jayachandran. 2019. "The price effects of cash versus in-kind transfers." *The Review of Economic Studies* 86 (1):240–281. 3.1

Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan. 2016. "Quality and Accountability in Health Care delivery: audit-study evidence from primary care in India." *American Economic Review* 106 (12):3765–99. 1

Dasgupta, Aditya and Devesh Kapur. 2020. "The Political Economy of Bureaucratic Overload: Evidence from Rural Development Officials in India." *American Political Science Review* 114. 7

De Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers. 2018. "Double for nothing? Experimental evidence on an unconditional teacher salary increase in Indonesia." *Quarterly Journal of Economics* 133 (2):993–1039. 7

Dhaliwal, Iqbal and Rema Hanna. 2017. "The devil is in the details: The successes and limitations of bureaucratic reform in India." *Journal of Development Economics* 124:1–21. 3

DiMaggio, Paul J and Walter W Powell. 1983. "The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields." *American Sociological Review* :147–160. 1, 6.2

Dunsch, Felipe A, David K Evans, Ezinne Eze-Ajoku, and Mario Macis. 2017. "Management, supervision, and health care: a field experiment." Tech. rep., National Bureau of Economic Research. 7

Easterly, William. 2007. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill And So Little Good*. Oxford University Press. URL https://EconPapers.repec.org/RePEc:oxp:obooks:9780199226115. 1

Easterly, William R. 2002. *The elusive quest for growth: economists' adventures and misadventures in the tropics*. MIT press. 1

Figlio, David and Susanna Loeb. 2011. "School accountability." In *Handbook of the Economics of Education*, vol. 3. Elsevier, 383–421. 1

Figlio, David N and Cecilia Elena Rouse. 2006. "Do accountability and voucher threats improve low-performing schools?" *Journal of Public Economics* 90 (1-2):239–255. 1

Gibbons, Robert and Rebecca Henderson. 2012. "Relational contracts and organizational capabilities." *Organization Science* 23 (5):1350–1364. 1

Glewwe, Paul and Karthik Muralidharan. 2016. "Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications." In *Handbook of the Economics of Education*, vol. 5. Elsevier, 653–743. 7

Gupta, Akhil. 2012. *Red tape: Bureaucracy, structural violence, and poverty in India*. Duke University Press. 1, 6.2

Hussain, Iftikhar. 2015. "Subjective performance evaluation in the public sector evidence from school inspections." *Journal of Human Resources* 50 (1):189–221. 1, 6

Iacovone, Leonardo, William Maloney, and David McKenzie. 2022. "Improving management with individual and group-based consulting: Results from a randomized experiment in colombia." *The Review of Economic Studies* 89 (1):346–371. 1

Kapur, Devesh. 2020. "Why Does the Indian State Both Fail and Succeed?" *Journal of Economic Perspectives* 34 (1):31–54. 7

Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken. 2016. "Tax farming redux: Experimental evidence on performance pay for tax collectors." *The Quarterly Journal of Economics* 131 (1):219–271. 1, 7

Khan, Adnan Q, Asim Ijaz Khwaja, and Benjamin A Olken. 2019. "Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings." *American Economic Review* 109 (1):237–70. 1, 7, 27

Kremer, Michael. 1993. "The O-ring theory of economic development." *The Quarterly Journal of Economics* 108 (3):551–575. 7

Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin. 2021. "Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools." *American Economic Review* 111 (7):2213–46. 1, 7

Lemos, Renata, Karthik Muralidharan, and Daniela Scur. 2021. "Personnel Management and School Productivity: Evidence from India." Working Paper 28336, National Bureau of Economic Research. 1

Levy, Brian, Robert Cameron, Ursula Hoadley, and Vinothan Naidoo. 2018. *The politics and governance of basic education: A tale of two South African provinces*. Oxford University Press. 1, 26

List, John A. 2022. *The Voltage Effect*. Penguin UK. 31

Majumdar, Sumon and Sharun W Mukand. 2004. "Policy gambles." *American Economic Review* 94 (4):1207–1222. 6.2

Martens, Bertin, Uwe Mummert, Peter Murrell, and Paul Seabright. 2002. *The institutional economics of foreign aid*. Cambridge University Press. 1, 7

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. "Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania." *The Quarterly Journal of Economics* 134 (3):1627–1673. 1, 1, 7

Morgan, James and Jeffrey K Liker. 2006. *The Toyota product development system: integrating people, process, and technology*. Productivity press. 1

Muralidharan, Karthik. 2017. "Field experiments in education in developing countries." In *Handbook of economic field experiments*, vol. 2. Elsevier, 323–385. 10

Muralidharan, Karthik, Jishnu Das, Alaka Holla, and Aakash Mohpal. 2017. "The fiscal cost of weak governance: Evidence from teacher absence in India." *Journal of Public Economics* 145:116–135. 2.1

Muralidharan, Karthik and Paul Niehaus. 2017. "Experimentation at Scale." *Journal of Economic Perspectives* 31 (4):103–24. 1, 3.1

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2016. "Building state capacity: Evidence from Biometric Smartcards in India." *American Economic Review* 106 (10):2895–2929. 28, 29

———. 2020. "General equilibrium effects of (improving) public employment programs: Experimental evidence from india." Working Paper 23838, National Bureau of Economic Research. 3.1

Muralidharan, Karthik, Paul Niehaus, Sandip Sukhtankar, and Jeffrey Weaver. 2021. "Improving last-mile service delivery using phone-based monitoring." *American Economic Journal: Applied Economics* 13 (2):52–82. 28

Muralidharan, Karthik, Abhijeet Singh, and Alejandro Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *The American Economic Review* 109 (4):1426–60. 13

Muralidharan, Karthik and Venkatesh Sundararaman. 2010. "The impact of diagnostic feedback to teachers on student learning: experimental evidence from India." *The Economic Journal* 120 (546):F187–F203. 6.1

———. 2011. "Teacher performance pay: Experimental evidence from India." *Journal of Political Economy* 119 (1):39–77. 1, 7

Pratham. 2017. *Annual Status of Education Report 2016*. Pratham, New Delhi. 2.1

Pritchett, Lant. 2009. "A review of Edward Luce's in spite of the Gods: the strange rise of modern India." *Journal of Economic Literature* 47 (3):771–80. 6.2

———. 2013. "The World Bank and public sector management: what next?" *International Review of Administrative Sciences* 79 (3):413–419. 1

Rasul, Imran and Daniel Rogger. 2018. "Management of bureaucrats and public service delivery: Evidence from the nigerian civil service." *The Economic Journal* 128 (608):413–446. 1

Rasul, Imran, Daniel Oliver Rogger, and Martin J Williams. 2018. "Management and Bureaucratic Effectiveness: Evidence from the Ghanaian Civil Service." Tech. rep., The World Bank. 1

Rockoff, Jonah and Lesley J Turner. 2010. "Short-run impacts of accountability on school quality." *American Economic Journal: Economic Policy* 2 (4):119–47. 1

Romero, Mauricio, Juan Bedoya, Monica Yanez-Pagans, Marcela Silveyra, and Rafael de Hoyos. 2022. "Direct vs indirect management training: Experimental evidence from schools in Mexico." *Journal of Development Economics* 154:102779. 2

Singh, Abhijeet. 2020. "Myths of Official Measurement: Auditing and Improving Administrative Data in Developing Countries." Tech. rep., RISE Programme, Oxford. 13, 14, 7

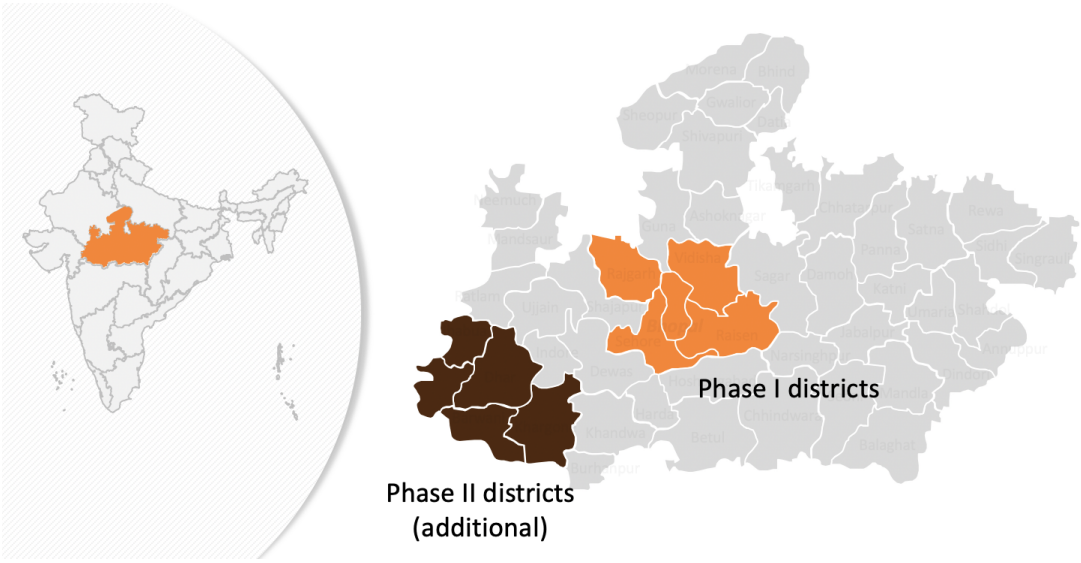
Singh, Abhijeet, Albert Park, and Stefan Dercon. 2014. "School meals as a safety net: an evaluation of the midday meal scheme in India." *Economic Development and Cultural Change* 62 (2):275–306. 29

Vivalt, Eva. 2020. "How much can we generalize from impact evaluations?" *Journal of the European Economic Association* . 3.1, 31

World Bank. 2003. *World development report 2004: making services work for poor people*. The World Bank. 1

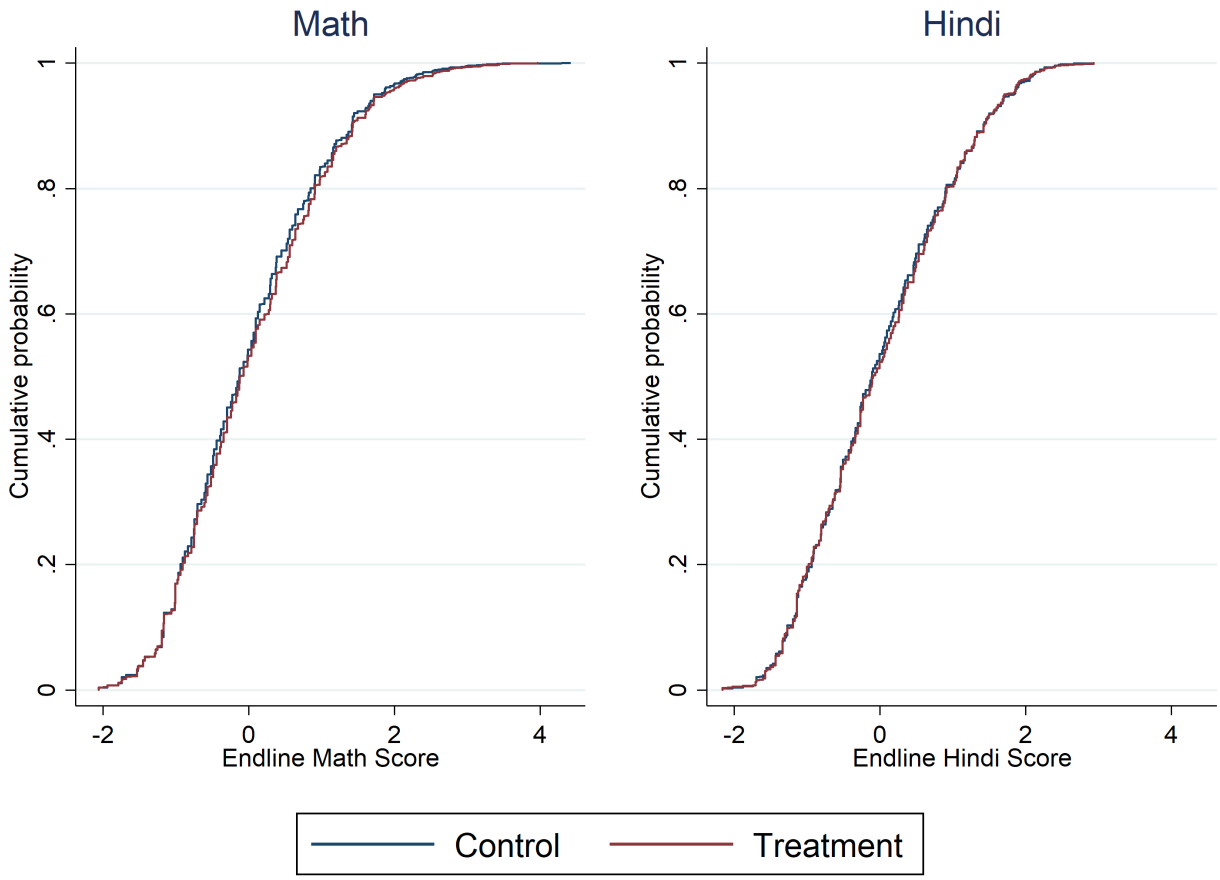
Figures and Tables

Figure 1: Districts in the MPSQA evaluation



Notes: This figure shows the five districts of the Bhopal region included in the Phase I (experimental) evaluation from Sept 2014 to March 2016 (in orange). The program was later scaled up across the whole state. We evaluated the scaled-up program in 2017-18 in the original five districts, plus tribal blocks from an additional five districts in the Indore region (in brown)

Figure 2: Distribution of student test scores at endline (Feb 2016)



Notes: This figure shows the distribution of individual student test scores for grades 1-8 in mathematics and Hindi from independent test data collection in February 2016. Test scores are standardized within grade with a control group mean of zero and standard deviation 1.

Table 1: Balance on observed characteristics

	All Study Schools				Subsample with primary data			
	Control Mean	Treatment Mean	Diff (T-C)	Diff SE	Control Mean	Treatment Mean	Diff (T-C)	Diff SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total enrolment	81.18	76.79	-5.59**	(2.17)	84.03	77.11	-7.14	(6.32)
No. of teachers	2.69	2.59	-0.11	(0.08)	2.71	2.67	-0.04	(0.19)
Pupil-teacher ratio	32.76	32.34	-0.73	(1.32)	33.39	31.77	-1.26	(2.74)
Proportion of Qualified teachers	0.91	0.90	-0.01	(0.01)	0.95	0.91	-0.03	(0.03)
Rural	0.92	0.93	0.01	(0.02)	0.91	0.89	-0.02	(0.03)
Electricity	0.15	0.13	-0.01	(0.02)	0.16	0.16	0.02	(0.04)
Visits from Block/Cluster officials	10.36	10.38	0.10	(0.34)	10.15	10.51	0.39	(0.77)
State test score (Pratibha Parv)	66.81	67.04	-0.06	(0.63)	67.05	66.39	-0.76	(1.52)
Observations	3661	1774			202	100		
F-test (p-value)			0.177				0.334	
F-test, number of schools			5084				280	

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent critical level. Differences in means, and associated standard errors, are reported from regressions that incorporate block fixed effects and cluster standard errors at the academic cluster level (level of randomization). The F-test refers to a test of joint significance for all reported variables and reports the F-statistics. The number of observations for the F-tests are lower than the total number of observations due to missing data for some individual variables.

Sources: 2013-14 school-level administrative data from school census (DISE) and state standardized tests (Pratibha Parv).

Table 2: Ranking of treatment schools by school inspectors

	Percentage of schools				
STANDARDS	Below standards	Close to standards	Meets standards	Above standards	N
1: School Development/Mentoring	17.5	46.9	31.6	4	1643
2: Management	6.1	51.2	37.9	4.8	1643
3: Teaching and learning	74.3	25.4	0.3	0.1	1643
4: Support for students	6	37.5	51.9	4.6	1643
5: SMC and engagement with parents	11	45.3	29.8	13.8	1643
6: Academic Outcomes	28.2	62.6	8.9	0.2	1643
7: Personal and Social Outcomes	7.4	62.2	28.3	2.1	1643
Overall	**16.2**	**74.9**	**8.9**	**0**	**1643**

Notes: These ratings are taken from the administrative data on program implementation and include all treatment elementary schools for which data was available. The data indicate that school inspections, and the corresponding input of assessments into the online portal, was done for 93% of the elementary schools assigned to treatment. SMC refers to School Management Committees.

Table 3: Informativeness of school assessments

	School Test scores †† (2014-15)		Teacher attendance†††	
	(1)	(2)	(3)	(4)
Close to standard†	0.072	0.097	0.172***	0.077*
	(0.070)	(0.068)	(0.052)	(0.046)
Meets standard†	0.260***	0.301***	0.258***	0.141
	(0.091)	(0.093)	(0.081)	(0.087)
School test scores (2013-14)††	0.388***	0.409***	0.020	-0.001
	(0.033)	(0.031)	(0.020)	(0.021)
Constant	-0.060	-0.083	0.490***	0.567***
	(0.062)	(0.062)	(0.046)	(0.038)
Block fixed effects	No	Yes	No	Yes
Observations	1,642	1,642	95	95
R-squared	0.179	0.261	0.147	0.634

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent critical level. This table relates future student achievement and teacher absence, aggregated at the school level, to the grades that schools received in the MPSQA school evaluation and pre-treatment test scores. The reference category is schools which were rated "Below standards". Columns 1-2 use post-treatment administrative test scores as the dependent variable. Columns 3-4 use data on teacher absence collected in a subsample of treated schools. Standard errors are clustered at school level.

Sources: †Administrative data on school ratings from the online program implementation portal.
†† Administrative data on aggregate test scores in the Pratibha Parv assessments.
††† Primary data on teacher absence from direct interviewer observations collected in three rounds.

Table 4: Treatment effects on monitoring and SMC functioning

	Control Mean (1)	Treatment Mean (2)	Diff (T-C) (3)	Diff SE (4)
Inspections†				
Visits by Block/Cluster officials (Full sample)	9.52	9.88	0.49	(0.41)
Visits by Block/Cluster officials (Study sample)	9.12	9.60	0.70	(0.75)
Time of last visit††				
Within last month	0.36	0.42	0.08	(0.06)
Within last 2 months	0.64	0.62	-0.02	(0.06)
Within 6 months	0.77	0.77	-0.00	(0.05)
Within last year	0.83	0.88	0.04	(0.04)
Time spent by inspector††				
Less than 30 minutes	0.21	0.19	-0.02	(0.05)
More than 30 minutes	0.79	0.81	0.02	(0.05)
More than 1 hour	0.42	0.37	-0.06	(0.06)
More than 3 hours	0.02	0.06	0.03	(0.03)
SMC Functioning†††				
School has SMC/PTA	1.00	1.00	-	-
SMC/PTA found useful	0.78	0.85	0.06	(0.04)
Last SMC/PTA meeting†††				
Within last month	0.32	0.40	0.08	(0.06)
Within last 2 months	0.70	0.77	0.07	(0.05)
Within last 6 months	0.96	0.96	-0.00	(0.02)
Within last year	0.98	1.00	0.02**	(0.01)
Observations (Full population)	3597	1751		
Observations (Subsample)	202	100		

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent critical level. Differences in means between treatment and control group, and associated standard errors, are reported from regressions incorporating block fixed effects. Standard errors are clustered at the academic cluster level (level of randomization).

Sources:† 2015-16 School-level administrative data from DISE.
†† Official 2015- school inspection register records, transcribed during school visits.
††† Primary data based on interviews with two teachers per school. Data on time of inspector visits was collected for a subsample of 302 elementary schools. Data on time spent by the inspector was reported by the headmaster/school in-charge. SMCs are governing bodies comprising of representatives from the school staff, parents and local authorities which are intended to exert community-based accountability on schools.

Table 5: Effects on teacher attendance, pedagogy and student engagement

	Control		Treatment		Diff	Diff SE
	N	Mean	N	Mean	(T-C)	
	(1)	(2)	(3)	(4)	(5)	(6)
Attendance†						
Teacher Attendance	2070	0.67	966	0.65	-0.03	(0.03)
Student Attendance (school-level)	201	0.53	100	0.53	-0.00	(0.02)
Pedagogical inputs††						
Textbooks used during class	353	0.72	177	0.71	-0.01	(0.04)
Workbooks used during class	353	0.12	178	0.10	-0.02	(0.03)
Teacher praised students	334	0.43	174	0.44	0.00	(0.05)
Percent of class time spent on:††						
— Lecture	353	43.80	178	39.89	-4.28*	(2.45)
— Silent Work	352	4.90	178	4.42	-0.40	(1.45)
— Group Call	353	16.75	178	16.78	-0.00	(1.82)
— Small Group Work	349	0.86	177	2.26	1.43**	(0.71)
— Big Group Work	345	1.85	175	2.21	0.45	(0.99)
— Class Discipline	352	2.52	178	4.07	1.58	(1.03)
— Out of class	353	2.73	178	1.69	-1.03	(0.76)
Child has HW notebooks which are checked†††	1511	0.40	737	0.44	0.03	(0.04)

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent critical level. Differences in means between treatment and control group and associated standard errors are reported from regressions incorporating block fixed effects and clustering standard errors at the academic cluster level (level of randomization). Round indicators are included as covariates in the estimations for teacher attendance. Student attendance is reported as average percentage of students across three rounds of data collection at the school level and includes only schools that were open at the time of observation.

Sources: † Attendance was collected for individual teachers over three rounds of primary data collection in 2015-16. Where a school was found closed at the time of the inspection (always during business hours), all teachers are marked absent. Student attendance was collected at a school-level aggregate only.

†† 2015-16 Direct interviewer observation of up to two classrooms per school. Time spent on activities was based on direct observation by surveyors and recorded in five categories: not conducted at all, below 25%, 25-50%, 50%-75%, above 75%. We take the mid-point of the categories to assign the percentage values here.

††† 2015-16 Student-level primary data collected for a random sample of students.

Table 6: Treatment effect on student achievement

	Independent tests†		Administrative school tests††					School-level PP scores†††	
	Math	Hindi	Math	Hindi	English	Science	Social Studies	Dec 2014	Dec 2015
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.031	0.004	0.067	-0.005	0.086	-0.010	0.065	-0.024	0.072
	(0.062)	(0.065)	(0.075)	(0.069)	(0.073)	(0.144)	(0.115)	(0.043)	(0.055)
Constant	0.005	0.003	-0.000	0.001	0.001	0.007	-0.013	-0.018	-0.030
	(0.023)	(0.025)	(0.033)	(0.030)	(0.033)	(0.054)	(0.047)	(0.021)	(0.024)
Observations	9,988	9,988	18,410	18,292	18,571	7,550	9,411	5,414	5,337
R-squared	0.060	0.052	0.042	0.039	0.070	0.111	0.048	0.060	0.101

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent critical level.
All regressions incorporate block fixed effects. Standard errors are clustered at the academic cluster level. The dependent variables in columns 1-7 are individual-level test scores normalized at mean 0 and standard deviation 1 within grade for the control group. The dependent variables for columns 8-9 are school-level scores normalized at mean 0 and standard deviation 1 within-year for the control group. The independent tests were administered to all students present on the day of the assessments. Individual student-level scores on the administrative school tests were transcribed from official registers for all students at the time of school visits. The number of observations is lower for independently-administered tests (in comparison to official tests) because of student absence, whereas official tests are mandatory. The number of observations for Science and Social Studies is lower than other subjects in the official data since these subjects are not tested in all grades.

Sources: †Student-level independent test data from February 2016.
†† Student-level data on administrative school tests (Pratibha Parv) from December 2015.
††† School-level data on administrative school tests (Pratibha Parv)

Table 7: Phase 2 Matched Study: Balance on observable characteristics

	Full Population				Matched Study			
	Untreated Mean (1)	Treated Mean (2)	Diff (2) - (1) (3)	SE (4)	Comp. Mean (5)	Treatment Mean (6)	Diff (6) - (5) (7)	SE (8)
Panel A: School characteristics (DISE)								
Middle school	0.16	0.49	0.33***	(0.00)	0.18	0.35	0.17**	(0.08)
Total Enrolment	53.92	99.12	44.94***	(1.01)	54.86	61.51	6.65	(4.05)
No. of teachers	2.25	3.29	1.02***	(0.03)	2.28	2.67	0.39**	(0.16)
Pupil-teacher ratio	27.16	35.22	8.28***	(0.46)	28.90	25.43	-3.47	(3.35)
Rural	0.96	0.92	-0.03***	(0.00)	0.94	0.90	-0.04	(0.03)
Electricity	0.07	0.19	0.11***	(0.01)	0.07	0.13	0.05	(0.06)
Visits from Block/Cluster officials	8.92	10.21	1.20***	(0.12)	9.72	10.28	0.56	(1.08)
Panel B: School-level test scores (Pratibha Parv)								
— 2012-13	51.85	58.83	6.86***	(0.26)	54.33	53.15	1.18	(2.68)
— 2013-14	57.62	63.11	5.43***	(0.24)	60.10	59.06	1.04	(2.12)
— 2014-15	58.08	64.05	5.86***	(0.25)	59.34	60.67	-1.33	(2.27)
— 2015-16	63.93	67.25	3.24***	(0.20)	64.20	63.76	0.44	(1.91)
— 2016-17	68.36	69.42	0.95***	(0.20)	67.86	67.42	0.42	(2.00)
Observations	18159	5264			100	100		

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.
Differences in means between treatment and comparison group in the matched study and associated standard errors are reported from regressions incorporating academic cluster fixed effects and clustering standard errors at the academic cluster level. The number of observations differs slightly due to occasionally missing information for individual schools. Test scores are on a scale from 0–100.

Source: Administrative data at the school level from school census (DISE) and standardized tests (Pratibha Parv).

Table 8: Phase 2 Matched Study: Treatment effect on student achievement

	Endline score (February 2018)†					
	Math			Hindi		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.064	0.031	0.00073	0.11	0.082	0.021
	(0.074)	(0.069)	(0.074)	(0.072)	(0.067)	(0.067)
School-level PP score (2015-16)††		0.16***			0.13**	
		(0.053)			(0.053)	
Individual PP score (2016-17)†††			0.13***			0.17***
			(0.0079)			(0.0082)
Observations	6,143	6,143	4,173	6,149	6,149	4,164
R-squared	0.215	0.223	0.381	0.184	0.189	0.420

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.
Treatment schools here refer to purposively-selected program schools where the intervention was implemented by the Government of Madhya Pradesh. Comparison schools are matched based on pre-program observable characteristics within the same academic cluster. All regressions include fixed effects for the cluster and for the grade of the student. PP scores refer to administrative Pratibha Parv test scores. The number of observations declines in Cols. 3 and 6 because individual-level lagged scores are only observed for students who were enrolled in the same school in the previous school year and whose scores can be matched. Students in grade 1 and grade 6 are all new entrants in a school in any given year.

Sources: † Student-level independent assessment data from February 2018.
††Administrative data on school-level Pratibha Parv test scores
††† Administrative data on student-level lagged test scores, transcribed manually from school records in 2018

Appendices

A School Management Interventions across the world

Goals of the review

Identifying universe of potentially relevant programs

We identified relevant education projects using the World Bank Education Projects Database, which includes all World-Bank funded projects within the education sector across the world with starting date between 1998-2017. To identify the projects that are relevant to this paper, a search was conducted in the database in June 2019. The search included projects conducted in primary and/or secondary schools and did not further filter on project country or start date.

We identified programs as potentially relevant if they included at least one of the following activities: (i) School-Based Management, (ii) Results-Based Management, (iii) School development Planning, (iv) School-Community Relationships, (v) School Supervision, (vi) Quality Assurance and Accreditation, (vii) Accountability Systems for Education Service Delivery, (viii) School-Based Evaluation of Learning Assessment, (ix) School Principal Performance Assessment, (x) Teacher Performance Assessment, (xi) Teacher Standards, (xii) Operational Standards for Schools, (xiii) Continuous Learning Assessment, (xiv) Management Information: Monitoring and Evaluation, (xv) Management Information: School Report Cards/Public Information on Quality of Provision, (xvi) Management Training and Professional Development. This resulted in a total of 160 potentially relevant projects.

Codified information

We used documentation from each project's webpage in the World Bank database to systematically collate the information on the characteristics of each potentially relevant project. We collected information on seven overview variables: (i) Project Name and website link (ii) Country, (iii) Project Period, (iv) Description of project objectives, (v) Scale, (vi) School Type (Primary VS Secondary) and (vii) Appraised Project cost (Mn $). Project-specific information was primarily acquired through the documentation available on the World Bank projects' websites, in particular the project appraisal documentation as well as the implementation completion reports. Note that these overview characteristics, such as objectives, scale and cost

refer to the full complement of activities within each country program and not just the aspects relevant to our evaluation.[32]

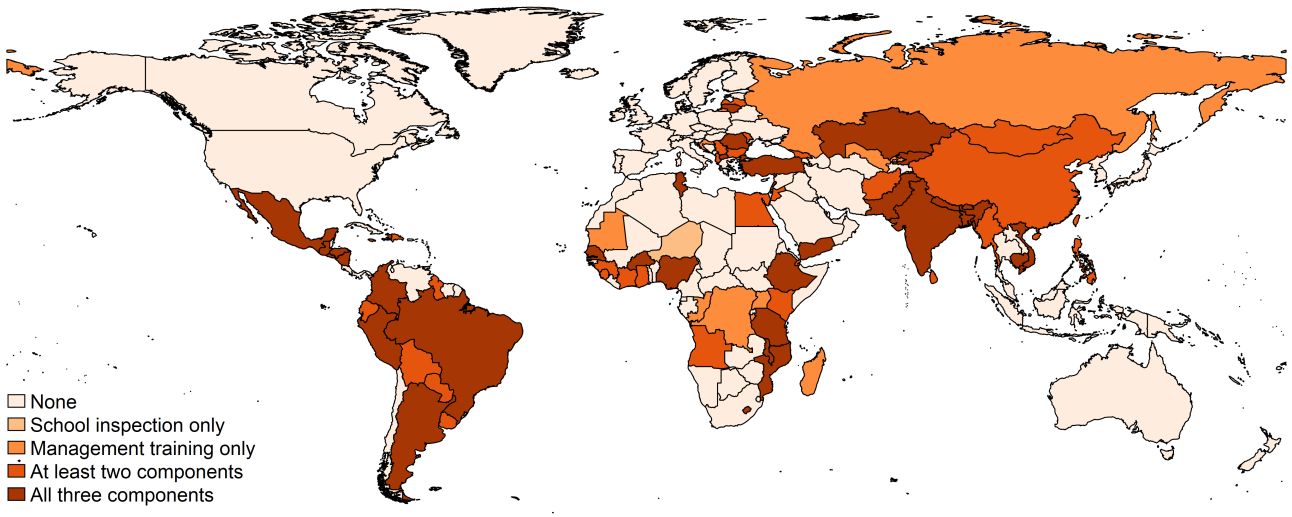
For each project, we then noted whether it involved (any of) the following components and, if so, a brief summary of activities under that component:

- **School Inspections:** Project activities related to conducting inspections of schools such as classroom observations, teacher performance assessments, supervision visits, follow-up on school development plan implementation.
- **School Development Plans:** Project activities related to introducing/enhancing schools' capacity to formulate and implement school development plans.
- **Management Training:** Project activities related to developing the leadership and school management skills of educational staff such as school principals, teachers, school inspectors or school supervisors.
- **School Report Cards:** Project activities related to the use of report cards to track and disseminate information on student- and school-level learning achievements.
- **School Management Committees:** Project activities related to the introduction and/or strengthening of school management committees consisting of both school staff and community members. These could also be referred to in the projects as parent-teacher associations, mother/father committees etc.
- **Monitoring of Learning outcomes:** Project activities related to monitoring the learning levels of students, such as continuous assessments in certain subjects, standardized testing, developing item banks for testing student learning, developing national learning assessment institutions etc.
- **Public Dissemination:** Project activities related to disseminating information on the output of the other project activities publicly, such as publishing school reports online on learning achievements, disseminating school development plans, community awareness campaigns etc.
- **Extra Incentives:** Project activities related to boosting school performance through various incentives such as performance-based teacher bonuses, incentive awards/grants to schools etc.

[32]Further, the scale of projects was not documented in a standardized way across projects and thus could be expressed in various metrics such as number of schools included, number of students to be benefited from the projects, or number of school personnel involved etc.

Following this review, we classified projects as relevant if they included at least one of the first three components, which are most similar to the MPSQA intervention. 160 projects in 84 countries were thus identified as relevant and are presented in Figure A.1. The 32 most relevant projects had all three of these project components; we present a summary of these projects in Table A.1 for illustration.

Figure A.1: School management interventions in developing countries



Notes: This map is based on a review of education-focused interventions supported by the World Bank. The map includes interventions in a total of 84 countries. We use the project documents available online to classify if an intervention included school inspections, management training and/or the use of school development plans. In the figure above, we classify projects as having one, two or all three of these features. All programs which included school development plans included at least one of the other features as well. The list of included programs that contain all three components which includes 32 programs in 27 countries is provided in Table A.1 .

Table A.1: Relevant World Bank projects with components of school management

Project Name	Country	Period	Scale	School Type	Cost (Mn $)	Inspections	Development Plan	Management Training	Report Cards	SMC	Testing	Dissemination	Incentives
Acre Social and Economic inclusion and Sustainable Development Project	Brazil	2008-2020	Target: 154 550 students benefiting from learning quality improvement.	Pri & Sec	150	School inspections to monitor the implementation of school development plans by state and municipal secretariats.	Developing plans to increase school accountability and community participation, and specify various quality improvement subprojects.	Providing training programs for teachers, state and municipal education officials, and supervisors.	No	No	Yes	Yes	No
Basic Education Development Program	Yemen	2004-2012	10 000 schools received >3 inspection visits per year. 47 674 teachers trained. 6666 school inspectors trained.	Pri	121.14	Investments in establishing and strengthening an effective school inspection function.	Building capacity within school communities to formulate and implement school development plans.	Ensuring all basic education inspectors and headmasters receive training supporting and evaluating teachers.	No	Yes	Yes	No	Yes
Ceara Basic Education Quality Improvement Project	Brazil	2000-2008	982 000 Students benefited. 55 municipalities adopted school improvement plans.	Pri	150	Introduction of a pedagogical monitoring and supporting system, including setting regular timetables for school visits, assessing student achievement, dropout rates, etc.	Supporting the design and implementation process of school development plans.	Training professionals responsible for training school staff and community members in designing school development plans and school management.	Yes	Yes	Yes	Yes	No
Cundinamarca Education Quality Improvement Project	Colombia	2003-2006	106 school development plans produced. Project implementation was limited.	Pri & Sec	21.4	Implementation of a comprehensive quality evaluation system, e.g. evaluating teacher and school management performance, learning resources etc.	Providing assistance in designing and implementing school development plans.	Training management teams in preparing their school development plans and entering courses specific to the administrative, financial and management needs of their schools.	No	No	Yes	Yes	No

Table A.1: Relevant World Bank projects with components of school management

Project Name	Country	Period	Scale	School Type	Cost (Mn $)	Inspections	Development Plan	Management Training	Report Cards	SMC	Testing	Dissemination	Incentives
Education Improvement Project	Lithuania	2002-2006	9000 teachers trained. 400 school principals trained.	Pri & Sec	45.41	Changing function of school inspectors to focus more on school improvement through self assessment and independent external evaluation.	Generating data on organizational culture and processes of individual schools in order to monitor their performance and to assist them in preparing school development plans.	Training principals in school management, leadership and community management skills.	No	No	Yes	Yes	No
Education Modernization Project	North Macedonia	2003-2011	26 038 teachers trained. 427 schools receiving improvement grants. 324 244 students benefiting from learning quality improvement.	Pri & Sec	19.5	Training inspectors to monitor and evaluate a school improvement grant program given to schools that develop school development plans. Also includes developing and monitoring standards for school effectiveness.	Providing training of staff to carry out self-evaluation and formulating school development plans.	Training school management to conduct school evaluations and implement school development plans.	No	No	Yes	No	Yes
Education Modernization Project	Kazakhstan	2017-2022	Target: 5400 rural and disadvantaged schools benefiting from project.	Pri & Sec	77	Development of an instrument to observe the pedagogical practice of teachers in the classroom, and improving school inspection practice.	Helping community members to contribute to and monitor school development plans.	Training staff, school leaders and community members to build their skills in school-based management, evaluation and inspections.	Yes	No	Yes	Yes	No
Education Quality and Secondary Education	Guatemala	2007-2015	1489 schools with trained school councils.	Pri & Sec	100	Training district staff on supervision and monitoring of school development plan implementation.	Supporting schools to prepare school development plans with strategies to guarantee on-time entry for new first grade students.	Provision of training and support for school principals and leadership staff.	Yes	Yes	Yes	No	No

Table A.1: Relevant World Bank projects with components of school management

Project Name	Country	Period	Scale	School Type	Cost (Mn $)	Inspections	Development Plan	Management Training	Report Cards	SMC	Testing	Dissemination	Incentives
Education Quality for Equality Project	Lesotho	2016-2021	377 schools with trained school boards, submitted development plans and school report cards.	Pri & Sec	25	Financing supervision of school development plan facilitators by regional inspectors, as well as audits of school development plan expenditures.	Supporting schools in using school developing plans aimed at increasing school performance with regard to quality, retention and equity of access.	Training school principals and boards in how to develop school development plans, training school development facilitators, district education officers and inspectors.	Yes	No	Yes	No	Yes
Elementary Education Project 3	India	2014-2018	4.1 million teachers trained. 261 100 headmasters trained (Nationwide project).	Pri	29833	Financing school performance assessments through development of indicators, and conducting internal and external evaluations of the schools.	Strengthening capacities of SMCs to prepare and implement school development plans.	Establishing a school leadership program to enhance the management competence of school headmasters and educational administrators.	Yes	Yes	Yes	Yes	No
Enhancing Education Development Project	Maldives	2013-2018	212 schools receiving school management training. 3685 teachers trained.	Pri & Sec	11	Introducing internal self-evaluation to help schools acquire information that feeds into their development plans.	Introducing school development plans as part of the school-based management model.	Training school board members and school directors in school-based management.	No	Yes	Yes	No	No
Female Secondary School Assistance Project 2	Bangladesh	2002-2008	6625 head teachers trained. 415 schools introduced to development plans. 128 000 SMC/PTA members trained.	Sec	144.62	Regular visits to schools by regional project office to ensure school standards and accountability.	Supporting quality improvements through the process of a school development planning exercise.	Academic supervision and management training for head teachers to support the improved effectiveness of teachers.	No	Yes	No	No	Yes
General Education Quality Improvement Project 2	Ethiopia	2013-2019	Target: 18 139 200 primary and 2 000 000 secondary school students benefiting from improved learning environment.	Pri & Sec	550	Development of inspection systems at various levels, training of inspectors and conducting quarterly school inspections in selected regions.	Training school leaders in using school development plans and developing a simplified framework for rural, isolated schools.	Leadership training for school directors and supervisors.	Yes	No	Yes	Yes	Yes

Table A.1: Relevant World Bank projects with components of school management

Project Name	Country	Period	Scale	School Type	Cost (Mn $)	Inspections	Development Plan	Management Training	Report Cards	SMC	Testing	Dissemination	Incentives
Lagos Eko Secondary Education Project	Nigeria	2009-2016	4609 school principals and 12606 teachers received management training.	Sec	95	Auditing schools to gather data on school performance that constitute the basis for a school performance award.	Introducing school development plans and school grants that are provided based on the plan proposals.	Training school leaders, district officers, and school-based management committees, particularly in using development plans and effective schooling strategies.	No	Yes	Yes	Yes	Yes
Rajasthan District Primary Education Project 1+2	India	1999-2008	11 956 SMCs established. 766 cluster resource centers established. 620 000 children enrolled.	Pri	188.8	Provision of school monitoring by cluster resource personnel through scheduled/unscheduled visits to observe teacher performance.	Assigning SMCs with responsiblity to plan school development and resource utilization.	Training district and block level management in participatory planning and supervision and providing SMC members with three-day management training.	No	Yes	Yes	No	No
Second Education Quality Assurance Project	Vietnam	2009-2016	2600 education managers and 3150 school principals received management and school inspection training.	Pri	181.4	Developing a teacher evaluation program and teacher methodologies that are strongly associated with improved student learning outcomes, e.g. conducting classroom observations.	Training supervisors and school principals in evaluation and implementation of school development plans.	Modules for in-service training in school management and the use of time for full-day schooling model.	No	Yes	Yes	No	No
School Based Management project 3	Mexico	2014-2018	18 447 school supervisors trained. 76 000 school directors trained.	Pri & Sec	819.95	Support capacity of school directors and supervisors to evaluate teacher performance using classroom observations.	Supporting schools to implement school development plans.	Providing school directors and supervisors with training in how to use the school dashboard tool to improve school management practices.	Yes	Yes	Yes	No	Yes

Table A.1: Relevant World Bank projects with components of school management

Project Name	Country	Period	Scale	School Type	Cost (Mn $)	Inspections	Development Plan	Management Training	Report Cards	SMC	Testing	Dissemination	Incentives
Secondary Education Project	El Salvador	1997-2005	4500 teachers trained. 78 000 students benefiting from improved learning envionrment.	Sec	65	Establishing a school accreditation system that monitors and reports data on school quality, including hiring and training school supervisors.	Enhancing school councils' capacities to prepare school development plans.	Providing management and curriculum training for school principals and school councils.	Yes	Yes	Yes	No	No
Secondary Education Project	Turkey	2005-2012	15 000 educational staff trained in school plan preparation. School development teams formed in 3500 schools.	Sec	104	Establishing a performance management system for continuous improvement of staff and institutions.	Training educational staff in developing and implementing school plans.	Establishing and training school development management teams in all secondary schools.	Yes	No	Yes	Yes	No
Sector Support For Education reform Project	Kyrgyzstan	2013-2019	10 000 teachers and 1500 school directors and inspectors trained.	Pri & Sec	16.5	Training inspectors in observing teaching-learning practices and classroom management, identifying weaknesses and providing supportive feedback.	Supporting school management in using school development plans through training and providing schools with templates for the plans.	Providing training to school directors in school leadership, development planning and community engagement.	Yes	No	Yes	Yes	No
Second Sindh Education Sector Project	Pakistan	2013-2018	8 600 000 students benefiting from improved learning environment (Project includes entire Sindh State).	Pri & Sec	400	Developing annual school census in which supervisors visit schools to collect information on infrastructure, student enrollment, teacher- and headmaster-level information.	Introducing SMCs who formulate and implement school development plans.	Recruitment and training of education managers and school headmasters.	Yes	Yes	Yes	Yes	No

Table A.1: Relevant World Bank projects with components of school management

Project Name	Country	Period	Scale	School Type	Cost (Mn $)	Inspections	Development Plan	Management Training	Report Cards	SMC	Testing	Dissemination	Incentives
Zanzibar Improving Student Prospects Project	Tanzania	2016-2021	Target: 1500 teachers receiving training. 170 000 direct benificiaries of the project.	Pri & Sec	35	Boosting classroom inspections so each teacher is inspected at least twice per year and inspection data is available in a more usable and timely fashion.	Assisting SMCs in formulating school development plans.	Capacity building of inspectors and head teachers through training.	Yes	Yes	Yes	Yes	Yes
Education Development Project II	Lebanon	2010-2018	300 project schools receiving training, inspection etc. 260 teachers trained.	Pri & Sec	42.6	Piloting a system of school self-evaluation or school-based review as the first step in developing school development plans.	Designing, piloting and evaluating a program for school-based planning in which school principals and school council members are taught to use school development plans.	Training and assisting school principals and other educational staff in designing development plans and conducting performance self-assessment.	No	Yes	No	No	No
Education Quality Improvement Program - (Phase 2)	Tunisia	2004-2010	800 principals and 42542 teachers trained.	Pri & Sec	290.92	Setting up a regional school supervision system to monitor and evaluate the implementation of school development plans.	Setting up school development plans and training school staff in developing such plans.	Training in school development planning to school staff and promoting decentralized school management.	No	No	Yes	No	No
Education Sector Support Project 1	Malawi	2005-2010	5200 schools received direct support.	Pri	32.2	Conducting interviews with students, teachers, parents and school administrators to identify factors in school and home environment affected by policy interventions.	Introducing the use of school development plans.	Training secondary school managers in their roles, responsibilities and school planning.	No	No	Yes	No	No
Education system Improvement Project	Kosovo	2015-2019	Target: 45 000 students in targeted schools and 25 000 teachers benefiting from project.	Pri & Sec	11	Developing capacities of teacher evaluators/inspectors to conduct teacher performance assessment.	Supporting primary schools to prepare and implement multi-year school development plans.	Training school staff in management, planning and self-evaluation.	No	No	Yes	No	No

Table A.1: Relevant World Bank projects with components of school management

Project Name	Country	Period	Scale	School Type	Cost (Mn $)	Inspections	Development Plan	Management Training	Report Cards	SMC	Testing	Dissemination	Incentives
Primary Education Development Project 2	Bangladesh	2004-2011	75 000 teachers and 8 000 head teachers received training. 18.5 million children estimated to benefit directly from project.	Pri	1815	Implementing a new school inspection system, steering away from mere policing towards more supportive school inspection.	Giving support to head teachers, SMCs and communities in implementing school development plans.	Providing SMC members with training in community engagement, and training head teachers in management, teacher support and supervision.	No	Yes	No	No	No
Secondary Education Improvement Project	Cambodia	2017-2022	Target: 130 target schools benefiting from upgraded learning environment and training provided to 310 school directors and 2200 teachers.	Sec	40.9	Conducting school-level assessments using the "Lower Secondary School Effectiveness Standards" tool, as well as implementing a self-assessment tool for teachers.	Provision of grants to schools that submit school development plans.	Providing school directors with leadership and management training.	No	Yes	No	No	No
Secondary Education Project	Romania	2015-2022	Target: 1160 schools receiving school-based interventions.	Sec	243.1	On-site monitoring of implementation of school improvement plans.	Technical assistance for the preparation and evaluation of school development plans.	Training teachers and public school directors on implementing a revised curriculum and an inclusive education.	No	No	Yes	No	No
The Third Secondary Education Project	Argentina	1998-2002	198 schools targeted in the project.	Sec	170	Training school supervisors in how to guide, monitor and supervise the use of school development plans, and establishing a permanent evaluation and monitoring system.	Technical assistance in implementing school development plans.	Training and technical assistance for school principals and vice-principals in school management, institutional administration and pedagogical supervision.	No	No	No	No	No
Nigeria State Education Sector Project	Nigeria	2007-2011	1523 schools targeted to implement school development plans.	Pri & Sec	75	Establishing a quality assurance inspection system in which training workshops are provided to inspectors to improve school monitoring, resulting in annual basic education reports.	Supporting the implementation of school development plans for which schools may receive grants.	Developing school leadership capacity through training in procurement, financial management and accountability.	Yes	No	No	No	No

Table A.1: Relevant World Bank projects with components of school management

Project Name	Country	Period	Scale	School Type	Cost (Mn $)	Inspections	Development Plan	Management Training	Report Cards	SMC	Testing	Dissemination	Incentives
Reaching all children with education in Lebanon support project	Lebanon	2016-2023	Target: 14 500 teachers per year receiving training. 780 schools to implement development plans.	Pri & Sec	234	Classroom observations to monitor teacher performance. Teacher observation records are digitalized so that teacher progress can be tracked over time.	Financing in the form of grants to schools for them to prepare and implement school development plans.	Training teachers and school leaders in effective instructional techniques, student assessment and school management, with focus on fostering diversity and gender equity.	No	No	Yes	No	No

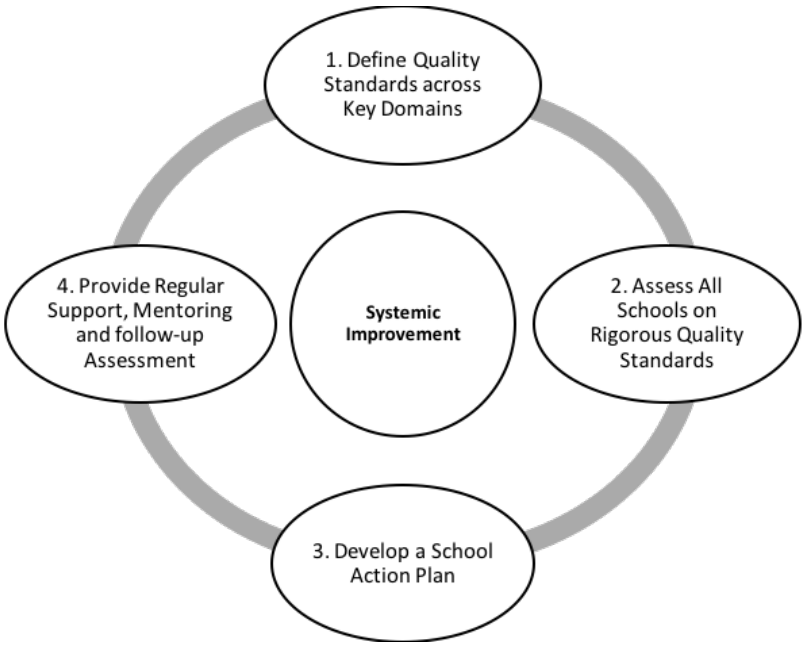
B Details on program design and implementation

This section contains details on the intervention – Madhya Pradesh *Shaala Gunvatta* (School Quality Assessment Program). The program was designed in partnership with the state government of Madhya Pradesh, India, the British DFID, and Absolute Return for Kids (ARK), based on procedures followed by the UK Office for Standards in Education, Children's Services and Skills (OfStEd).

B.1 Theory of Change

MPSQA was conceived with the aim of devolving power to the school for overall improvement in school quality. Schools were supported to identify the challenges hindering quality. Once areas for improvement were identified by the school itself, they were provided with regular support and mentoring to improve upon their processes. This would lead to an improvement in the quality of the school (including school level processes and outcomes). The process for overall school improvement is described in the chart below.

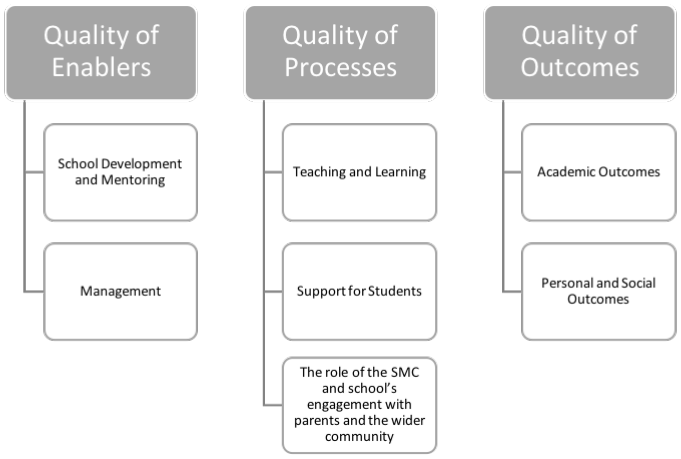
Figure B.1: Process of school improvement in MPSQA



B.2 Rubric for school assessments

Schools were assessed on 7 standards that were spread over 9 sets of questionnaires (called Tools). Chart 2 shows the division of the standards and elements they aimed to measure. Based on the assessment, for each standard the school was graded as one of the following levels (a) Exceeds expectation (b) Meets Expectation (c) Close to Expectation (d) Below Expectation. The overall rating for the school was generated as a weighted average of these standard-specific ratings.

Figure B.2: Components of school assessments



We provide the details on each standard below, with a description of indicators for all 4 grades/ levels. The assessors would fill up 9 different questionnaires/ tools which included questions around school level processes, outcomes and interviews with different stakeholders. Questions from across these tools mapped onto these 7 standards. The score of each standard would then be generated using a pre-decided weightage. A weightage would then be provided to each standard to compute the final grade of the school.

Std 1: School Development and Mentoring

Description: The Head Teacher effectively promotes improvement in the quality of teaching and student outcomes.

Description of standards			
Exceeds	Meets	Close	Below
a. The school has produced a clear and specific School Development Plan which is strongly focused on improvement of student outcomes and the quality of teaching.	a. The school has produced a clear and specific School Development Plan which includes the improvement of student outcomes and improvement in teaching.	a. The school has produced a clear and specific School Development Plan, although the priorities identified are limited to improvement of infrastructure.	a. The school does not have a School Development Plan.
b. This is supported by systematic classroom observation by the head teacher, followed by feedback and guidance to the teachers to improve their practice.	b. This is supported by regular visits to the classroom by the head teacher, followed by feedback and guidance to the teachers to improve their practice.	b. The head teacher makes visits to the classrooms occasionally, and these are sometimes followed by feedback and guidance to the teachers to improve their practice.	b. The head teacher seldom visits the classroom and provides with little or no feedback and guidance to the teachers to improve their practice.

Std 2: Management

The school is managed well, with funding, facilities and human resources used efficiently, satisfactory administrative routines, and a good team spirit among the staff.

Description of standards			
Exceeds	Meets	Close	Below
a. The school is managed very well, with funding, facilities and human resources used very efficiently,	a. The school is managed well, with funding, facilities and human resources used efficiently,	a. In most respects the school is managed adequately, with funding, facilities and human resources generally used efficiently, but there are a few examples of inefficiencies.	a. The school is managed poorly with inefficiencies in the use of funding, facilities and human resources,
b. Good administrative routines (for example: record keeping and time keeping),	b. Satisfactory administrative routines (for example: record keeping and time keeping)	b. Most but not all administrative routines are satisfactory (for example: record keeping and time keeping)	b. Weaknesses in administrative routines (for example: record keeping and time keeping)
c. And a very good team spirit among the staff, with focus on school development and optimum use of facilities and resources.	c. And a good team spirit among the staff.	c. And the team spirit among the staff is satisfactory.	c. And limited team spirit among the staff.

Std 3: Teaching and Learning

The teaching demonstrates good pedagogical skills, good subject knowledge and good interactions between teachers and students.

Description of standards			
Exceeds	Meets	Close	Below
a. Almost all of the teachers demonstrate good pedagogical skills including the use of formative assessment to meet the needs of different students.	a. Most of teachers demonstrate good pedagogical skills including the use of formative assessment to meet the needs of different students.	a. The majority of the teachers demonstrate good pedagogical skills including the use of formative assessment to meet the needs of different students.	a. Only a minority of the teachers demonstrate good pedagogical skills including the use of formative assessment to meet the needs of different students.
b. Almost all of the teachers' subject knowledge is accurate and up to date.	b. Most of the teachers' subject knowledge is accurate and up to date.	b. The majority of the teachers' subject knowledge is accurate and up to date.	b. The subject knowledge of only a minority of teachers is accurate and up to date.
c. Almost all the teachers question the students effectively, communicate with them well and establish positive relationships in the classroom.	c. Most of the teachers question the students effectively, communicate with them well and establish positive relationships in the classroom.	c. The majority of the teachers question the students effectively, communicate with them well and establish positive relationships in the classroom.	c. Only a minority of the teachers question the students effectively, communicate with them well and establish positive relationships in the classroom.

Std 4: Support for students

The school is inclusive, provides equal opportunities and academic support for all students, and promotes the students' health, safety and personal development.

Description of standards			
Exceeds	Meets	Close	Below
a. The school is exceptionally effective in providing equal opportunities and ensuring no discrimination occurs.	a. The school is inclusive, providing equal opportunities and discouraging discrimination.	a. The school is inclusive in most respects in providing equal opportunities and discouraging discrimination.	a. The school falls well short of being fully inclusive. There is some evidence of discrimination.
b. The progress of all students, including those with special educational needs, is effectively tracked and very well supported.	b. The progress of all students, including those with special educational needs, is tracked and well supported.	b. The progress of students, including those with special educational needs, is tracked and some support is given.	b. The progress of students, including those with special educational needs, is not tracked effectively and little support is given.
c. The school promotes the students' health, safety and personal development actively and highly effectively.	c. The school promotes the students' health, safety and personal development.	c. The school makes some effort to promote the students health, safety and personal development.	c. The school does little to promote the students' health, safety and personal development.

Std 5: The role of the SMC and school's engagement with parents and the wider community

The school has an effective SMC, engages well with the wider community, communicates well with parents and secures their support.

Description of standards			
Exceeds	Meets	Close	Below
a. The SMC performs its role effectively.	a. The school has a functioning SMC (head and members).	a. The school has an SMC but only the head is active.	a. The SMC is inactive, and there is little contact with the wider community.
b. The school engages well with the wider community, communicates well with the parents, and secures their active support for their children's education.	b. The school initiates wider community involvement, provides information to the parents, and encourages them to take an interest in their children's education.	b. There is some contact with the wider community. The school provides a little information to the parents and tries to make contact with them.	b. The school provides no information to the parents and does not encourage contact.

Std 6: Academic Outcomes

Description of standards			
Exceeds	Meets	Close	Below
a. 90% or more students are in grades C or above in Hindi, and Maths for class 5 or 8 Pratibha Parv results. Or 75% or more students are in division 2 or above in Hindi, and Maths for grade 10 Or 55% or more students are in division 2 or above in all the streams for grade 12	a. 75% or more students are in grades C or above in Hindi, and Maths for class 5 or 8 Pratibha Parv results. Or 65% or more students are in division 2 or above in Hindi, and Maths for grade 10 Or 65% or more students are in division 2 or above in all the streams for grade 12	a. 50% or more students are in grades C or above in Hindi, and Maths class 5 or 8 Pratibha Parv results. Or 55% or more students are in division 2 or above in Hindi, and Maths for grade 10 Or 55% or more students are in division 2 or above in all the streams for grade 12	a. Less than 50% students are in grades C or above in Hindi and Maths class 5 or 8 Pratibha Parv results. Or Less than 45% or more students are in grades C or above in Hindi and Maths for grade 10 Or Less than 45% students are in division 2 or above all the streams for grade 12
b. In most of the classes the students make very good progress and the students generally demonstrate understanding consistent with or better than their test results.	b. In most of the classes the students make good progress and the students generally demonstrate understanding consistent with their test results.	b. In most of the classes the students make acceptable progress and the students generally demonstrate understanding slightly below consistent with their test results.	b. In most of the classes the students make weak progress and the students generally demonstrate understanding well below their test results.

Std 7: Personal and social outcomes

Description of standards			
Exceeds	Meets	Close	Below
a. Almost all of the students attend well and are punctual to school and for lessons.	a. Most of the students attend well and are punctual to school and for lessons.	a. The majority of the students attend well and are punctual to school and for lessons.	a. Only a minority of the students attend well and are punctual to school and for lessons.
b. Students have very positive attitudes and values, and respectful relationships with adults.	b. Students have positive attitudes and values, and good relationships with adults.	b. Students have positive attitudes and values, and good relationships with adults.	b. Poor behaviour and lack of respect by a minority of the students disrupts the learning environment.
c. They show respect and understanding of local history and culture.	c. They show interest in and some understanding of local history and culture.	c. Students show a little interest in but a limited understanding of local history and culture.	c. Students show no interest in and very little understanding of local history and culture.

B.3 Implementation of project components

1. **Training of school assessors:** School assessors were trained using a cascade model of training, which is common in large interventions in this setting. A State Resource Group (SRG) was created, who were designated as master trainers. This resource group of about 50 individuals represented each of the 10 regions (each region is a group of districts) in MP. This SRG, in turn, trained the District Resource Group (DRG) at the regional level comprising of 35- 40 individuals. The DRG then trained the assessors. These trainings were spread over 4 days and included one day of field visit, where the trainees conducted a mock assessment process. During this field trial the assessors collected the necessary data and then returned to generate the school-specific reports. There was no feedback that was given to schools and no action plan was developed as part of this training.

2. **Assessments of Schools:** Each school was assessed by a team of two assessors, one internal and one external. The internal assessor was the schools' corresponding Cluster Resource Coordinator or Block Resource Coordinator and the external assessor was a retired head teacher or a fresh Bachelor's of Education Graduate. This team of two spent 1-2 days to complete the assessments: these assessments used 9 tools (structured data collection

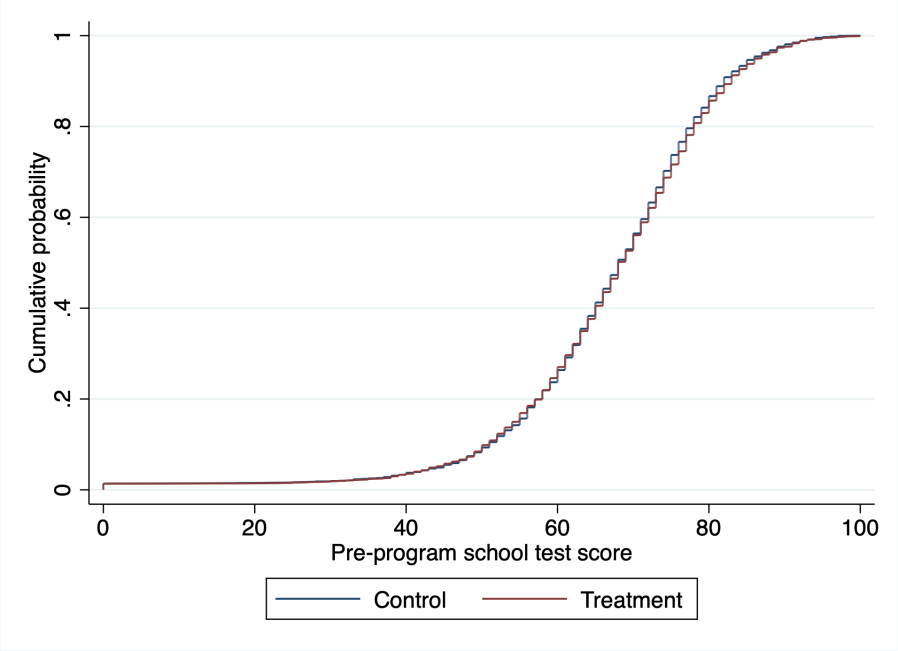
questionnaires or protocols) to arrive at judgements on the 7 domains. The assessors could enter the data on their android phones/ tablets in the schools directly or came back to the base station and entered it on the online application. At the end of the data entry a ∼4 page report was generated which listed out the key recommendations for the schools to work on to improve their quality standards. As the next step, the assessors went back to the schools and spent half-a-day with the Head Teacher, School Management Committee president and members, and teachers to discuss the report and recommendations in detail. Based on these recommendations the group then collaboratively worked to create the School Improvement Plan. The entire exercise was neither prescriptive nor punitive.

3. **School Development Plan and Follow up Visits:** As described above, the School Improvement Plan was developed collaboratively. This action plan listed out all the specific actions that the school was to undertake over the next academic year. During this time, the schools were supposed to be provided with continued mentoring, monitoring and support from the internal assessors by quarterly follow up visits. These visits were intended to be part of the regular schools visits that these officials undertake but with the specific task of discussing the progress on the school action plan.

C Additional figures and tables

C.1 Phase I evaluation: Elementary schools

Figure C.1: Distribution of pre-program test scores



Notes: This figure shows the distribution of school-level test scores administrative test data (Pratibha Parv) in March 2013 for the full population of randomly-assigned treatment and control schools in Phase I of the evaluation. These were the most recent public data available at the the time of randomization in 2014. We cannot reject the equality of distributions across the two groups (p-value of 0.67 in a Kolmogorov-Smirnov test).

Table C.1: Quantile treatment effects on independent test scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mathematics					**Hindi**				
VARIABLES	q10	q30	q50	q70	q90	q10	q30	q50	q70	q90
Treatment	-0.04	0.03	0.00	0.07*	0.03	-0.01	-0.05	-0.00	0.07*	0.03
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.05)
Pre-treatment PP school score	0.03***	0.07***	0.09***	0.10***	0.02	0.05***	0.07***	0.09***	0.07***	0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	-1.25***	-0.87***	-0.41***	0.21*	1.08***	-1.36***	-0.91***	-0.33***	0.35***	1.29***
	(0.10)	(0.07)	(0.09)	(0.11)	(0.08)	(0.04)	(0.07)	(0.07)	(0.09)	(0.08)
Observations	9,901	9,901	9,901	9,901	9,901	9,901	9,901	9,901	9,901	9,901

Source: Student-level data from independent learning assessments conducted in February 2016.
Note: Treatment is an indicator variable denoting schools which were randomly-assigned for program implementation. Dummy variables for each block are included in all regressions. Bootstrapped standard errors in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

C.2 Phase I evaluation in secondary schools

Table C.2: Balance on observable characteristics

	All Study Schools				Subsample with primary data			
	Control Mean	Treatment Mean	Diff (T-C)	Diff SE	Control Mean	Treatment Mean	Diff (T-C)	Diff SE
Total enrollment	256.00	250.30	-12.32	(22.20)	265.28	250.30	-25.53	(30.11)
No. of teachers	6.23	6.46	0.21	(0.43)	6.31	6.46	-0.06	(0.58)
Pupil-teacher ratio	44.14	43.87	-0.99	(3.42)	44.38	43.87	0.81	(3.84)
Proportion of Qualified teachers	0.89	0.89	-0.01	(0.01)	0.89	0.89	-0.01	(0.02)
Rural	0.78	0.80	0.00	(0.05)	0.80	0.80	0.02	(0.05)
Electricity	0.52	0.48	-0.05	(0.05)	0.52	0.48	-0.06	(0.06)
Class X Board Score								
Total Score					266.18	265.13	-7.94	(6.59)
Hindi Score					46.07	45.94	-1.39	(1.26)
Math Score					33.75	35.54	-0.37	(1.31)
Observations	273	116			116	116		
F-test of joint significance (F-stat)			0.431				0.938	
F-test, number of observations			346				191	

Source: 2013-14 school-level administrative data from DISE and Board scores.
Note: Differences in means between treatment and control group and associated standard errors are reported from regressions incorporating block fixed effects and clustering standard errors at the academic cluster level (level of randomization). The F-test refers to a test of joint significance for all reported variables and reports the F-statistics. The number of observations for the F-tests are lower than the total number of observations due to missing data for pupil-teacher ratio, proportion of qualified teachers, electricity indicator and Board scores. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table C.3: Ranking of treatment schools by school inspectors

	Percentage of schools				N
	Below standards	Close to standards	Meets standards	Above standards	
Std.1: School Development and Mentoring	8.7	22.3	60.2	8.7	103
Std.2: Management	3.9	36.9	46.6	12.6	103
Std.3: Teaching and learning	53.4	39.8	5.8	1	103
Std.4: Support for students	1	21.4	52.4	25.2	103
Std.5: SMC and engagement with parents	2.9	32	34	31.1	103
Std.6: Academic Outcomes	17.5	66	13.6	2.9	103
Std.7: Personal and Social Outcomes	4.9	28.2	58.3	8.7	103
Overall	4.9	63.1	29.1	2.9	103

Source: Administrative ARK data on school assessment grades.
Note: These ratings are taken from the administrative data on program implementation and include all treatment elementary schools for which data was available. The data indicate that school inspections, and the corresponding input of assessments into the online portal, was done for 89% of the treatment schools at the secondary level.

Table C.4: Informativeness of school assessments

	(1)	(2)
	Board Score 2014-15††	
Close to standard†	0.109	0.115
	(0.106)	(0.189)
Meets standard†	0.150	0.159
	(0.128)	(0.241)
Constant	0.062	0.075
	(0.095)	(0.180)
Pre-treatment Board Score††	0.684***	0.573***
	(0.063)	(0.072)
Observations	87	84
R-squared	0.626	0.793
Block FE	No	Yes

Source: † Administrative ARK Data on school assessment grades.
†† Administrative student-level data on board exam scores.
Note: This table relates future student achievement aggregated at the school level to the grades that schools receive in the MPSQA school assessments and lagged test scores. Standard errors are clustered at school level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table C.5: Program effects on frequency and intensity of monitoring and School management committee functioning

	Control Mean	Treatment Mean	Diff (T-C)	Diff SE
Visits by Block/Cluster officials (Full sample)†	3.05	2.42	-0.61	(0.42)
Visits by Block/Cluster officials (Sub sample)†	2.50	2.42	-0.18	(0.59)
Time of last visit††				
Within last month	0.27	0.27	-0.01	(0.06)
Within last 2 months	0.49	0.42	-0.06	(0.07)
Within 6 months	0.70	0.68	-0.06	(0.06)
Within last year	0.84	0.90	0.08*	(0.04)
Time spent by inspector††				
Less than 30 minutes	0.14	0.19	0.05	(0.04)
More than 30 minutes	0.86	0.81	-0.05	(0.04)
More than 1 hour	0.51	0.48	-0.04	(0.07)
More than 3 hours	0.04	0.05	0.01	(0.02)
Observations	116	116		

Source: † 2015-16 School-level administrative Dataset from DISE.
†† Official 2015 school inspection register records.
Note: Differences in means between treatment and control group and associated standard errors are reported from regressions incorporating block fixed effects and clustering standard errors at the academic cluster level (level of randomization). The sample for the visits by block/cluster officials in the full sample consists of 271 control schools and 116 treatment schools. Data on time of inspector visits was collected for the subsample of 232 secondary schools. Data on time spent by the inspector was reported by the headmaster/school in-charge. If no details were available, a school was visited up to 3 times to complete the survey. The content of the inspector's comments were also transcribed and do not reveal any important differences between the treatment and control schools. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table C.6: Effects on teacher attendance, pedagogy and student engagement - School Level

	Control		Treatment		Diff	Diff SE
	N	Mean	N	Mean	(T-C)	
Teacher Attendance*	116	0.78	115	0.76	-0.03	(0.02)
Pedagogical inputs†						
Textbooks used during class	113	0.92	113	0.95	0.03	(0.02)
Workbooks used during class	114	0.09	113	0.05	-0.04	(0.02)
Teacher praised students	112	0.42	113	0.42	-0.00	(0.05)
Time spent on activities†						
% Spent on Lecture	114	47.70	113	47.23	-2.36	(2.82)
% Silent Work	114	2.36	113	3.15	0.74	(0.94)
% Group Call	114	14.04	113	16.98	4.12**	(1.87)
% Small Group Work	114	0.11	113	0.00	-0.13	(0.09)
% Big Group Work	114	2.03	113	2.05	0.06	(1.04)
% Class Discipline	114	2.19	113	1.22	-0.69	(0.58)
% Out of class	114	1.43	113	0.61	-0.70	(0.56)
Child has HW notebooks which are checked††	114	0.35	113	0.33	-0.02	(0.04)
Student Attendance†††	116	0.51	114	0.49	-0.03	(0.03)
Student Engagement						

Source: *2015-16 teacher-level primary data from direct interviewer observations collected in one round. † 2015-16 teacher-level primary data from direct interviewer observation of two classrooms per school. †† 2015-16 Primary data of homework review collected at student-level. ††† 2015-16 School-level primary data collected in three rounds.

Note: All data has been aggregated at the school-level. Differences in means between treatment and control group and associated standard errors are reported from regressions incorporating block fixed effects and clustering standard errors at the academic cluster level (level of randomization). Round indicators are included as covariates in the estimations for teacher attendance. Where a school was found closed at the time of the inspection (always during business hours), all teachers are marked absent. Data on pedagogical inputs and time spent on activities are collected for two classrooms per school in grade 9 divided equally between mathematics and hindi. For a subset of 7 schools - only one classroom could be observed. The variables indicating time spent on activities are based on questionnaire items which asks if the time spent in classroom for each of these activities was below 25%, above 25% and below 50%, above 50% and below 75%, above 75% or not conducted at all, and take on five values: 0%, 12.5%, 37.5%, 62.5% and 87.5%. Student attendance is reported as average percentage of students across three rounds of data collection at the school level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table C.7: Treatment effect on student achievement

	Class X Board Exams	
	(1)	(2)
	2015	2016
Treatment	0.081	0.058
	(0.056)	(0.060)
Constant	0.027	0.020
	(0.041)	(0.043)
Observations	15,346	17,242
R-squared	0.105	0.094

Source: Administrative student-level data on test scores in Class 10 Board examinations.
Note: All regressions incorporate block fixed effects. Standard errors are clustered at the academic cluster level. The dependent variables are individual-level test scores normalized at mean 0 and standard deviation 1 within year for the control group. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

D Details of Data Collection

D.1 Intervention implementation and data collection timelines

S.No.	Activity	Timeline
1.	Piloting of Shaala Gunwatta tools in 100 schools in Madhya Pradesh, India to finalise tools and evaluation model	2012-2013
2.	Selection of treatment clusters and control clusters for implementation of Shaala Gunwatta across 5 districts in Madhya Pradesh	July-August, 2014
3.	Assessments of treatment schools	September-November 2014
4.	Pratibha Parv standardized student tests 2014-15	December 2014
5.	Round 1 of collection of process variables from sub-sample of elementary and secondary schools (treatment and control)	September – November, 2015
6.	Pratibha Parv standardized student tests 2015-16	December 2015
7.	Round 2 of collection of process variables from sub-sample of elementary schools (treatment and control)	January, 2016
8.	Round 3 of collection of process variables from sub-sample of elementary schools (treatment and control)	February, 2016
9.	Independent student tests in sub-sample	March, 2016
10.	Phase 2: Training of District/School staff	June, 2016
11.	Phase 2: Completion of assessments and SIPs	November 2016
12.	Phase 2: Independently-administered student tests in elementary schools in matched sample	March 2018

Note: The Phase 2 data collection refers only to the independent data that is available for both treated and comparison schools in the matched study.

D.2 Student Learning Outcomes

Our primary outcome, student learning outcomes, was measured through independently administered tests in Hindi (language) and maths supervised by the research staff in March 2016. Students in all elementary schools in the sub-sample were tested. No testing was conducted in secondary schools.

D.2.1 Test Design & Content

The test items were a combination of questions from the state administered learning survey (called Pratibha Parv) as well as questions from test booklets of research studies like Young Lives and APREStt. Each test was a combination of items below and at grade level. The items below grade level include questions that tested for basic math and literacy skills allowing us to understand levels of learning. The grade level items helped us assess proficiency of students at the level that the state government expected them to be. All tests were piloted in non-study elementary schools to remove questions with little or no variation as well as to fine tune test administration protocols.

D.2.2 Test Administration

All tests took place in the school premises within school hours. All students present on the day of the assessment were tested. No revisits were made to assess students who were absent on the day of the original assessment. Three different methods of test administration were used – individual oral, group oral and written assessments. This allows us to avoid floor effects in assessments. The table below provided the details of the mode of administration in each grade:

GRADE	HINDI	MATHS
1	Group Oral & Individual Oral	Group Oral & Individual Oral
2	Group Oral & Individual Oral	Group Oral & Individual Oral
3	Group Oral & Individual Oral & Written assessment	Group Oral & Individual Oral & Written assessment
4	Group Oral & Written assessment	Written assessment
5	Written assessment	Written assessment
6	Written assessment	Written assessment
7	Written assessment	Written assessment
8	Written assessment	Written assessment

1. **Individual Oral Assessment**: one on one assessment of each student by survey staff outside the classroom. This test mostly included reading of letters, words and sentences. It also included questions on counting objects to test basic numeracy skills. Individual oral was restricted to students in class 1, 2 and 3. In each class a maximum of 15 students were tested using this method. If more than 15 students were present, a random number generator was used to select the 15 students to be tested.

2. **Group Oral Assessment**: all students present in the class were tested together. The survey staff read out the question to the class and students marked answers in their individual test booklets. This test contained questions on vocabulary, mathematical operations and word problems. Group oral was restricted to students in primary grades (1,2,3,4) It was ensured that during administration at least two survey staff were present in the classroom to ensure that children followed the questions being read to them.

3. **Written assessments**: all students present in the class were provided the test booklet and were provided defined time to complete the assessment. Survey staff were instructed to let children complete the assessment even if the maximum time allowed had elapsed. Students from classes 3-8 were provided written assessments. No written assessments were conducted with students in class 1&2.

D.3 Details of data collection and protocols

In addition to outcome data from student achievement, we collected information on school functioning over multiple visits (refer timeline provided in Table C.1) to the 302 elementary across 5 districts.:

D.3.1 Principal/ Headmaster details and characteristics

1. Included details on professional experience, methods of monitoring and evaluating staff. All information was collected based on in-person interview of the school principal/headmaster (HM)/in-charge during school hours.

2. In some schools where the principal or headmaster hadn't been appointed, the teacher designated as in-charge was interviewed. If the principal/ headmaster/ designated in-charge was not present on the day of the visit, upto 3 visits were conducted.

D.3.2 Teacher attendance

1. Staff attendance was collected thrice during the academic year 2015-16. Each visit was unannounced. In the first round of collection, several other modules were also administered (example – classroom observation, homework review, details on school management committees etc.). It was ensured that staff attendance was collected in the first/ unannounced visit. During the first round of data collection, permanent teachers in the state went on strike. This strike led to many schools being closed during data collection. In these schools, attendance data was collected in the second/ third visit. All these visits were unannounced.

2. In the second and third round of attendance collection, only one visit to each school was made. This visit was made during school hours (data on this had been collected previously). If the school was found closed during school hours, all staff were marked as absent.

D.3.3 Details on inspections and visits by government officials

1. Details of last and second to last visit by a government official were noted, including details of comments made in the official inspection register.

2. This information was collected using the official records maintained by the school (primarily inspection register). In cases where documentation was not available but school in-charge/ senior teachers could provide details, survey was completed. If no details were available, a school was visited up to 3 times to complete the survey.

D.3.4 Details of School Management Committees/ PTAs

1. Details of SMC/ PTAs including details of last two meetings and their perceived usefulness were collected. The school principal/ headmaster or any senior teacher were typical respondents. Under the Right to Education, 2009, each elementary school is mandated to have a School Management Committee. In secondary schools, details of Parent Teacher Associations (SMC equivalent were collected).

2. All administrative details of SMC (present or not, number of members, details of meetings) were noted down form the official documentation maintained in the school. In cases where documentation was not available on the day of the visit but school in-charge/ senior teachers could provide details, survey was completed. If no details were available a school was visited up to 3 times to complete the survey.

D.3.5 Classroom observations

1. In each school, two classrooms were observed. Classes observed were based on type of school:

Type of school	Grades observed	Subject observed
Primary school (schools from class 1 to 5)	2 and 4	Hindi (language) and Maths
Middle school (schools from class 6 to 8)	6 and 8	Hindi (language) and Maths
Secondary schools (schools from 9 to 12)	9	Hindi (language) and Maths

2. For primary and middle schools, the selection of the combination of class and subject was done randomly by the research staff (who also ensured that equal number of class*subject combinations existed). In each school only teachers teaching the subject regularly to the class in the academic year 2015-16 were observed. No substitute teachers were observed. Up to a maximum of three visits were made to complete the classroom observation. In some schools, the same teacher was observed twice as they taught both subjects to be observed. Each classroom was observed for a maximum of 1 hour. If the teacher had already started teaching (10 minutes of class starting), survey staff revisited the school but did not enter midway during the class.

3. Staff was trained through in-class training (by using classroom videos) as well as field practice by visiting government schools.

4. Percentage of time spent on task: During observation, survey staff also noted down time spent on different activities by the teacher. The time spent was not exact but recorded in ranges:

   (a) Activity not conducted

   (b) Less than 25% time spent on the activity

   (c) More than 25% but less than 50% time spent on the activity

   (d) More than 50% but less than 75% time spent on the activity

   (e) More than 75% time spent on the activity

D.3.6 Homework Review

1. In each school, two classrooms were observed. In each observed classroom, homework copies of five randomly selected students were reviewed. A random number list was used for this

selection. In the review, survey staff looked for whether the copy had been checked, and if checked the nature of feedback provided.

2. Information on whether or not these five children had homework notebooks was noted down. No substitution was made if a child did not have a homework notebook. If less than five students were present, notebooks of all students were reviewed. In some schools, no separate homework notebook was maintained. Survey staff reviewed any notebook that the teacher regularly corrects (which may or may not necessarily be referred to as a homework notebook).

D.3.7 Teacher Characteristics

1. All information was collected based on in-person interview of a school teacher during school hours. In each school a maximum of two teachers were interviewed. These teachers were observed teaching before the interview. In a primary school, teachers teaching class 2 and 4 were interviewed, in middle schools, teachers teaching class 6 and 8 were interviewed and in secondary schools teachers teaching class 9 were interviewed. In each school, one hindi/ language and one maths teacher was interviewed. The selection of the combination of class and subject was done randomly.

2. In each school only teachers teaching the subject regularly to the class in the academic year 2015-16 were interviewed. No substitute teachers were interviewed. Up to a maximum of three visits were made to complete the teacher interview. If the teacher observed was the school in-charge, they were not interviewed a second time as many of the questions between the principal and the teacher interview are the same. In some schools, the same teacher was observed twice as they taught both subjects observed. These teachers were only interviewed once.

D.3.8 School Infrastructure

1. Infrastructure details were collected by inspecting the school building and facilities. A school staff was asked for clarifications when needed.

D.3.9 Student attendance

1. Student enrollment and attendance were recorded from the student attendance register present in the school on the day of survey. In some schools, summary attendance is also maintained. If the attendance register was not present, summary attendance records were used. Attendance was collected three times over the academic year through unannounced

visits. It was ensured that student attendance was collected in the first visit. During the first round of data collection, permanent teachers in the state went on strike. This strike led to many schools being closed during data collection. In these schools, attendance data was collected in the second/ third visit. All these visits were unannounced.

2. In the second and third round of data collection, only one visit to each school was made. This visit was made during school hours (data on this had been collected previously). If the school was closed, no student attendance data could be collected.

D.3.10 Initial school observation

1. Details of classes being held and general state of affairs as observed by survey staff on entering the school premises.

2. No clarification was sought from the school staff while noting down the state of affairs (including classrooms in which teachers weren't present but children were). This section was skipped if morning assembly or lunch hour was going on when the survey staff obtained consent.

D.3.11 Student surveys

For students in classes 4-8 we also conducted, at the time of the assessments, a survey to collect details on assets owned at home, support received at home to complete homework, details of any tuition attended, and the frequency of certain classroom practices and student-teacher engagement.

E Example of a School Improvement Plan (Phase 2 implementation)

MOTTO OF THE SCHOOL: EDUCATION IS A SEA OF KNOWLEDGE									
Domain 1 (A): Resources Available at the School (Availability and Sufficiency)									
Standard	Level	Marked Areas for Improvement	Priority Order	Points of Proposed Actions	Responsible Person at School Level		Expected Departmental Assistance		Timeline
					Name	Designation	From Who	What	
1. School Campus	1	Boundary Wall/ Fencing	Medium	Fence will be made shaping locally available plants like henna, ipomoea, leucina, lantana etc. and will be taken care of.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	03/08/2018
2. School Campus	1	School Garden	Medium	In absence of space for garden, plants will be grown in flower pots. As an alternative arrangement for flower pots, old pitchers, buckets, plastic containers, bottles etc. will be used after decoration. In absence of space for garden and fencing, indoor plants will be grown in flower pots. In case of insufficient space being available, small plants will be grown. The students will be provided plants as rewards.	Mr. XX Mrs. XX Children's Cabinet All Staff	School HM SMC President Members		Financial Assistance	05/01/2017
3. School Campus	1	Space for School Assembly	Medium	The biggest hall of the school will be used for assembly. The assembly will be done at available open space.	All Staff	All Staff	Blank	Blank	01/04/2017

4. Playground, Sports Material & Appliances	1	Sports Material & Appliances	Medium	Material for indoor and outdoor games will be procured.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	12/08/2017
5. Classrooms & Other Rooms	1	Size of Rooms	High	After forming the sections suitable for rooms available in the school, the arrangements will be made that the sizes of the rooms match the ratio of the students. Bigger sized additional rooms will be demanded.	All Staff	All Staff	Blank	Blank	01/04/2017
6. Classrooms & Other Rooms	1	Arrangement of Furniture	High	Sufficient number of furniture will be demanded for teachers and students. Furniture will be arranged with community assistance. Furniture will be acquired through School Gift Scheme (Shala Upahar Yojana).	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	04/01/2018
7. Classrooms & Other Rooms	1	Room for Head of the School	High	One room of the school will be arranged for School Head. Additional room will be demanded.	Mr. XX Mrs. XX	School HM		Financial Assistance	01/09/2017
8. Power & Appliances	1	TV & Radio	Low	TV and radio sets will be demanded. TV and radio sets will be arranged with community assistance. TV and radio will be received through School Gift Scheme.	Mr. XX Mrs. XX	SMC President HM SMC President		Financial Assistance	04/07/2018
9. Power & Appliances	1	Arrangement of Power	Low	Electrify connection will be taken for the school.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	05/01/2018

				Electric fittings will be made in the building, for which help will be taken from community electrician. Electric fittings will be demanded.					
10. Power & Appliances	1	Electric Appliances	Low	Lighting arrangement and fans will be demanded for all rooms. Light and fans will be arranged with community assistance. Light and fans will be acquired through School Gift Scheme.	Mr. XX Mrs. XX	School HM SMC President		Connection will be demanded through...	04/01/2018
11. Library	2	Acquisition of Newspapers & Magazines	High	Local newspapers and other educational magazines will be subscribed. Newspapers and magazines available at teachers'/ students' residence will be collected.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	05/01/2018
12. Library	2	Hall & Reading Space for Library	High	Closed or unused rooms will be developed as library. If rooms are not available, through the block, we will demand that that is included in the district plan.	Mr. XX Mrs. XX	School HM SMC		Financial Assistance	01/09/2018
13. Library	2	Acquisition of Books	High	Books (other than text books) will be acquired @ 10 books per student. Books will be protected from moisture.	Mr. XX Mrs. XX	President School HM SMC		Financial Assistance	01/05/2017
14. Laboratory (where applicable)	1	Procuring Basic Apparatus	High	We will prepare a list of experiments mentioned in the science textbooks for class VI to VIII. According to the list, we will provide apparatus required to do these experiments. We will use mathematics kits to clarify the concepts of mathematics.	All Staff	President All Staff	Blank	Blank	01/03/2017

				We will prepare mathematics kits with easily available material like wood, paper, cardboard etc.					
15. Laboratory (where applicable)	1	Developing Science and Mathematics Laboratory	High	We will develop one of the closed/ unused rooms of the school for laboratory. We will procure instruments and apparatus from the fund available under "Tod-Fod-Jod" (Split, break and assemble) Club programme.	Mr. XX Mrs. XX	School HM SMC President	Blank	Blank	04/01/2018
16. Computer (where applicable)	No Provision								
17. Ramp	1	Ramp Construction	High	Ramp will be constructed as per standards set by the state.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	24/05/2018
18. MDM, Food and Utensils (where food is cooked in school campus)	1	Availability of Kitchen Shed	Low	Kitchen shed will be constructed as per standards set by the state If any unused/ additional room is available, we will be develop it as a store room cum kitchen shed.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	02/02/2018
19. MDM, Food and Utensils (where food is cooked in school campus)	1	Cooking Utensils	Low	In order to cook and serve food, utensils of appropriate sizes will be procured in appropriate number based on number of students. Utensils will be acquired with community participation and under school gift scheme.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	19/04/2018

20. Drinking Water	1	Availability of Sufficient Drinking Water Regularly	High	Sufficient number of pitchers, tanks, drums, cans etc. will be arranged for drinking water storage. In case of tap water supply to the school, help of students, children's cabinet members, teachers, and SMP members will be taken for storage. In case handpump is installed in the school, students will be assigned responsibility on rotation basis. Help of SHG(s) will be taken for drinking water storage.	Children's Cabinet All Staff	Children's Cabinet All Staff	Blank	Blank	01/06/2017
21. Hand washing facilities	1	Water Supply	High	Water for hand-washing will also be stored along with drinking water.	Children's Cabinet All Staff	Children's Cabinet All Staff	Blank	Blank	07/07/2017
22. Hand washing facilities	1	Place for Hand-washing	High	Separate place will be fixed for hand-washing of students. Appropriate arrangement will be made for pre-lunch hand-washing so that all students can wash their hands properly: — Hand-washing in queues – class-wise hand-washing – hand-washing under supervision of teachers/ monitors/ children's cabinet. Based on number of students, arrangement of taps or buckets/ utensils will be made at the place fixed for hand-washing.					

23. Toilets	1	Separate toilet will be made for students with physical disability.	Low	Separate toilet will be made for students with physical disability.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	03/08/2018
Domain 1 (B): Resources Available at the School (Quality and Utility)									
1. School Campus	1		Low						
2. Playground, Sports Material & Appliances	1		Low						
3. Classroom & Other Rooms	1		Low						
4. Power & Appliances	1	Wiring & Switch Board	Medium	Good quality wiring and ISI marked switch boards will be fixed. Their periodic checkup and required repair will be done for safety.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	04/01/2018
5. Library	2		Low						
6. Laboratory (where applicable)	1		Low						
7. Computer (where applicable)	No Provision								

8. Ramp	1	Quality of Ramp	Medium	Ramp will be constructed as per standards set by the state.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	05/02/2018
9. MDM, Cooking & Utensils (where food is cooked in school campus)	1	Cleanliness & Hygiene	High	MDM related utensils will be kept clean. Food will be kept covered. Cooking and food serving place will be cleaned regularly so that food is hygienic. Clean water will be used to wash vegetables, pulses, rice etc.	All Staff	All Staff	Blank	Blank	01/04/2017
10. Drinking Water	1		Low						
11. Hand-washing Facilities	1	Importance of Hand-washing	High	Teachers will explain the importance of hand-washing to students.	All Staff	All Staff	Blank	Blank	20/01/2017
12. Hand-Washing Facilities	1	Monitoring of Hand-washing	High	Hand-washing of students will be monitored regularly with the help of monitor/ children's cabinet.	Children's Cabinet All Staff	Children's Cabinet All Staff	Blank	Blank	02/05/2017
13. Toilets	1	Repair	Medium	Toilets will be made functional through their required repair.	Mr. XX Mrs. XX	School HM SMC President		Financial Assistance	06/05/2017
Domain 2: Teaching-Learning & Their Assessment									

1. Teachers' Understanding about students	2	Discussion on Academic Achievements and Educational Requirements of Students	High	Teachers will introduce the students with their subject-wise educational requirements/ achievements and provide them learning material accordingly. In teacher-parent meeting/ on other occasions, teachers will discuss with parents and introduce them with educational requirements and achievements of students. Educational requirements and achievements of students will be mentioned in the portfolio.	All Staff	All Staff			01/04/2017
2. Teachers' Understanding about Students	2	Individual Distinction based Assistance	High	After recognising individual distinctions of the students, different educational activities will be organised. For instance: Listening activities (stories, poems etc.) for the students understanding by listening. Picture, chart, graph, comics etc. for the students understanding by observing. Individual/ group home-work/ project work will be given to the students based on their interest/ aptitude.	All Staff	All Staff			01/05/2017
3. Teachers' Subject & Educational Knowledge	2	Sharing Subject Knowledge & Teaching Skills	High	School Head will identify the teachers with better required teaching skills. In the weekly meetings, the identified teachers will present before all persons the skills in which they are more competent. The list of identified teachers with their specialization will be provided to the cluster centre so that subject knowledge and teaching skills can be shared at cluster level.	Mr. XX	School HM			01/04/2017

4. Teaching Plans	2	Providing Experience of Educational Programmes Prevailing in the State	High	Educational visits of students will be made to to the neighbouring schools engaged in innovative schemes of the department. Activity Based Learning (ABL)/ Active Learning Method (ALM) Headstart Smart Classes Mathematics-Science Kit The teachers of the schools, engaged in effective use of those schemes, will be invited in our school for experience sharing/ demonstration.	All Staff	All Staff	Blank	Blank	01/12/2017
5. Teaching Plans	2	Teaching Plan Preparation as per Learning Needs	High	Prior to developing teaching lesson plan, teachers will read the learning needs of the students mentioned in the portfolio. Activities suitable for students' learning needs will be identified, like – listening activities for students learning by listening, group work for group learners, etc. learning needs based activities will be included in the plan.	All Staff	All Staff	Blank	Blank	02/03/2017
6. Teaching Plans	2	Teaching Plan Preparation Linking Local Environment	High	Examples based on local language/ dialect, local culture, customs, crops, climate, business etc. will be selected during the teaching plan preparation.	All Staff	All Staff	Blank	Blank	01/03/2017
7. Learning Environment	2	Creating Interactive Environment in Classrooms	High	Students' interaction in learning activities will encouraged and not treated as indiscipline.	All Staff	All Staff	Blank	Blank	06/05/2017

				Those activities will be adopted in which students can learn with their friends, e.g. pair work, role play, group work etc. We will ask such questions which can enable the students linking new knowledge with their previous knowledge while answering the same. We will appreciate the viewpoint of every student.					
8. Teaching-Learning Process	2	Encouraging the Students for Self-study	High	In order to develop reading habit in students, we will provide them easy and interesting books. After reading an essay/ story/ poem individually, in pair or in group, we will tell them to – Repeat the same in their own words. Develop questions based on the same. Develop any similar event or story. Search the solution in a lesson or a book if a problem is given.	All Staff	All Staff	Blank	Blank	01/07/2017
9. Teaching-Learning Process	2	Getting Teaching-aid Prepared by Students	High	We will teach the students to prepare the following things: - Poster - Motto/ thought writing - Static model - Dynamic model - Flash cards - Word wheels etc. After organising competition based on skills required for teaching material preparation, we will reward the students.	All Staff	All Staff	Blank	Blank	01/05/2017

10. Classroom Management	2	Classroom Discipline	High	Teachers and students will form the rules of classroom discipline. Corrective actions for the students avoiding these rules will also be decided by teachers and students. We will associate class monitors to follow the classroom rules.	Children's Cabinet All Staff	Children's Cabinet All Staff	Blank	Blank	01/03/2017
11. Classroom Management	2	Meeting Arrangement	High	As per need of the activity, we will make seating arrangements in queues, pairs, small group circles, or big circles.	All Staff	All Staff	Blank	Blank	01/04/2017
12. Students' Assessment	2	Improvement in Assessment Outcome Based Learning Process	High	In staff meetings and parent-teacher association meetings, we will discuss the required improvement measures in individual social qualities and other sectors.	All Staff	All Staff	Blank	Blank	01/03/2017
13. Students' Assessment	2	Assessment of Different Dimensions	High	The class teacher or subject teacher will make remarks on every student's academic achievement and individual social qualities in the portfolio once in every 15 days.	All Staff	All Staff	Blank	Blank	01/04/2017
14. Use of Teaching-Learning Resources	2	Use of Educational Sources	High	When needed, the students will be shown video lessons compiled on a smartphone. Apart from the textbook content, the teachers will study available learning materials to increase their own and well as students' understanding and use the same linking with the lesson. For training in effective use of educational sources, public teachers (Janshikshak) will apprise BAC, DIET through the School Head.	All Staff	All Staff	Blank	Blank	06/05/2017

15. Use of Teaching-Learning Resources	2	Sharing Educational Sources	High	In weekly staff meetings, we will discuss the use of educational sources and share the same. After forming social media group of teachers, we will share them digital material. We will provide educational material compiled in the school to respective JSK/ cluster. If demanded by other schools, we will share them educational material.	All Staff	All Staff	Blank	Blank	01/06/2017
16. Self-reflection of Teachers on Teaching-Learning Material	2	Classroom Teaching Experiences	High	We will organise monthly reflection meetings, in which the teachers will share their remarkable classroom teaching experiences. We will seek solution to problems in classroom teaching through group thinking. We will change teaching plans and methods based on findings of the thinking.	All Staff	All Staff	Blank	Blank	06/05/2017
Domain 3: Progress, Achievement and Development of Students									
1. Students' Attendance	1	Marking Absent/ Irregular Students	High	We will prepare class-wise list of students who are irregular or absent for a long time. We will also prepare the list of students coming to school late and leaving early.	All Staff	All Staff	Blank	Blank	01/04/2017
2. Students' Attendance	1	Information on Absence	High	Parents will be informed of student's absence/ irregular presence in PTA meetings. Parents will be informed of student's absence/ irregular presence through classmates/ children's cabinet. Parents will be informed of student's absence/ irregular presence through SMS on their mobile phones.	All Staff	All Staff	Blank	Blank	01/04/2017

				Students' absence will be recorded on their diaries and their parents' signature on the same will be taken every month.					
3. Students' Attendance	1	Attendance Record	High	The column of absent student will neither left blank nor marked with (.). Instead, L will be marked for prior information of absence and A will be marked for absence without information.	All Staff	All Staff	Blank	Blank	01/04/2017
				Class-wise data of absence will be displayed in attendance register and school notice board within one hour from opening of the school.					
4. Students' Participation & Engagement	1	Participation in Cultural and Co-educational Programmes	High	A calendar of cultural and co-educational programmes will be prepared.	Mr. XX	School HM	Blank	Blank	02/03/2017
				Cultural and co-educational programmes to be organised in the school will be organised in a way that more and more students could participate and parents could see them.					
				Each student will be assigned different task in the programmes.					
				Inter-school games and cultural competitions will be organised in the JSK schools.					
				Best performing students of cultural and co-educational programmes will be rewarded.					
				One teacher of the school will be selected cultural in-charge. Similarly, students will also be selected class-wise in-charge.					

5. Students' Progress	2	Monitoring of Progress	High	All subject teachers will review the progress of the students from each class in every three months. Along with all subjects, students' co-educational areas and individual social qualities will also be discussed. The ups and downs in progress of each student in all areas will be noted. Based on the findings, students' learning progress and attitudes in the subjects will be noted, like – students of class VII feel difficulty in theorems in Math, or students of class V-A are unable to understand the concept of preposition in English; after annual function, students of class VI-B have started reciting poems, etc.	All Staff	All Staff	Blank	Blank	01/04/2017
6. Students' Progress	2	Change in Teaching Methods	High	Based on learning progress and attitudes, the teachers will make necessary changes in their teaching; for instance: solar/ lunar eclipse by role play, local values by match sticks etc.	All Staff	All Staff	Blank	Blank	01/04/2017
7. Students' Individual and Social Development	1	Discussion with Parents	High	Individual and social indicators will be developed in PTA meetings. Parents will be invited in different celebrations of the school. Parents will be requested to take in these indicators in their domestic environment also.	Mr. XX	School HM	Blank	Blank	01/03/2017
8. Students' Achievements	1	Improving Basic Capabilities	High	In order to acquire basic abilities of Hindi, English and Mathematics, opportunity will be provided for regular dictation, script reading, oral and written everyday mathematics activities and reading library books.	All Staff	All Staff	Blank	Blank	01/04/2017

9. Students' Achievements	1	Identifying Tough Points of Subjects	High	Based on continuous assessment, subject-wise tough points for every child will be identified.	All Staff	All Staff	Blank	Blank	01/04/2017
Domain 4: Performance of Teachers and Their Professional Up-gradation									
1. Orientation of New Teachers	1	Introduction of New Teachers in School	Medium	Newly appointed, promoted or transferred teacher will be introduced with all teachers organising special meeting. Newcomer teachers will be provided information regarding the school, like – what facilities are available at what place in the school or may be provided by which teacher. Newcomer teachers will be explained and assigned their responsibilities.	Mr. XX	School HM	Blank	Blank	04/05/2017
2. Teachers' Attendance	1	Action on Absent Teachers	High	In case of a teacher's absence without information, the school head will assure at his/ her level whether the teacher's absence has any proper reason. Written complain of the teachers absent without any proper or prior information will be made to the JSK in-charge.	Mr. XX	School HM	Blank	Blank	01/09/2017
3. Teachers' Attendance	1	Alternative Arrangement of Teachers	High	For alternative educational arrangement, a register will be maintained, in which period-wise signature of teachers substituting absent/ on-leave teachers will be taken.	Mr. XX	School HM	Blank	Blank	01/06/2017
4. Teachers' Attendance	1	Teachers' Attendance Record	High	All teachers including the school head will sign the teachers' attendance register twice – after arriving and before departing.	Mr. XX	School HM	Blank	Blank	01/03/2017

				Prior information of leave will be provided by application, but in special circumstances, information through M-Mitra App, SMS, Email or telephone will be valid, which will be mentioned in teachers' attendance register. Within half an hour after the school starts, the teachers' attendance register will be finalised.					
5. Targets for Distribution of Functions and Performance	2	Performance Target Setting	Medium	Teachers will set their performance targets themselves. All teachers of the school will discuss the innovations to be done in education and work according to the decisions made. Teachers will monitor their targets themselves and submit its written report to the school head periodically.	All Staff	All Staff	Blank	Blank	01/04/2017
6. Targets for Distribution of Functions and Performance	2	Distribution of Functions	Medium	Teachers will be assigned responsibilities based on their ability, specific interest in work and consent.	Mr. XX	School HM	Blank	Blank	02/11/2017
7. Preparation of Teachers As Per Changing Needs of Syllabus	2	Change in Teaching-Learning Process As Per Change	Medium	Based on findings of their discussion, teachers will introduce required changes in teaching-learning process according to change in syllabus or textbooks. In case of any difficulty in any new content or teaching method, the school head will inform Janshikshak, BAC, DIET.	Mr. XX	School HM	Blank	Blank	01/11/2017
8. Monitoring of Teachers' Performance	2	Review of Teachers' Performance	High	Apart from observation by school head and prescribed proforma, students' progress and achievements will be included in the review of performance and will be discussed with teachers.	Mr. XX	School HM	Blank	Blank	01/03/2017

				The school head will discuss the performance of teachers with students, parents and SMC members and update the teachers with their suggestions.					
9. Monitoring of Teachers' Performance	2	Reflection on Teachers' Performance	High	Teachers will discuss and reflect on their performance in weekly staff meetings. They will help each-other with mutual cooperation and suggestions.	Mr. XX	School HM	Blank	Blank	01/03/2017
10. Professional Up-gradation of Teachers	2	Encouraging Innovation	Medium	Innovations being introduced at various levels in the field of education will be discussed in weekly staff meetings. Teachers introducing any innovation will be rewarded at school level and their activities will be reported to cluster/ JSK and portal so that others are also benefited with the same.	Mr. XX	School HM	Blank	Blank	04/08/2017
11. Professional Up-gradation of Teachers	2	Sharing Training Inputs	Medium	After arrival of a teacher from a training or professional up-gradation programmes, knowledge and skills acquired by him/ her will be shared with other staff teachers. Teaching material received in these programmes like modules, handouts, CDs, digital media, softwares will be shared with all colleague teachers.	Mr. XX	School HM	Blank	Blank	01/10/2017
Domain 5: School Leadership and School Management									
1. Vision & Direction Setting	2	Preparation of School Up-gradation Action Plan	High	Suggestions of children's cabinet and SMC members will be invited for preparing school up-gradation action plan. Their suggestions for the action plan will be noted and included In the same based on their merit.	Mr. XX	School HM	Blank	Blank	02/05/2017

2. Vision & Direction Setting	2	Review of Up-gradation Action Plan	High	Review meetings will be organised to review the completion and quality of work as per priority, in which members of SMC and children's cabinet will also participate. Solution of incomplete works and difficulties will be sought through discussion. If needed, Janshikshak/ BAC/ BRC/ DIET will be informed for additional assistance.	Mr. XX	School HM	Blank	Blank	05/05/2017
3. Leading for Change & Improvement	2	Execution of Change	High	In order to effect change in identified areas, we will allocate the responsibilities on the basis of interest, experience and ability. We will record the changes occurring during the implementation.	Mr. XX	School HM	Blank	Blank	06/04/2017
4. Leading for Change & Improvement	2	Making Aware on Requirements of Change	High	School head will organise workshops time to time to update all teachers, students, children's cabinet members and SMC members with the changes occurring in the field of education and their requirements. Understanding towards these changes will be developed during these workshops, which will be mentioned as reports at the end of the workshops. During these workshops and meetings, clear goals will be set for the school and improvement strategy will be designed.	Mr. XX	School HM	Blank	Blank	02/05/2017
5. Leading Teaching & Learning	2	Reflection on Students' Progress	High	In every three months, the school head will lead collective thinking with respective teachers on students' progress.	Mr. XX	School HM	Blank	Blank	02/05/2017

				During the collective thinking, popular methods or teaching and learning will be considered to be continued or changed.					
6. Leading Teaching & Learning	2	Improvement in Teaching-Learning Methods	High	Class-wise and subject-wise 'Learning Indicators' will be compiled and teachers will be told to study the same. Available literature on student-centric methods and innovation based learning will be compiled and teachers will be told to study the same. After thinking on needs of improvement in teaching and learning, effective methods will be adopted to effect positive changes in the same.	Mr. XX	School HM	Blank	Blank	02/05/2017
7. Leading the School Management	2	Collective Assessment of Progress	High	The responsibility to act on and monitor the areas of improvement identified in the school up-gradation action plan as per point of action will be assigned to staff and SMC members according to their interest, consent and ability.	Mr. XX	School HM	Blank	Blank	05/04/2017
8. Leading the School Management	2	Designing Strategy for Management	High	THe school head will hold regular discussion with teachers, parents, children's cabinet, and SMC members. During the discussion, they will be invited to provide suggesions for development and progress of the school. Based on suggestions received, the strategy for development and progress will be designed as school up-gradation action plan.	Mr. XX	School HM	Blank	Blank	02/05/2017
Domain 6: Inclusion, Health and Safety									
1. Environment for Inclusion	2	Equal Opportunity for Participation	High	In any situation, the students will not let feel that they have been marked due to any special ability or weakness.	Mr. XX	School HM	Blank	Blank	02/05/2017

				In different educational programmes like lecture, essay competition, etc. and in co-educational programmes like games and sports, cultural programmes etc., the three groups will be provided equal opportunity to participate. Teachers' consultation sessions will be organised in which they will guide the students to participate in the activities based on their ability and interest.					
2. Inclusion of Students with Special Needs (SWSN)	2	Inspiring SWSN	Medium	In children's assembly and prayer assembly, we will tell inspiring stories on success of SWSN. We will arrange screening of inspiring films based on SWSN.	All Staff	All Staff	Blank	Blank	01/03/2017
3. Inclusion of Students with Special Needs (SWSN)	2	Training of Teachers	Medium	JSK/ cluster/ DIET will be made aware with the demand for teachers' training for inclusion of SWSN in the classes of other students.	Mr. XX	School HM	Blank	Blank	25/05/2017
4. Students' Safety	2	Disaster Management	Medium	In order to deal with disasters, a campaign will be organised to create awareness in the school and community through songs, plays, films, discussion, rally etc. A plan will be prepared to deal with emergencies and the plan will be reviewed time to time. For emergency exit, additional doors/ gates will be constructed in classrooms or school building.	All Staff	All Staff	Blank	Blank	07/07/2017
5. Students' Safety	2	Safety Arrangements	Medium	Phone/ mobile numbers for emergency contact will be kept stored in mobile phones of all teachers. Contact numbers of parents will be recorded for making contact with them.	All Staff	All Staff	Blank	Blank	05/07/2017

				After developing contacts with Nirbhaya unit, we will get demonstration and training for assistance and safety.					
				For repair of building for safety reasons, we will inform the JSK/ BRC/ DPC/ DEO at once.					
				We will get help of village panchayat/ municipality for safety from rodents, insects, and venomous animals.					
6. Emotional Safety	2		Medium	Counseling sessions for students and parents will be organised to reduce the fear of study and good performance.	Mr. XX	School HM	Blank	Blank	04/07/2017
				Academic assistance will be received from DIET faculty members towards age based adolescent education.					
				Counseling sessions for students and parents will be organised on adolescence related problems.					
				After developing contacts with Nirbhaya unit, we will get demonstration and training for assistance and safety.					
				Any sensitive female teacher will be given the responsibility of emotional counseling, grievance redressal and dealing with requirements.					
7. Health and Hygiene	2	Creating an Environment for Health, Sanitation & Hygiene	High	Teachers, students and SMC members will be given responsibility for monitoring of health, sanitation & hygiene.	Children's Cabinet All Staff	Children's Cabinet All Staff	Blank	Blank	01/04/2017
				A committee will be formed to implement the Clean School Scheme, which will organise awareness programmes like workshops, street shows, slogans, film screening etc. for students and parents.					
				Health workers, doctors, and voluntary organisation will be linked with these events.					

				If any health problem is noticed in students, we will inform and counsel the parents.					
Domain 7: Community Participation									
1. SMC Formation & Management	1	Meeting Agenda Setting	High	Any SMC meeting will not be organised without setting an agenda.	Mr. XX	School HM	Blank	Blank	05/09/2017
				The school head will set the agenda of the SMC meeting based on teachers, children's cabinet members, requirements of school and other sources like feedback from parent-teacher association.					
				The SMC members will be made aware of the agenda before the meeting.					
				Apart from financial and basic issues, educational, co-educational and other issues will be included in the agenda.					
				Every month, at least one SMC meeting will be organised with prior information and fixed agenda.					
2. SMC Formation & Management	1	Presence of Members in the Meeting	High	While fixing the date/ time of the meeting, it will be kept in mind that there should be no special festival or fair on the day of meeting in which the members are likely to be busy.	Mr. XX	School HM	Blank	Blank	05/08/2017
				Respective SMC members will be informed of the date and time of the SMC meeting through the teachers or children's cabinet members.					
				In SMC meetings, incentive programmes like, distribution of free uniforms, textbooks, money for bicycles, etc. will be done by the members.					

				For attending meeting, the members will convince students to try to call their parents in the meeting. On these days, the school will organise cultural activities and rewards will be distributed by members.					
3. SMC Empowerment	1	Compliance of Educational Programmes/ Provisions/ Acts	High	We will make the SMC members aware towards education related programmes, provisions and acts like Right to Education, SSA, incentive schemes etc. For awareness, we will discuss with members, read out documents and explain them, and if facilitated, we will provide them handouts. We will deliver the message of provisions and acts to the members through street plays. We will request the members that in case of violation of provisions/ acts, they inform the school head immediately.	Mr. XX	School HM	Blank	Blank	04/04/2017
4. SMC Empowerment	1	Preparation & Implementation of School Up-gradation	High	We will involve the SMC members in preparation of school up-gradation action plan and respect their ideas/ suggestions. After preparation of the plan, we will inform the members in SMC meeting about the same.	Mr. XX	School HM	Blank	Blank	12/08/2017
5. School-Community Correlation	2	Action Plan Collaboration with Industry Houses/ Community	High	After providing the list of school requirements, the SMC members will be told to contact neighbouring NGOs/ corporate bodies (big industries) and alumnis and get their help. In order to fulfill the requirements, we will publicize the School Gift Scheme and get its benefits.	Mr. XX	School HM	Blank	Blank	06/04/2017

6. Community as Learning Resource	2	Linking Local Knowledge/ Skills with Subject	High	Subject teachers will mark the topics in the syllabus that can be linked to the local milieu. For instance, aquatic animals for students of riverside schools, birds and animals for near forest schools, industry related information for schools of industrial towns/ cities etc.	Mr. XX	School HM	Blank	Blank	12/05/2017
7. Community Empowerment	2	Organising the Programmes Jointly	High	School and community will organise various programmes jointly. The venue may be school or any local place. Active participation of community members will be attempted in these programmes as in program chairmanship, prize distribution, moderation, vote of thanks.	Mr. XX	School HM	Blank	Blank	18/05/2017