RESEARCH ARTICLE

Using Low-Tech, Behaviorally Informed Solutions to Improve Teacher Adoption of Foundational Pedagogical Practices: Evidence from a Field Experiment in India

[version 1]

Sneha Shashidhara, Sonal Garg

Centre for Social and Behaviour Change, Ashoka University, New Delhi, New Delhi, 110057, India

v1 **First published:** 15 Sep 2025, **2**:283
https://doi.org/10.12688/verixiv.1853.1

Latest published: 15 Sep 2025, **2**:283
https://doi.org/10.12688/verixiv.1853.1

Abstract

Background

Early-grade teachers in India are now supplied with structured pedagogy resources, yet day-to-day uptake remains low. We tested two low-tech, behaviourally informed strategies delivered over WhatsApp to simplify access to those resources and nudge adoption of evidence-based classroom practices.

Methods

In a cluster-randomised controlled trial spanning 216 Nyay Panchayat clusters (N = 1,872 Grade-3 teachers) in rural Uttar Pradesh, schools were allocated to: (1) a chatbot that served bite-sized, audio-enabled lesson-plan summaries and habit-building reminders; (2) fortnightly micro-practice videos showing a relatable teacher-actor modelling targeted techniques; or (3) business-as-usual control. Primary outcomes—teachers' adoption of prescribed practices, valuation of teaching aids, and knowledge of structured-pedagogy content—were measured at baseline (September 2023) and after six months (April 2024).

Results

Neither intervention moved the primary outcomes for the full sample

($p > .10$). Exploratory subgroup analyses, however, revealed modest gains in the teacher guides' valuation, motivation, and beliefs about effective practice. Negative or null effects were concentrated in Hardoi, where union resistance suppressed engagement. Attrition and baseline covariates were balanced across arms.

Conclusions

Making evidence-based resources "easy, attractive, social, and timely" via WhatsApp was not sufficient, on its own, to shift practice at scale. Light-touch digital supports need complementary in-person coaching and systemic problem-solving, mainly where local credibility, systemic burdens or labour-relations issues constrain engagement.

Keywords
Foundational literacy & numeracy; teacher professional development; WhatsApp chatbot; micro-practice video; behavioural science; cluster RCT.

Corresponding author: Sneha Shashidhara (sneha.shashidhara@ashoka.edu.in)

Author roles: Shashidhara S: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Garg S**: Data Curation, Formal Analysis, Project Administration, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project was funded (INV-053842) by the Bill & Melinda Gates Foundation (BMGF).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Shashidhara S and Garg S. **Using Low-Tech, Behaviorally Informed Solutions to Improve Teacher Adoption of Foundational Pedagogical Practices: Evidence from a Field Experiment in India [version 1]** VeriXiv 2025, **2**:283 https://doi.org/10.12688/verixiv.1853.1

First published: 15 Sep 2025, **2**:283 https://doi.org/10.12688/verixiv.1853.1

Introduction

Foundational literacy and numeracy (FLN)—the ability to read connected text and perform basic arithmetic by Grade 3—underpin later learning (Belafi et al., 2020; Darling-Hammond, 2006). Yet many children worldwide, especially from disadvantaged households, complete primary school without these skills (World Bank, 2018; Glewwe & Muralidharan, 2016). Therefore, India's National Education Policy (NEP, 2020) mandates universal FLN by Grade 3 (MHRD, 2020). However, 84% of Grade-3 pupils in rural government schools still cannot read Grade-2 text, and 78% cannot subtract two-digit numbers (ASER Centre, 2022). Improving teachers' day-to-day practice is critical because instruction quality strongly predicts learning, self-efficacy, and long-run outcomes (Bau & Das, 2020; Chetty et al., 2014; Kraft, 2019; OECD, 2020).

Teacher professional development in low- and middle-income countries (LMICs) needs reform to prioritise evidence-based practices and focus on core teaching skills rather than solely on content knowledge (Molina et al., 2018). Structured-pedagogy programmes—packages that couple scripted teacher guides, training, classroom materials, formative assessment and mentorship—have raised learning twice as much in low- and middle-income countries (LMICs) as in the United States (average d = 0.44 vs 0.22; Stern & Piper, 2019; Chakera et al., 2020). Core practices include explanation, modelling, guided and independent practice, and ongoing student assessment (Piper & Dubeck, 2024). India's NIPUN Bharat mission has begun scaling such resources; Uttar Pradesh, for example, supplies teacher guides, workbooks, training and mentorship to all primary teachers (Sharma Kukreja, 2024; CSBC, 2023). Adoption, however, remains stubbornly low: heavy administrative loads (Sankar & Linden, 2014; Bhatty & Saraf, 2016), large multigrade classes and prevailing beliefs that learning is limited by home background (CSF, 2021; CSBC, 2023) all depress uptake.

Behavioural science suggests that present-bias, information overload and misperceived social norms can further deter change (Hahn et al., 1992). Guided by these insights, this study tests two low-tech, light-touch WhatsApp strategies designed to make existing structured-pedagogy resources Easy, Attractive, Social and Timely, to boost adoption. Grade-3 teachers in 216 *Nyay Panchayat* clusters were randomised to (i) a chatbot delivering bite-sized, audio-enabled lesson-plan summaries, (ii) fortnightly micro-practice videos (MPVs) demonstrating targeted techniques, or (iii) business-as-usual control. We preregistered three hypotheses: the interventions would raise (H1) knowledge, (H2) valuation and (H3) classroom adoption of prescribed practices. As secondary outcomes, we also evaluated shifts in teachers' motivation, beliefs and attitudes toward structuring pedagogy resources into their daily teaching practices. The outcomes were measured at baseline (September 2023) and after six months (April 2024).

By simplifying access rather than adding tasks, we ask whether light-touch digital nudges can shift practice in resource-constrained public schools.

Literature review

Barriers to effective teacher engagement

Teaching is widely reported as highly stressful—around one quarter of teachers call their job "very or extremely" stressful (Travers & Cooper, 1996; Dunham & Varma, 1998; Kyriacou, 2000). Stressors include unmotivated students, discipline, time pressure, role conflict and poor working conditions (Benmansour, 1998; Pithers & Soden, 1998). In LMICs, these pressures intensify: Indian primary teachers devote 20–25% of work hours to non-teaching duties such as data entry and mid-day-meal oversight (Sankar & Linden, 2014), and many resent such tasks (CETE, 2023). Training quality is often weak (World Bank, 2018; Mbiti, 2016; Ramachandran et al., 2016), and resources are cumbersome—teachers in both South Africa and Uttar Pradesh found FLN teacher guides text-heavy and overwhelming (Flaschen et al., 2024; CSBC, 2023). Such challenges can lead to teaching inefficiencies and a lack of necessary knowledge (Bold et al., 2017). Another issue is ineffective supervision, which weakens accountability as such inspections focus on administrative duties over teaching, with teachers prioritising record completion over learning (Mbiti, 2016; Ramachandran et al., 2016). Behavioural barriers compound structural ones. Teachers may procrastinate, cling to status-quo routines, underestimate peer uptake of new methods, or attribute low learning to pupils' backgrounds (CSBC, 2023; CSF, 2021; Menon et al., 2017). Many believe they cannot help children of uneducated parents (Sabarwal et al., 2022; Kaur, 2023), and such beliefs correlate with lower effectiveness (Filmer et al., 2021). Tackling both structural and behavioural obstacles is therefore essential.

Interventions to improve practice

Structured-pedagogy packages integrate lesson plans, intensive training and coaching. Kenya's PRIMR raised early-grade literacy via scripted guides plus 1015 days of training, and ongoing feedback (Piper et al., 2014). Variants combining resources with behavioural nudges also show promise: in U.S. middle schools, access to online lessons modestly improved math (d = 0.06 SD), but adding reminder emails and peer-collaboration webinars achieved even greater gains (d = 0.09; Jackson & Makarin, 2018). Their findings indicated that reminders and support kept teachers engaged to put lessons into practice instead of procrastinating. In Kenya, weekly instructional SMS plus motivational messages enhanced literacy instruction and reading scores (Jukes et al., 2017).

Low-cost digital delivery scales reach. UNESCO pilots in Nigeria and Pakistan sent short daily texts, images and videos that teachers reported improved activity-based learning (McAleavy et al., 2018). Bangladesh's English in Action pushed audio-visual content via cheap phones and SD cards in remote areas; primary-teacher English use rose from 33% to 71% of lesson time (McAleavy et al., 2018). Tablet-based lesson libraries lifted achievement in Kenya (Gray-Lobe et al., 2022) and Pakistan's eLearn classrooms (Beg et al., 2022). MOOCs likewise help Indian and Filipino teachers upskill (Wolfenden et al., 2017; Oakley et al., 2018).

Complementary social-motivation levers also show a meaningful effect. Peer-recognition banners, certificates and WhatsApp poll discussions boosted teacher motivation in India (STiR & IDinsight, 2018) and Delhi's Teacher Development Coordinator programme (Teotia & Kumar, 2019). Integrated online-offline coaching models yield even larger learning gains (de Hoop et al., 2025).

Gaps addressed by the present study

Government teachers in India already receive training, mentoring, scripted guides and workbooks, yet adoption remains low due to the intertwined structural and behavioural barriers outlined above (CSBC, 2023). Evidence on ultra-light, mobile-first behavioural supports in such contexts is still sparse by testing WhatsApp-based chatbot and MPVs—each embedding reminders, social proof, loss-aversion stickers and peer modelling—our cluster-RCT probes whether simplifying information alone can unlock the potential of structured-pedagogy resources, and for whom.

Methods

Experimental design & sample size estimation

We ran a three-arm, cluster-randomised controlled trial in two rural districts of Uttar Pradesh, India—Sitapur and Hardoi. Randomisation was conducted at the level below the block, Nyay Panchayat (NP), a cluster of villages. A total of 216 NPs (72 per arm) were randomly assigned to (i) a WhatsApp chatbot, (ii) fortnightly micro-practice videos (MPVs) or (iii) business-as-usual control (Figure 1). The Ashoka University IRB approved the study (23-E-10065-Sharma) and pre-registered on the AEA/Social-Science Registry (#12060).

From ~5,850 government primary schools, we randomly drew **10 schools per NP plus two replacements**, targeting 2,592 across 18 blocks each in Hardoi and Sitapur. Grade 3 teachers from the selected schools formed the study participants. Resistance from one teachers' union forced us to drop 5 of the 18 Hardoi blocks. A total of 1,872 teachers completed baseline (Sept 2023) and 1,601 completed end-line (Apr 2024) with an overall attrition of 15%, which did not differ by arm. Baseline covariates (age, gender, experience, etc.) were balanced across arms.

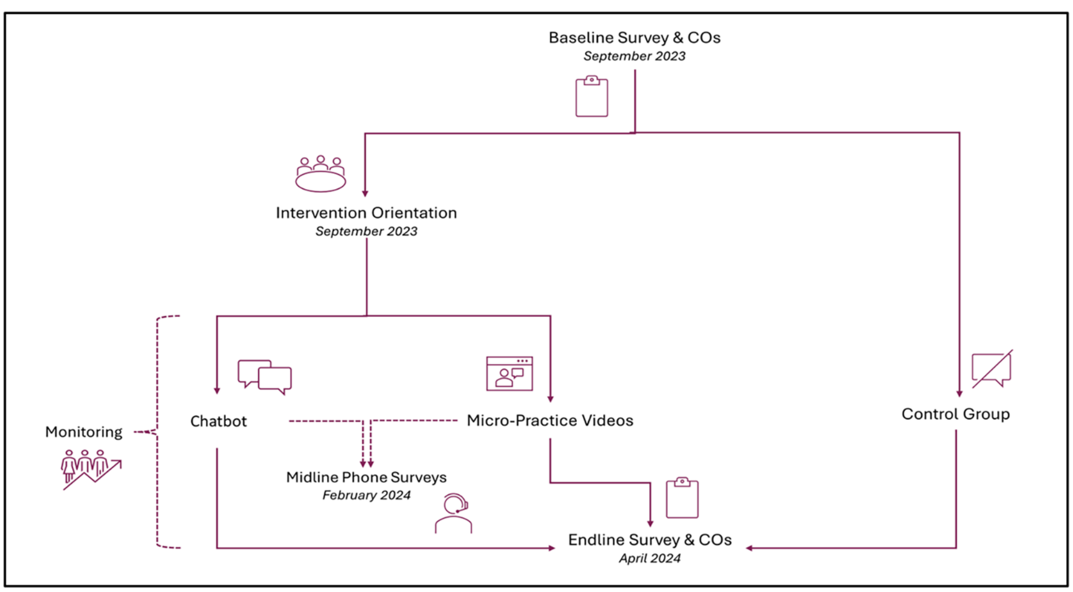

Figure 1. Experimental design.

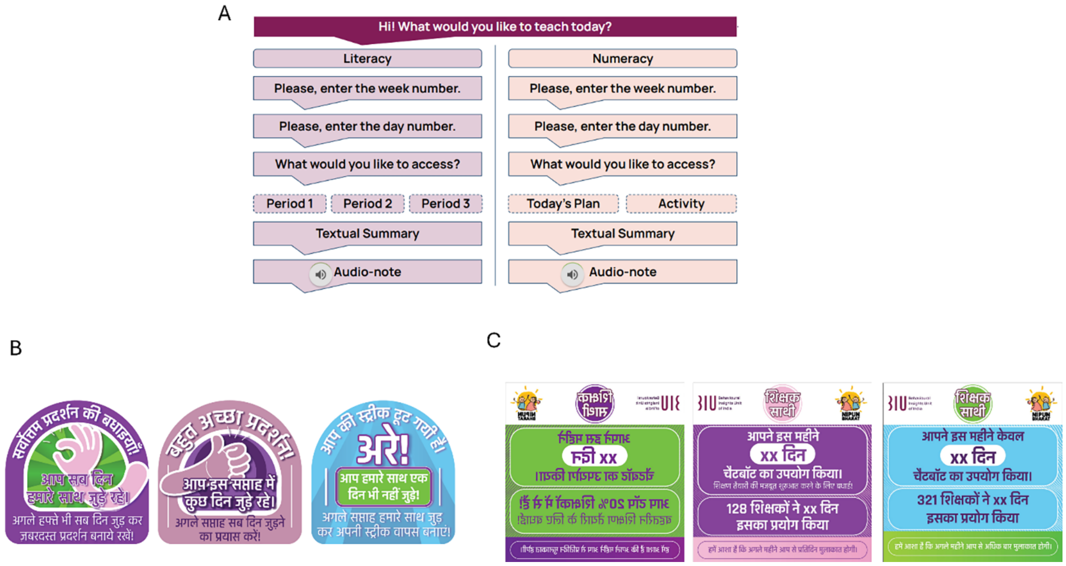

Figure 2. WhatsApp chatbot intervention design. A. Conversation flow of the WhatsApp Chatbot. B. 3 types of weekly streak stickers in the order: (1) Recognising users who regularly use the bot by congratulating them. (2) Encouraging the mid-level users to use the bot more frequently. (3) Reminding the inactive users of their broken streak and asking them to use the bot. C. Monthly report card types in the order: (1) Recognising the top 20% of users, (2) Appreciating the 50-80th percentile, (3) Reminding the bottom 50% of low engagement compared to others.

Interventions

Both interventions were delivered entirely through WhatsApp and were built on the "EAST" behaviour-change framework (BIT, 2014), aiming to make state-issued FLN resources easy, attractive, social and timely. The first arm used a WhatsApp chatbot that served bite-sized text and audio versions of the government's daily Grade-3 teacher-guide lessons of math and Hindi. Teachers summoned a lesson with a few taps, received an early-morning prompt ("What would you like to teach today?") daily, and accumulated weekly "streak" stickers. Every Sunday, all users got a sticker congratulating consistent engagement or highlighting a broken streak to evoke loss aversion, a behavioural strategy (Kahneman and Tversky, 1977) to encourage action by leveraging people's tendency to avoid losses more strongly than they seek gains. Monthly report cards that compared individual usage with peers and occasional testimonial videos from mentors or high-performing teachers were also shared to reinforce social proof (Figure 2).

The second arm comprised twelve live-action micro-practice videos (≈3 minutes each) that modelled evidence-based techniques such as balanced-literacy routines, eliciting student responses effectively and the Gradual-Release-of-Responsibility cycle. Videos were posted to closed WhatsApp groups of fifteen to twenty teachers every fifteen days. Between releases, polls, infographics and view-count snapshots were shared to keep the teachers engaged. Additionally, fortnightly "Leader Banners" and bi-monthly certificates publicly recognised the most active participants on the WhatsApp groups (Figure 3). Although a few clips referenced the teacher guide, they focused on making discrete practices salient rather than on the guide itself. In contrast, the chatbot intervention promoted teacher guide lesson plans by nudging teachers toward default use of the practices on which the TGs were based.

Data collection

Data collection began with an in-person, self-administered baseline survey in September 2023 at each Block Resource Centre. Teachers completed a 40- to 50-minute tablet questionnaire that measured knowledge, adoption, valuation, confidence calibration, motivation, beliefs and demographics. They provided written consent as a part of both baseline and endline surveys. Treatment teachers then received a 20- to 30-minute orientation: chatbot teachers practised the bot's menu flow, and MPV teachers watched a trailer and were enrolled in their WhatsApp group. Control teachers provided feedback to mask their status. The interventions ran from October 2023 to March 2024, with a phone-based mid-line perception survey with a sub-sample five months in and an in-person end-line in April 2024. Attrition from 1,872 baseline teachers to 1,601 at end-line (13.6 per cent for chatbot, 13.3 per cent for MPV, 16.2 per cent for control) did not differ significantly across arms. Back-checks with 10–12 per cent of respondents confirmed data accuracy. Interventions were closely monitored using engagement dashboards, and minor adjustments such as modifying messages' frequency and timing were made to course correct. Data, codes, and supplementary sections are available in the OSF repository (https://osf.io/3r48g/?view_only=ecfc1480faef4df2abed903720cf095e).

Figure 3. Micro-practice video intervention design. A. Snippet from a micro-practice video that shows our classroom set-up, and the breakdown of each practice into three steps, with step 1 highlighted on the screen. B. An infographic summarising the steps shown in the video. C. Teachers participating in polls in the WhatsApp groups. D. Certificate appreciating the bi-monthly top performers. Panel A is a screenshot from the micro-practice video showing a teacher-actor modelling a targeted technique, and the actors (teachers and students) gave their permission to be published in academic research.

Sample size determination

Drawing on the literature and experimental studies with similar research and study populations (Piper et al., 2014; Muralidharan and Sundararaman, 2010; STiR and IDinsight, 2018), the sample size for the current study was determined using power calculations. Previous studies suggested sample sizes ranging from 100 to 300 schools per arm. Power calculations assumed 72 Nyay-Panchayat clusters per arm, an ICC of 0.33 and SD = 10, yielding a minimum detectable effect (MDE) of 0.30 SD. A teachers' boycott in five Hardoi blocks reduced the realised sample, but baseline data showed an ICC of 0.04 and SD $\leq$ 1.26, improving the post-hoc MDE to 0.18 SD (0.19 SD with 25 per cent attrition) across 183 clusters and roughly ten teachers each.

Outcomes measured

The primary outcomes assessed effects on adoption, knowledge and perceived value of the prescribed techniques and resources. These were (i) adoption, captured through six scenario vignettes that required choosing the action aligned with accurate application of teacher guide and effective teaching practices; (ii) knowledge, assessed with four direct content questions on teachers guides and associated practices; and (iii) valuation, elicited via a revealed-preference game in which teachers picked either the printed guide, a 30-minute training session on effective teaching techniques or a sliding cash amount (₹0–₹500).

Secondary outcomes included a confidence-bias score (over- or under-confidence relative to vignette accuracy), a binary motivation indicator derived from a fourteen-item job/teaching-related index, and four belief items contrasting traditional with new student-centred teaching methods. Additional self-efficacy, intention, and self-reported student progress measures were also included.

A small sample of classroom observations was also conducted as described in the Supplementary; baseline classroom observations were discarded after quality checks, and endline observations ($\approx$30 schools per arm) are reported only descriptively because of limited power.

Additionally, a phone-based midline perception survey was administered five months into the intervention to approximately 200 treatment-arm teachers (100 per arm) with varying levels of engagement to understand the challenges to uptake and elicit perceptions of the interventions while teachers are still using them.

Model specifications

Analyses followed the pre-registered intent-to-treat plan. Each outcome was regressed on treatment indicators, its baseline value, a vector of teacher demographics (age, gender, caste, religion, experience, designation, education, household size, and school staff size[1]) and district fixed effects; standard errors were clustered at the Nyay-Panchayat level. As appropriate, continuous, binary, and ordinal outcomes were estimated using OLS, logistic, or ordered-logistic models. The regression equation is as follows.

$$\mathbf{y}_i = \boldsymbol{\beta}\mathbf{T}_s + \boldsymbol{\gamma}_1 \mathbf{y}_i^B + \boldsymbol{\gamma}_2 \mathbf{X}_i + \boldsymbol{\delta}_\beta + \boldsymbol{\varepsilon}$$

Where y_i is the relevant outcome for the participant, T is an indicator variable capturing whether the participant is assigned to the appropriate treatment arm or control. β is the coefficient of interest - it captures the effect of the treatment on the outcome. $\mathbf{y}_i^B$ is the outcome for the participant at baseline. X_i is demographic control. δ_β are the district-fixed effects. ε are the standard errors clustered at the *Nyay Panchayat* level.

The outcome variables were constructed as follows. The knowledge and vignette scores were calculated as proportions of correctly answered questions out of 4 and 6, resulting in discrete numeric values ranging from 0 to 1. The perceived valuation of the TG was measured by the amount teachers chose for the TG over monetary compensation. Similarly, the perceived value of 30-minute training was measured by the amount selected for training over financial compensation.

Confidence bias was measured by having teachers rate their confidence in their response (0%, 25%, 50%, 75%, or 100%) after answering a vignette question. A bias score was calculated by comparing confidence with accuracy for each question (the difference between confidence and accuracy, divided by the number of questions). A positive score indicated overconfidence, while a negative score indicated underconfidence, with values ranging from -1 to 1. A motivation score was calculated by averaging responses across 14 statements, with agreement coded as one and neutrality or disagreement coded as 0. For the analysis, we used a binary outcome, assigning a value of 1 for those whose motivation scores improved from baseline and zero otherwise.

For four questions assessing beliefs on student-centred teaching, the outcome was binary: one if the teacher preferred new approaches and zero if they favoured traditional methods. For the intention to use TG and effective practices, as well as the self-rated ability rating to contribute to student learning goals (measured on a 5-point ordinal scale), the outcome was coded as one if the rating improved from baseline and zero otherwise. For the other two self-reports, teachers indicated the percentage of students achieving FLN goals in the previous year. They predicted to achieve in the current year using an ordinal scale (<50%, 50%, 51–70%, >70%), coded as 1 to 4, respectively.

All subgroup comparisons, such as those between teacher types and between districts, were exploratory in nature. We examined **heterogeneity** along two fault lines. **Districts**: Sitapur and Hardoi, the latter showing a union-led push-back at baseline that likely dampened engagement. **Teacher status**: *permanent* Assistant Teachers (ATs) versus *contractual* Shiksha Mitra (SMs). ATs reported monthly pay of roughly ₹4060 k and easier access to training, whereas SMs, hired to plug shortages with limited access to training, earned about ₹10–20 k. These structural gaps could shape uptake and impact, so subgroup analysis was essential.

As a robustness check, we re-ran every model on the ≈72% of teachers who passed an attention check question at both waves of data collection; results were consistent with the full sample (Tables 2–11).

We also did a **factor analysis** to investigate further the factors driving our primary endline outcome measure. We used the same variables as before and dropped age due to its correlation with teaching experience. After list-wise deletion of rows with missing data, we (i) z-scored every remaining numeric column so that variables were on a standard scale, then (ii) estimated a maximum-likelihood exploratory factor model. The outcome was a combined knowledge + vignette score. The number of factors was chosen with the Kaiser criterion (eigen-value > 1), which—under the final variable roster—yielded 11 substantive factors. We rotated the unstandardised loading matrix with an orthogonal varimax rotation to achieve a simpler, more interpretable structure and calculated regression factor-scores and correlation scores for interpretation.

[1]Due to the high non-response rate, teacher salary and household income could not be included as covariates. However, Cramér's V association tests revealed association between these variables and teacher designation. Hence, we excluded these variables from the final regression analysis and used teacher designation only instead.

Results

Sample demographics

At endline, 47.6% teachers were women, 40% belonged to the Other Backward Classes, the median age was 39, 89.2% identified as Hindu, and their households averaged five members. One quarter declined to report salary, but of those who did, 46% earned more than ₹40,000 (US$470 per month); 39% withheld household income, while 41% reported above ₹45,000 (US$530). Eighty-six per cent held at least a bachelor's degree and reported roughly twelve years of teaching experience; one teacher in three had worked in a private school. The cadre split was 54% permanent Assistant Teachers and 32% contractual Shiksha Mitra. Schools averaged five staff members and classes 23 pupils, with 78% of teachers handling multigrade rooms. Daily time-use reports showed 1–2 hours commuting, 4–6 hours teaching, less than an hour on non-teaching duties, 2–4 hours household work and 1–2 hours lesson preparation. Slightly more than half used WhatsApp for under an hour daily, yet 93% searched online for teaching materials. Balance tests (Table 1A) reveal no

Table 1A. Summary statistics of demographics and balance check results.

	(1) Chatbot		(2) MPV		(3) Control		t-statistic		
Variable	**N**	**Mean [SD]**	**N**	**Mean [SD]**	**N**	**Mean [SD**	**(1)-(2)**	**(2)-(3)**	**(1)-(3)**
Age	551	39.29	543	39.06	498	39.7	0.515	1.367	0.902
		[7.188]		[7.562]		[7.475]			
Household Size	550	5.893	547	5.958	500	5.876	0.413	0.491	0.103
		[2.555]		[2.679]		[2.722]			
Experience (years)	552	11.96	547	11.75	502	12.52	0.483	**1.722**+	1.265
		[7.016]		[7.037]		[7.329]			
School Staff Size	552	4.587	546	4.52	502	4.47	0.53	0.407	0.928
		[2.137]		[2.035]		[1.93]			
Variable	**(1) Chatbot**		**(2) MPV**		**(3) Control**		**chi-square statistic**		
Gender	**N**	**Prop**	**N**	**Prop**	**N**	**Prop**	**(1)-(2)**	**(2)-(3)**	**(1)-(3)**
Male	309	0.5608	273	0.5009	255	0.509	**3.708**+	0.039	2.628
Female	242	0.4392	272	0.4991	246	0.491			
Salary									
(Did not answer)	136	0.2464	136	0.2486	127	0.253	1.119	6.629	7.137
(less than 10k)	70	0.1268	64	0.117	59	0.1175			
(10,000-20,000)	80	0.154	81	0.1481	95	0.1892			
(20,001-40,000)	2	0.0036	2	0.0037	5	0.001			
(40,001-60,000)	170	0.308	166	0.3035	127	0.253			
(60,001-80,000)	77	0.1395	82	0.1499	75	0.1494			
(more than 80,000)	12	0.0217	16	0.0293	14	0.0279			
District									
(Hardoi)	233	0.4221	233	0.426	212	0.4223	0.005	0.003	0
(Sitapur)	319	0.5779	314	0.574	290	0.5777			
Caste									
(General)	192	0.3478	204	0.3729	183	0.3645	4.603	0.227	**5.654**+
(0BC)	207	0.375	222	0.4059	211	0.4203			
(SC/ST)	153	0.2772	121	0.2212	108	0.2151			
Religion									
(Hindu)	496	0.8986	484	0.8848	448	0.8924	0.403	0.086	0.05
(Other)	56	0.1014	63	0.1152	54	0.1076			

Table 1A. *Continued*

	(1) Chatbot		(2) MPV		(3) Control		t-statistic		
Variable	**N**	**Mean [SD]**	**N**	**Mean [SD]**	**N**	**Mean [SD**	**(1)-(2)**	**(2)-(3)**	**(1)-(3)**
Household Income									
(Did not answer)	208	0.3768	213	0.3894	200	0.3984	7.65	1.595	5.797
(30,000)	110	0.1993	89	0.1627	93	0.1853			
(30,001-45,000)	8	0.0145	12	0.0219	11	0.0219			
(45,001-60,000	113	0.2047	96	0.1755	82	0.1633			
(60,001-75,000)	63	0.1141	69	0.1261	59	0.1175			
(75,001-100,000)	30	0.0543	38	0.0695	30	0.0598			
(More than 100,000)	20	0.0362	30	0.0548	27	0.0538			
Education									
High School/ Diploma	78	0.1413	83	0.1517	66	0.1315	1.418	4.211	2.691
Graduate (BA/BSc)	176	0.3188	175	0.3199	184	0.3665			
Bed	142	0.2572	125	0.2285	122	0.243			
Masters/PhD	156	0.2826	164	0.2998	130	0.259			
Type of Teacher									
Assistant Teacher	309	0.5598	309	0.5649	268	0.5339	2.211	1.245	2.392
Shiksha Mitra	165	0.2989	158	0.2888	160	0.3187			
Teacher-In charge	44	0.0797	36	0.0658	34	0.0677			
Head-teacher	34	0.0616	44	0.0804	40	0.0797			

We conducted t-tests for continuous variables and chi-square tests for categorical variables.
Statistical significance level:
$^{+}p < 0.10$.

Table 1B. **Attrition test.**

	(1) Chatbot		(2) MPV		(3) Control		t-statistic		
Variable	**N**	**Prop**	**N**	**Prop**	**N**	**Mean**	**(1)-(2)**	**(2)-(3)**	**(1)-(3)**
Stayed	552	0.86385	547	0.866878	502	0.838063	0.006	1.8099	1.4273
Attrited	87	0.13615	84	0.133122	97	0.161937			

We conducted chi-square tests comparing proportions attrited across arms.

significant baseline differences across study arms, confirming successful randomisation. End-line attrition was 13.6% in the chatbot arm, 13.3% in the video arm and 16.2% in the control; χ^2 tests show these differences are insignificant at the 5% or 10% levels (Table 1B).

Intervention monitoring

Teacher engagement was tracked in real time: a dashboard recorded every chatbot interaction, while trained moderators logged each micro-practice-video (MPV) response on an engagement sheet. Among the 552 teachers offered the chatbot, 88.7% triggered the flow at least once, and almost 70% completed it at least once. Sustained use was rarer, with only 16% completing a flow every month, rising to 22% after the first month. Engagement was skewed. While the median teacher accessed the chatbot on just 7 days, the top 20% used it on 51 days, indicating that only a small fraction showed high engagement.

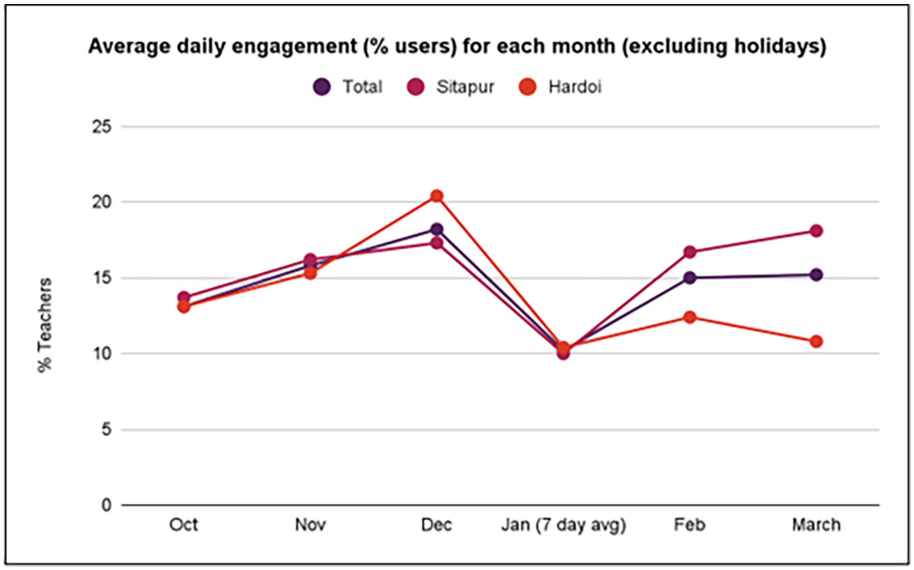

Figure 4. Average daily user engagement (%) per month (excluding holidays) on Chatbot.

Monthly unique-user rates oscillated between 10% (in January due to school holidays) and 18% (November–December, when streak stickers were introduced and working days were longest); Figure 4 plots this arc and shows Sitapur consistently out-performing Hardoi—for example, in March 18.1% of Sitapur teachers opened the bot compared with 10.8% in Hardoi, lifting the overall daily rate to 15.2%. Nudges such as report cards produced sharp but brief spikes. For example, there was a significant increase in engagement the day after the first report card, with a 117% rise from the day before.

Because video views cannot be verified on WhatsApp, MPV engagement was inferred from in-group activity. Forty-eight per cent of teachers answered at least one of the 12 MPV-related polls, 20% reacted to a video with emojis, 17% posted comments or questions, and only a handful shared the optional demonstration clips of themselves, an effort-heavy task (Figure 5A). Figure 5B shows the usual Sitapur edge, with poll participation being 55% there versus 39% in Hardoi.

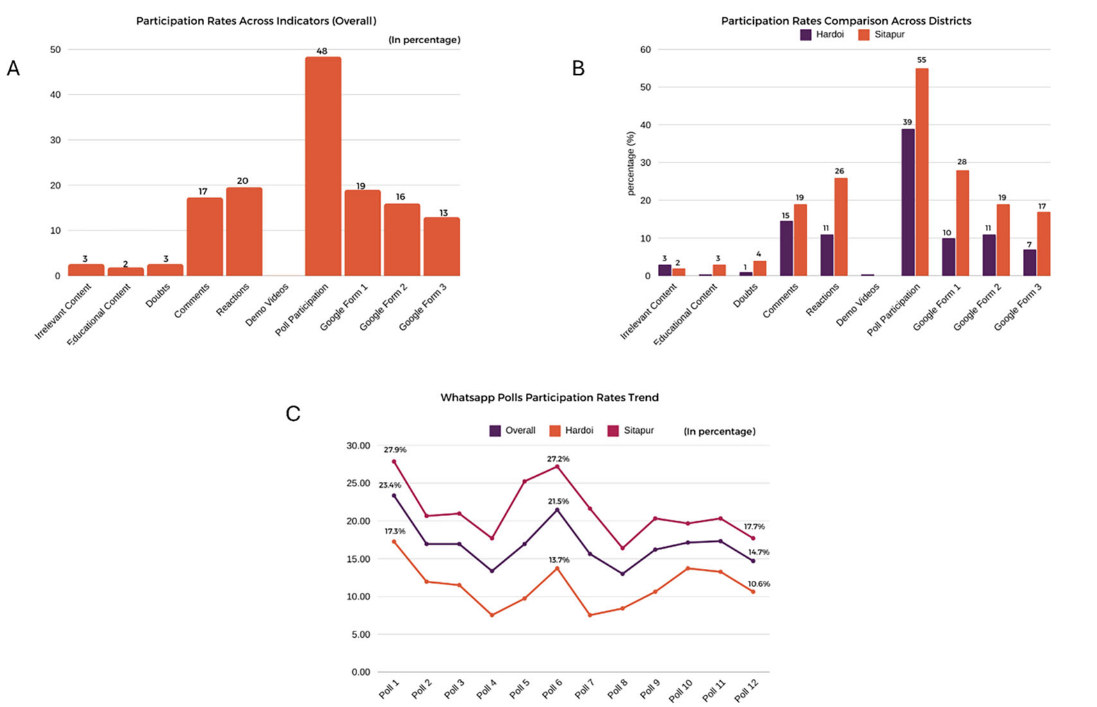

Figure 5. User engagement (%) on micro-practice videos whatsapp group, i.e. percentage of teachers responding in whatsApp groups. A. Percentage of Teachers Responding in WhatsApp Groups by message category. B. Same as in A, split across the two districts. C. Percentage of Teachers Responding to each of the 12 WhatsApp Polls.

Social-reward dynamics are evident in Figure 5C, as engagement peaked on the poll following the sixth video, likely driven by increased interest after a certificate winner's testimonial was shared two days earlier.

Primary outcomes

As mentioned above, primary outcomes measured the effects of intervention on the *adoption* (Table 2), *knowledge* (Table 3) and *valuation* of teacher guides (Table 4) and/or effective pedagogical practices (Table 5). Adoption was measured using vignette scenarios, which presented realistic situations to assess how teachers might apply specific techniques and practices. These vignettes allowed teachers to demonstrate their understanding and behaviour more nuancedly than direct survey questions. These questions were linked to using TGs and effective practices shown in the videos and covered both numeracy and literacy practices. Each vignette option was carefully designed to avoid being obviously incorrect or socially undesirable to capture actual behaviour. For example, one vignette reads: *Sanjay is a primary school teacher with 17 students. He reads new stories aloud, explains unfamiliar words, and adds elements. Despite this, many students still struggle with speaking clearly. What should Sanjay do to improve their speaking skills? (i) Ask open-ended questions, (ii) Write new words on the blackboard, explain their meanings, and ask students to repeat them, (iii) Play videos with engaging stories, (iv) None of the above.* While none of the options appear overtly incorrect, based on evidence of effectiveness and the intended target of the intervention, the correct response is to ask *open-ended questions.*

Adoption, measured by six numeracy- and literacy-linked vignettes, did not improve. Endline accuracy means were 0.452 (0.2060) in the chatbot arm, 0.455 (0.2107) in MPV and 0.464 (0.2068) in control; no treatment or subgroup coefficient was significant. Knowledge, captured by four aligned quiz items, likewise showed no gains. Mean scores were 0.451

Table 2. Effect on adoption (of TG & effective practices) measured through vignette accuracy.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
T1: Chatbot	-0.0183	-0.0106	-0.029	-0.0119	-0.0343	-0.0143
	[0.0116]	[0.0152]	[0.0190]	[0.0144]	[0.0211]	[0.0137]
T2: MPV	-0.0122	0.0006	-0.029	-0.00618	-0.0268	-0.0129
	[0.0110]	[0.0133]	[0.0179]	[0.0133]	[0.0222]	[0.0134]
Control (raw mean)	0.464	0.474	0.44	0.497	0.393	0.468
N	1584	915	669	1103	481	1149

Notes: Robust standard errors are in brackets.
Treatment effects are reported relative to the control group. The outcome variable is a discrete proportion ranging from 0 to 1. Models include controls for demographic variables, district fixed effects, and baseline conditions with clustered standard errors.

Table 3. Effect on knowledge of TG and effective practices.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
T1: Chatbot	0.00894	0.00872	0.00413	0.00622	0.0111	0.0146
	[0.0119]	[0.0157]	[0.0179]	[0.0142]	[0.0207]	[0.0151]
T2: MPV	-0.00903	0.0107	-0.0357^{+}	-0.0206	0.0186	-0.0113
	[0.0126]	[0.0167]	[0.0183]	[0.0147]	[0.0213]	[0.0147]
Control (raw mean)	0.443	0.446	0.44	0.477	0.37	0.447
N	1584	915	669	1103	481	1149

Notes: Robust standard errors are in brackets.
Significance level:
$^{+}p < 0.10$.
Treatment effects are reported relative to the control group. The outcome variable is discrete, ranging from 0 to 4. Models include controls for demographic variables, district fixed-effects, and baseline conditions, with clustered standard errors.

Table 4. Effect on TG valuation/price.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
T1: Chatbot	8.986	**18.26**+	-6.34	10.45	6.491	13.77
	[6.934]	[10.23]	[7.227]	[9.801]	[5.256]	[8.553]
T2: MPV	9.117	17.44	-5.725	**15.69**+	-4.825	9.956
	[7.195]	[10.60]	[7.082]	[9.048]	[8.932]	[8.928]
Control (raw mean)	482.62	476.62	491.34	481.47	485	482.25
N	926	544	382	629	297	680

Notes: Robust standard errors are in brackets.
Significance level:
+p < 0.10.
Treatment effects are reported relative to the control group. The outcome variable is continuous. Models include controls for demographic variables, district fixed effects, and baseline conditions, with clustered standard errors.

Table 5. Effect on valuation/price of attending training on effective teaching methods.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
T1: Chatbot	8.579	**55.27**+	-20.79	7.814	-9.866	-16.28
	[17.43]	[30.29]	[20.46]	[13.85]	[77.03]	[16.04]
T2: MPV	6.111	30.11	-13.39	1.46	13.03	-27.13
	[22.33]	[37.30]	[17.51]	[20.77]	[56.28]	[17.66]
Control (raw mean)	475.55	459.75	495.16	491.67	463.16	477.14
N	158	89	69	129	29	117

Notes: Robust standard errors are in brackets.
Significance level:
+p < 0.10.
Treatment effects are reported relative to the control group. The outcome variable is continuous. Models include controls for demographic variables, district fixed-effects, and baseline conditions with clustered standard errors.

(0.2036), 0.435 (0.2011) and 0.443 (0.2042) for chatbot, MPV and control. Only one effect emerged: in Hardoi, the MPV arm fell by –0.0357 (β = -0.0357, SE 0.0183, $p < .10$).

To measure whether micro-practice videos/chatbot's TG-based lessons led to any changes in the perceived value of receiving training on effective practices or using teacher guide, a modified "willingness to pay/accept" contingent valuation method (Carson & Hanemann 2005) was used - a game pitting cash (₹500→₹1), a teacher-guide (TG) and a training session on effective practices against one another. To simulate realistic choices, the decision was tied to a lottery, where lottery winners would receive their selected option: cash in the form of a phone recharge, training, or TG.

At end-line 73% chose the TG, 26% training and 1% neither. TG was valued at an average of ₹489.8 (57.17) in chatbot, ₹487.5 (68.10) in MPV and ₹482.6 (84.23) in control; training values were ₹480.1 (92.15), ₹480.6 (93.30) and ₹475.5 (104.80). Regressions show no overall effects, yet three 10%-level gains appeared from subgroup analyses. In Sitapur, the chatbot arm increased TG valuation by ₹18.3 (SE 10.2), while the MPV arm led to a ₹55.3 (SE 30.3) increase. Among Assistant Teachers, the MPV arm resulted in a ₹15.7 (SE 9.1) increase in TG valuation.

Secondary Outcomes

Overconfidence bias, the gap between self-rated certainty about the accuracy and vignette accuracy, remained high, with 91–92% displaying overconfidence. End-line means of bias scores for chatbot, MPV and control group teachers were 0.36 (0.24), 0.35 (0.24) and 0.37 (0.24). The only significant shift was a –0.0286 reduction (SE 0.017, $p < .10$) for Assistant Teachers in MPV (Table 6).

Table 6. Effect on confidence bias.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
T1: Chatbot	-0.00992	-0.0132	-0.00985	-0.0168	0.00548	-0.00618
	[0.0126]	[0.0163]	[0.0213]	[0.0160]	[0.0230]	[0.0154]
T2: MPV	-0.0222	-0.0275	-0.0174	-0.0286^{+}	-0.0073	-0.019
	[0.0139]	[0.0191]	[0.0204]	[0.0167]	[0.0252]	[0.0163]
Control (raw mean)	0.374	0.355	0.398	0.36	0.402	0.369
N	1584	915	669	1103	481	1149

Notes: Robust standard errors are in brackets.
Significance level:
$^{+}$p < 0.10.
Treatment effects are reported relative to the control group. The outcome variable is continuous (-1 to 1). Models include controls for demographic variables, district fixed-effects, and baseline conditions with clustered standard errors.

Motivation (to do their job) was measured through teachers' agreement to 14 statements related to intrinsic and extrinsic motivational factors (IDinsight 2015). These statements included examples such as, "I do this job because I care about benefiting others through my work", and "My supervisor praises me for my efforts in the school. Motivation, measured by the improvement from baseline in an index constructed from these 14 items, improved for 43.0% of chatbot teachers, 40.6% in MPV and 38.3% in control. Overall, no statistically significant effects were observed across the treatment arms. However, among Shiksha Mitra contractual teachers, odds of improvement rose 1.52 times (SE 0.254, p < .10) in chatbot and 1.68 times (SE 0.521, p < .10) in MPV (Table 7).

Beliefs about student-centred pedagogy shifted marginally (Table 8). Chatbot teachers were 1.34× more likely (SE 0.164, p < .10) to think parents favoured new methods. Assistant Teachers in MPV were 31.6% less likely to feel the same (SE 0.202, p < .10). For Shiksha Mitra, the chatbot raised the odds of believing new methods boost learning by 1.73× (SE 0.304, p < .10).

Intentions and self-efficacy, measured through self-reported ability to contribute to student outcomes (Table 9) and teachers' intention to use teacher guides and adopt new practices, mostly declined in MPV (Table 10). Relative to control, MPV teachers were less likely to upgrade self-assessed ability ($\beta = -0.555$, $p < .01$) and intention to use the TG ($\beta = -0.4768$, $p < .01$).

Table 7. Effect on motivation.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
T1: Chatbot	0.166	0.179	0.0831	0.06	**0.417**$^{+}$	0.115
	[0.132]	[0.182]	[0.190]	[0.175]	[0.254]	[0.159]
T2: MPV	0.0675	0.0893	0.0381	-0.113	**0.521**$^{+}$	-0.053
	[0.129]	[0.179	[0.183]	[0.160]	[0.275]	[0.154]
Control (raw mean)	0.3828	0.365	0.4107	0.3972	0.3542	0.398
N	1365	827	538	945	420	982

Notes: Robust standard errors are in brackets.
Significance level:
$^{+}$p < 0.10.
Treatment effects are reported relative to the control group. The outcome variable is binary. It takes the value of 1 if motivation increased from baseline and 0 otherwise. Models include controls for demographic variables and district fixed-effects with clustered standard errors.

Table 8. **Effect on teacher beliefs on effective teaching practices.**

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher's own preference for new methods						
T1: Chatbot	0.172	-0.0215	0.447	0.19	0.0761	0.3
	[0.216]	[0.278]	[0.330]	[0.331]	[0.326]	[0.258]
T2: MPV	0.0422	0.137	-0.0985	-0.317	0.552	0.0175
	[0.209]	[0.264]	[0.344]	[0.256]	[0.381]	[0.272]
Control (raw mean)	0.8983	0.894	0.8846	0.9311	0.8025	0.902
N	1489	867	622	1043	446	1089
Teacher's perception of other teachers' preference						
T1: Chatbot	0.157	0.146	0.144	0.262	-0.0815	0.154
	[0.188]	[0.249]	[0.308]	[0.238]	[0.314]	[0.218]
T2: MPV	-0.0592	-0.104	0.0559	-0.138	0.152	-0.222
	[0.194]	[0.237]	[0.337]	[0.250]	[0.341]	[0.225]
Control (raw mean)	0.8166	0.8195	0.8125	0.8259	0.7974	0.836
N	1309	772	537	906	403	950
Teacher's preference for new methods to improve student outcomes						
T1: Chatbot	0.287	0.149	0.454	0.0609	**0.548**$^{+}$	0.246
	[0.198]	[0.242]	[0.358]	[0.300]	[0.304]	[0.255]
T2: MPV	0.0878	0.148	0.019	-0.0709	0.303	0.03
	[0.210]	[0.269]	[0.336]	[0.304]	[0.302]	[0.280]
Control (raw mean)	0.8593	0.8655	0.8505	0.9079	0.7597	0.881
N	1369	806	563	946	423	996
Teacher's perceptions of parental acceptance of new methods						
T1: Chatbot	**0.292**$^{+}$	0.264	0.391	0.217	0.306	**0.408**$^{+}$
	[0.164]	[0.208]	[0.268]	[0.228]	[0.246]	[0.217]
T2: MPV	-0.218	-0.216	-0.236	**-0.380**$^{+}$	0.0599	-0.165
	[0.153]	[0.195]	[0.250]	[0.202]	[0.279]	[0.191]
Control (raw mean)	0.7541	0.7398	0.7735	0.777	0.7114	0.764
N	1186	685	501	800	386	865

Notes: Robust standard errors are in brackets.
Significance level:
$^{+}p < 0.10$.
Treatment effects are reported relative to the control group. The outcome variables are binary (0 or 1). They take the value 1 if the teacher answered in favour of new methods and 0 if the teacher answered in favour of traditional methods. Models include controls for demographic variables, district fixed effects, and baseline conditions with clustered standard errors.

Subgroup analysis showed that Assistant Teachers in the MPV were less likely to improve their self-assessed ability ($\beta = -0.5$, $p < 0.01$) and intention to use the TG ($\beta = -0.385$, $p < 0.05$). Among Shiksha Mitras, those in both the chatbot ($\beta = -0.397$, $p < 0.1$) and MPV ($\beta = -0.49$, $p < 0.05$) arms were less likely to improve ability assessments, with additional negative effects in the MPV arm on their intentions to use the TG ($\beta = -0.663$, $p < 0.05$) and implement effective practices ($\beta = -0.504$, $p < 0.1$).

These declines were driven by Hardoi, where self-assessed ability fell in the chatbot arm ($\beta = -0.571$, $p < .01$) and the MPV arm ($\beta = -0.910$, $p < .001$), intention to use TG declined in the chatbot ($\beta = -0.567$, $p < .01$) and MPV ($\beta = -0.707$, $p < .01$) arms, and intention to implement practices dropped in the MPV arm ($\beta = -0.566$, $p < .05$). None of these negative effects were observed in Sitapur.

Table 9. Effect on self-assessment of ability to contribute to learning goals.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
T1: Chatbot	-0.0571	0.0473	**-0.571****	-0.112	**-0.397**+	**-0.299**+
	[0.174]	[0.178]	[0.216]	[0.172]	[0.238]	[0.165]
T2: MPV	**-0.555****	-0.26	**-0.910*****	**-0.500****	**-0.490***	**-0.607*****
	[0.186]	[0.187]	[0.209]	[0.183]	[0.230]	[0.179]
Control (raw mean)	0.3028	0.2724	0.3443	0.269	0.375	0.3028
N	1584	915	669	1103	481	1149

Notes: Robust standard errors are in brackets.
Significance levels:
+p < 0.10,
*p < 0.05,
**p < 0.01,
***p < 0.001.
Treatment effects are reported relative to the control group. The outcome variable is binary (0 or 1). It takes the value 1 if ability rating is improved from baseline and 0 otherwise. Models include controls for demographic variables, district fixed effects with clustered standard errors.

Table 10. Effect on self-assessment of intention to adopt TG/Effective teaching methods.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
Increase in intention to use TG						
T1: Chatbot	-0.223	0.0867	**-0.567****	-0.172	-0.314	**-0.285**+
	[0.140]	[0.186]	[0.194]	[0.169]	[0.243]	[0.165]
T2: MPV	**-0.468****	-0.24	**-0.707****	**-0.385***	**-0.663***	**-0.482****
	[0.146]	[0.186]	[0.224]	[0.165]	[0.274]	[0.169]
Control (raw mean)	0.295	0.2483	0.3585	0.2865	0.3125	0.29
N	1584	915	669	1103	481	1149
Increase in intention to implement effective practices						
T1: Chatbot	-0.0744	0.28	-0.403	0.131	-0.328	-0.00137
	[0.177]	[0.256]	[0.245]	[0.209]	[0.282]	[0.192]
T2: MPV	-0.181	0.176	**-0.566***	0.0457	**-0.504**+	-0.177
	[0.177]	[0.236]	[0.276]	[0.209]	[0.277]	[0.201]
Control (raw mean)	0.2251	0.1655	0.3067	0.1667	0.35	0.2194
N	1584	915	669	1103	481	1149

Notes: Robust standard errors are in brackets.
Significance levels:
+p < 0.10,
*p < 0.05,
**p < 0.01,
***p < 0.001.
Treatment effects are reported relative to the control group. The outcome variable is binary (0 or 1). It takes value 1 if the intention rating is improved from the baseline and 0 otherwise. Models include controls for demographic variables and district-fixed effects with clustered standard errors.

Both self-reported FLN attainment in previous academic years and self-reported predictions of FLN attainment in the current year rose with teachers claiming more than 70% of pupils met learning goals, jumping from 14% at baseline to 49% (chatbot) and 43% (MPV & control) at endline. A similar trend was seen in FLN outcome predictions, with 65% (chatbot), 64% (MPV), and 61% (control) of teachers expecting over 70% students to achieve FLN goals in their current academic year, up from 38% predicting the same at baseline. No overall treatment effects were found, but among

Table 11. Effect on self-reported FLN outcome achieved & outcomes predicted to achieve.

	Full sample	Sitapur	Hardoi	Assistant teacher	Shiksha Mitra	Sample that passed attention check
	(1)	(2)	(3)	(4)	(5)	(6)
Self-reported FLN outcomes achieved						
T1: Chatbot	0.21	0.187	0.265	**0.296**$^{+}$	0.0963	**0.263**$^{+}$
	[0.144]	[0.187]	[0.234]	[0.180]	[0.210]	[0.159]
T2: MPV	-0.0178	0.0512	-0.136	-0.0453	0.043	-0.0428
	[0.131]	[0.170]	[0.207]	[0.168]	[0.202]	[0.148]
Control (Median)	3	3	4	3	3	3
N	1485	871	614	1037	448	1083
Self-reported FLN outcomes predicted to be achieved in the next year						
T1: Chatbot	0.187	0.206	0.171	0.221	0.0988	0.271
	[0.162]	[0.211]	[0.262]	[0.207]	[0.223]	[0.182]
T2: MPV	0.0823	0.0949	0.0932	-0.0201	0.163	0.077
	[0.158]	[0.198]	[0.263]	[0.203]	[0.237]	[0.172]
Control (Median)	4	4	4	4	4	4
N	1409	801	608	975	434	1028

Notes: Robust standard errors are in brackets.
Significance level:
$^{+}p < 0.10$.
Treatment effects are reported relative to the control group. The outcome variable is ordinal (1-4). Models include controls for demographic variables, baseline conditions and district fixed-effects with clustered standard errors.

subgroups, Assistant Teachers in the chatbot arm were likelier to report more students achieving higher learning outcomes (β = 0.296, $p < 0.10$; Table 11).

Instrumental variable analysis

Minimal engagement was partial with59.4% of chatbot teachers accessing a lesson at least twice, and 57.2% of MPV teachers reacting in their WhatsApp group at least once. Hence, we ran an instrumental variable analysis that used random assignment as the instrument for these uptake thresholds (Duflo et al., 2007). In the case of the MPV WhatsApp group, since it was difficult to confirm if teachers watched the videos on WhatsApp, we used group responsiveness (e.g., comments, poll participation) as a proxy for engagement. IV estimates were indistinguishable from the intent-to-treat results and are therefore not shown.

Factor analysis

Factor analysis yielded 11 factors. Bivariate correlations showed that Factor 1, "Knowledgeable Junior assistant teachers" (high for assistant teachers, low for Shiksha Mitra), and more junior teachers had the strongest positive link with the outcome (r = 0.34). Six of the eleven latent factors were associated with higher or lower outcomes (significant effects in the regression). The regression also included the treatment variable separately to look for treatment heterogeneity, which was not significant. On the positive factors, F 01 "Knowledgeable Junior assistant teachers" with heavier loadings on Assistant teachers, less experience, and greater knowledge and vignette score at baseline, has the most considerable effect (β = 0.60, p < 0.001). F 04 "Endline Intent & Self-efficacy" (β = 0.20, p < 0.001) is strongest in teachers with higher Endline measures of intent to use TG and effective classroom practices, and higher self-reported ability. F 09 "Baseline Intent & Self-efficacy" (β = 0.06, p < 0.01) combines teachers' *baseline intent* to adopt the effective classroom practices and TG, and higher self-reported ability, showing that early commitment backed by self-efficacy still matters. Conversely, the other three factors were negative predictors: F 03 "Belief in traditional teaching" (β = -0.16, p < 0.01) loads heavily on *beliefs*, teachers who choose traditional over student-centric approaches to teaching at Endline on the vignette question from their own, other teachers, parents, and learning outcome perspectives underperform. F 10 "Gender" (β = -0.18, p < 0.01) is dominated by gender, indicating that—once latent traits are controlled—being a male teacher slightly depresses scores. Finally, F 11 "School effects" (β = -0.37, p < 0.001) lacks any loading above 0.3, but still captures a residual school fixed effects disadvantage, a low number of teachers in the school at baseline, and low baseline motivation scores.

Self-reported use and perceptions of interventions

As part of our endline survey, we also asked participants in the treatment arms about their views and perceptions regarding the use and effectiveness of the interventions. Teachers with low or limited engagement with chatbot (n = 277) cited the following reasons for low adoption: poor internet connectivity (54%), preference for teaching guides (34%), and having too many existing resources (19%). Among those with low or high engagement (n = 489), 86% found the chatbot easy or very easy to use. Most of these teachers (60-75%) considered the intervention successful in improving classroom instruction, aiding classroom preparation, helping them learn new things, and improving student outcomes. Furthermore, 64% of these teachers stated they would 'definitely recommend' the chatbot to others. However, only 41% agreed to share the contact information of another teacher with whom the chatbot could be shared, indicating that the willingness to share contact information was lower than the intention to recommend.

In the MPV arm (n = 520), 97% of teachers reported watching the MPV videos, and 96% of those who reported watching indicated that they had implemented the teaching practices demonstrated in the videos. While a high percentage (96%) found the teaching practices easy or very easy to understand, a slightly lower rate (86%) found it easy or extremely easy to implement them, suggesting a gap between comprehension and practical application. Approximately 87% of teachers found the videos successful or very successful in improving classroom instruction, assisting with classroom preparation, engagingly presenting teaching methods, and learning new things. A slightly smaller percentage (78%) found the videos successful or very successful in improving student outcomes. When asked if they would recommend the videos to other teachers, 80% said they would, but only 45% agreed to share the contact information of another teacher with whom the videos could be shared.

Discussion

We tested two light, WhatsApp-based interventions—a bite-sized lesson-plan chatbot and fortnightly micro-practice videos—in a cluster-RCT across rural Uttar Pradesh. Both wrapped existing structured-pedagogy resources in reminders, motivational messages and recognition-style rewards.

Neither tool shifted the three pre-registered primary outcomes. End-line accuracy on vignette-based adoption scenarios averaged 0.452 (SD 0.2060) in the chatbot arm, 0.455 (0.2107) in MPV and 0.464 (0.2068) in control—well below a 50% success rate. Knowledge scores were similarly low: 0.451 (0.2036), 0.435 (0.2011) and 0.435 (0.2042). Limited classroom observations (≈30 per arm) data confirmed that most target teaching practices were still absent. Broadly, motivation and beliefs did not improve. One positive finding was that teachers in the chatbot group were 1.34 times more likely ($p < .10$) to believe that parents would support new teaching methods. Conversely, MPV teachers lowered their self-rated ability to meet FLN goals ($\beta = -0.555$, $p < .01$) and their intention to use the teacher guide ($\beta = -0.4768$, $p < .01$).

Teachers did, however, value printed teacher guides. At endline 73% chose the guide over training or cash, 26% chose training, and just 1% chose neither. Mean valuations were ₹486.8 for the guide and ₹478.8 for training, yet neither intervention raised these figures overall.

Context matters: Hardoi versus Sitapur

Subgroup analysis pointed to Hardoi as the primary driver of negative effects, with teachers in both arms lowering self-assessment of ability to contribute to FLN outcomes and intentions to adopt practices. In contrast, these effects were absent in Sitapur, where, on the contrary, the chatbot arm slightly improved valuation of both TG and receiving 30-minute training on effective practices. These differences likely reflect contextual factors. Hardoi saw baseline resistance and lower buy-in, possibly due to stronger teacher unions. Monitoring data supports this explanation, with Sitapur logging higher engagement almost every month. Prior research shows regional and institutional factors such as union strength, governance structures, and delivery systems can significantly shape program outcomes (Bruns et al., 2011). For example, a study in Delhi and UP found teacher networks improved motivation only in Delhi but had no significant effects in UP, underscoring that effectiveness is highly context-dependent (STiR & IDinsight, 2018).

Teacher type matters, too

Assistant Teachers and Shiksha Mitra responded differently. In ATs the MPV reduced over-confidence (–0.0286, SE 0.017, $p < .10$) and lifted TG valuation by ₹15.7 (9.1, $p < .10$); the chatbot nudged up reported FLN attainment ($\beta = 0.296$, $p < .10$). SMs, typically less qualified and lower paid, gained motivation in both arms (odds +1.52 in chatbot, +1.68 in MPV, $p < .10$) and, in the chatbot, became 1.73 × more likely to believe new methods raise learning ($p < .10$).

These findings indicate that the interventions affected distinct teacher types differently. Better alignment of confidence with knowledge was seen in our previous study on improving parental engagement in Uttar Pradesh, and these gains were observed in the arm that was better equipped and resourced, as in the current study (Shashidhara et al., 2025). The modest improvements in motivation and beliefs among Shiksha Mitras highlight the importance of investing in tailored capacitybuilding strategies for diverse teacher profiles, rather than focusing solely on increasing teacher numbers. The heterogeneity aligns with findings from other studies. For instance, Barrera-Osorio et al. (2022) found that non-monetary rewards improved motivation among female but not male teachers, while Ajzenman et al. (2024) reported that intrinsic motivators and identity-based approaches were more effective in increasing the willingness of high-performing teachers (compared to low-performing ones) to work in disadvantaged schools.

Factor analysis reinforces this heterogeneity. “Knowledgeable junior ATs” and teachers with high intent and self-efficacy at baseline or end-line were strong positive predictors. Traditional-teaching beliefs, male gender and poorly staffed schools with low baseline motivation predicted lower scores.

Engagement: the 20% problem

Monitoring data revealed that only a small segment of teachers (15–20%) were actively engaged with the interventions, while the majority exhibited low engagement. This limited uptake may explain the overall lack of impact. The interventions were designed to be minimal additions to existing systems to avoid overburdening teachers. Still, this light-touch approach may have been insufficient to reach the broader base, particularly the ‘swing teachers’ who are neither highly motivated nor fully disengaged (Piper et al., n.d.). Instead, the interventions primarily resonated with teachers at the extremes i.e. those already highly motivated and open to innovation. This pattern is consistent with broader evidence: many voluntary education programs targeting teachers or school leaders, regardless of their format or content, tend to engage only about 20% of the target group (Romero et al., 2022; STiR & IDinsight, 2018; UNESCO, 2019). In Nepal, Nigeria, and Uganda, 14–15% of schools drove nearly 80% of early-grade learning gains, often led by teachers with empathy, strong communication, and the ability to navigate social and systemic barriers (Gove & King, 2023). The key challenge, therefore, lies in developing approaches that engage the “average” teacher, not just the early adopters.

Non-pecuniary rewards show promise

Monitoring data showed that teachers responded well to recognition and social proof. Engagement rose from 20% to 100% after reward-based messages, like report cards highlighting top performers, though effects were short-lived. This aligns with findings from northern India, where 60% of teachers said peer recognition boosts motivation (Hamm-Rodríguez et al., 2018). While less effective than in-kind incentives, non-monetary rewards can still matter, especially for intrinsically motivated teachers (Barrera-Osorio et al., 2022). Given the mixed evidence on and known limitations of monetary incentives (Barrera-Osorio & Raju, 2017), more research is needed on integrating non-monetary rewards into teacher development models in contexts where the implementation of financial incentives poses a challenge.

Why teachers tuned out, what they liked, and wanted more of

A mid-line survey (≈200 teachers) highlighted time pressure, resource overload and poor internet as key barriers to engagement. Low-engagers in the chatbot arm cited these obstacles at rates of 54%, 34% and 19%, respectively. Moreover, third-party delivery (rather than official channels) may have dampened credibility.

Nevertheless, teachers who engaged with the interventions generally reported positive feedback. The micro-practice videos (MPV) were valued for their engaging format and ease of understanding. However, some teachers found the videos idealistic and hard to apply in their contexts, suggesting more content for multi-grade and large-class settings. The chatbot was valued for its quick, user-friendly access to lesson summaries, with suggestions to add more TG content, activity videos, downloadable lesson plans, and a query helpline.

Implications for design

Light-touch nudges alone rarely move the middle. Evidence from different countries shows that sustained, in-person training and coaching, and peer networks yield accountability and hence larger gains than virtual or cascade models (Romero et al., 2022; Wolf et al., 2019; Cilliers et al., 2022). For example, an in-person coaching program in Peru improved teaching practices by 0.20 SD and boosted student learning (Castro et al., 2025).

A blended model combining online and offline elements may increase student learning gains (de Hoop et al., 2025). For example, Pakistan’s eLearn program paired video lessons with in-person training to help teachers use them effectively (Beg et al., 2022). Designed to support rather than replace instruction, this approach can benefit less experienced teachers like Shiksha Mitras.

A more comprehensive intervention includes behavioural strategies that simplify information, aligning with teachers’ beliefs and working conditions. To avoid superficial compliance or mechanical adoption of new pedagogies, it is crucial to address the socio-cultural dimensions of pedagogy, including extant belief systems, professional norms, and classroom contexts, while helping them understand the rationale behind new methods (Hoadley, 2024; Kaur, 2023; Miyazaki, 2015). More research is needed on maintaining these attitudinal shifts and converting them into sustained practices.

Conclusion

Our findings yield three takeaways. First, effects varied by context and teacher type. One district's low initial buy-in reduced engagement, while the other saw modest gains, such as higher valuation of teaching aids. Contractual teachers, who are typically less qualified and lower paid, showed slight but encouraging boosts in motivation, highlighting the value of targeted support. Second, only a self-selected group of highly motivated teachers engaged, while the average teacher largely tuned out. Third, daily information overload easily crowds out light-touch formats. These tools are likely useful only in systems or among teachers who already meet a baseline level of preparation. For example, micro practice videos work better as refreshers after solid training, and chatbot lesson plans help when teachers already understand the core content. Scaling success will require pairing low-cost nudges with trusted communication channels, sustained coaching, and meaningful system-level support, while carefully weighing resource demands against long-term improvements in teaching and student learning.

Acknowledgements

This study was conducted by authors from the Centre for Social and Behaviour Change (CSBC), Ashoka University, in collaboration with the Central Square Foundation (CSF). It was generously funded (INV-053842) by the Bill & Melinda Gates Foundation (BMGF).

References

Ajzenman N, Bertoni E, Elacqua G, *et al.*: **Altruism or money? Reducing teacher sorting using behavioral strategies in Peru.** *J. Labor Econ.* 2024; **42**(4): 1049–1091.
Publisher Full Text

ASER Centre: *Annual Status of Education Report (Rural) 2022.* New Delhi, India: ASER Centre; 2022.
Reference Source

Barrera-Osorio F, Raju D: **Teacher performance pay: Experimental evidence from Pakistan.** *J. Public Econ.* 2017; **148**: 75–91.
Publisher Full Text

Barrera-Osorio F, Cilliers J, Cloutier MH, *et al.*: **Heterogenous teacher effects of two incentive schemes: Evidence from a low-income country.** *J. Dev. Econ.* 2022; **156**: 102820.
Publisher Full Text

Bau N, Das J: **Teacher value added in a low-income country.** *Am. Econ. J. Econ. Pol.* 2020; **12**(1): 62–96.
Publisher Full Text

Beg S, Halim W, Lucas AM, *et al.*: **Engaging teachers with technology increased achievement, bypassing teachers did not.** *Am. Econ. J. Econ. Pol.* 2022; **14**(2): 61–90.
Publisher Full Text

Belafi C, Hwa Y-Y, Kaffenberger M: *Building on solid foundations: Prioritising universal, early, conceptual, and procedural mastery of foundational skills.* Oxford, UK: 2020.
Publisher Full Text

Benmansour N: **Job satisfaction, stress, and coping strategies among Moroccan high school teachers.** *Mediterranean Journal of Educational Studies.* 1998; **3**: 13–33.

Bhatty K, Saraf R: *Does government's monitoring of schools work?* Centre for Policy Research (CPR); 2016.
Reference Source

Bold T, Filmer D, Martin G, *et al.*: **Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa.** *J. Econ. Perspect.* 2017; **31**(4): 185–204.
Publisher Full Text

Bruns B, Filmer D, Patrinos HA: *Making schools work: New evidence on accountability reforms. Human Development Perspectives.* World Bank; 2011.
Reference Source

Carson RT, Hanemann WM: **Contingent valuation.** Mäler K-G, Vincent JR, editors. *Handbook of environmental economics.* Elsevier; 2005; Vol. **2**: pp. 821–936.
Publisher Full Text

Castro JF, Glewwe P, Heredia-Mayo A, *et al.*: **Can teaching be taught? Improving teachers' pedagogical skills at scale in rural Peru.** *Quant. Econ.* 2025; **16**(1): 185–233.
Publisher Full Text

Central Square Foundation: *Systemic drivers of foundational learning outcomes.* Central Square Foundation; 2021.
Reference Source

CETE: **Status of teachers in the workforce in eight states: A report based on SOTTTER 23 Survey. Background research report 5: State of Teachers, Teaching Education Report 2023.** 2023.

Chakera S, Haffner D, Harrop E: *UNICEF Eastern and Southern Africa Region Working Paper – Structured Pedagogy: For real-time equitable improvements in learning outcomes.* UNICEF; 2020.

Chetty R, Friedman J, Rockoff J: **Measuring the impact of teachers II: Teacher value-added and student outcomes in adulthood.** *Am. Econ. Rev.* 2014; **104**(9): 2633–2679.
Publisher Full Text

Cilliers J, Fleisch B, Kotze J, *et al.*: **Can virtual replace in-person coaching? Experimental evidence on teacher professional development and student learning.** *J. Dev. Econ.* 2022; **155**: 102815.
Publisher Full Text

CSBC: *Improve Teacher Uptake of Effective FLN Instruction Diagnostic Report.* Centre for Social and Behaviour Change; 2023.
Reference Source

Darling-Hammond L: *Powerful Teacher Education: Lessons from Exemplary Programs.* Jossey-Bass; 2006.

Duflo E, Glennerster R, Kremer M: **Using randomization in development economics research: A toolkit.** *Handb. Dev. Econ.* 2007; **4**: 3895–3962.
Publisher Full Text

Dunham J, Varma V: *Stress in teachers: Past, present, and future.* London: Whurr; 1998.

Filmer D, Nahata V, Sabarwal S: *Preparation, practice, and beliefs: A machine learning approach to understanding teacher effectiveness.* Development Economics, Development Research Group & Education Global Practice, World Bank; 2021. (Policy Research Working Paper No. 9847).

Flaschen K, Amjad Z, Better C, *et al.*: **Improving teacher uptake of pedagogical best practices for foundational literacy and numeracy: Key behavioral barriers and tips to address them. ideas42.** 2024.
Reference Source

Glewwe P, Muralidharan K: **Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications.** Hanushek EA, Machin S, Woessmann L, editors. *Handbook of the Economics of Education.* North-Holland; 2016; Vol. **5**: pp. 653–743.

Gove A, King S: **Replicating Success: How Behavioral Science Can Help Education Systems Recover from Learning Loss, (blog).** 2023.
Reference Source

Gray-Lobe G, Keats A, Kremer M, *et al.*: **Can education be standardized? Evidence from Kenya. Evidence from Kenya (September 16, 2022).** *University of Chicago, Becker Friedman Institute for Economics Working Paper, (2022-68).* 2022.

Hahn M, Lawson R, Lee YG: **The effects of time pressure and information load on decision quality.** *Psychol. Mark.* 1992; **9**(5): 365–378.
Publisher Full Text

Hamm-Rodríguez M, Richardson E, Guajardo J: **Teacher motivation and behavior change: Results of the teacher motivation diagnostic tool in Northern India.** Pouezevara S, editor. *Cultivating dynamic educators: Case*

studies in teacher behavior change in Africa and Asia. RTI Press; 2018; pp. 249–288.
Publisher Full Text

Hoadley U: **How do structured pedagogy programmes affect reading instruction in African early grade classrooms?** *Int. J. Educ. Dev.* 2024; **107**: 103023.
Publisher Full Text

de Hoop T, Ring H, Siwach G, *et al.*: *Endline report for the mixed-methods cluster-randomized controlled trial of Impact Network's eSchool 360 model in rural Zambia.* American Institutes for Research; 2025.

IDInsight: **Quantitative assessment of teacher motivation, classroom practices, and student learning.** 2015.
Reference Source

Jackson K, Makarin A: **Can online off-the-shelf lessons improve student outcomes? Evidence from a field experiment.** *Am. Econ. J. Econ. Pol.* 2018; **10**(3): 226–254.
Publisher Full Text

Jukes MC, Turner EL, Dubeck MM, *et al.*: **Improving literacy instruction in Kenya through teacher professional development and text messages support: A cluster randomized trial.** *J. Res. Educ. Effect.* 2017; **10**(3): 449–481.
Publisher Full Text

Kahneman D, Tversky A: **Prospect Theory. An Analysis of Decision Making Under Risk.** 1977.

Kaur J: **How much do I matter? Teacher self-beliefs, effort, and student learning.** *Job Market Paper.* 2023.
Reference Source

Kraft MA: **Teacher effects on complex cognitive skills and social-emotional competencies.** *J. Hum. Resour.* 2019; **54**(1): 1–36.
Publisher Full Text

Kruger J, Dunning D: **Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments.** *J. Pers. Soc. Psychol.* 1999; **77**(6): 1121–1134.
PubMed Abstract | **Publisher Full Text**

Kyriacou C: *Stress-busting for teachers.* Cheltenham: Stanley Thornes; 2000.

Mbiti IM: **The need for accountability in education in developing countries.** *J. Econ. Perspect.* 2016; **30**(3): 109–132.
Publisher Full Text

McAleavy T, Hall-Chen A, Horrocks S, *et al.*: *Technology-supported professional development for teachers: Lessons from developing countries.* Education Development Trust; 2018.
Reference Source

Menon S, Krishnamurthy R, Kutty S, *et al.*: **Literacy Research in Indian Languages (LiRIL): Research report of a study of literacy acquisition in Kannada and Marathi.** 2017.
Reference Source

Ministry of Human Resource Development (MHRD): *National Education Policy 2020.* Government of India Government of India; 2020.
Reference Source

Miyazaki T: **Is changing teaching practice the mission impossible? A case study of continuing professional development for primary school teachers in Senegal.** *Compare: A Journal of Comparative and International Education.* 2015; **46**(5): 701–722.
Publisher Full Text

Molina E, Pushparatnam A, Rimm-Kaufman SE, *et al.*: **Evidence-based teaching: Effective teaching practices in primary school classrooms.** *World Bank Policy Research Working Paper, (8656).* 2018.

Muralidharan K, Sundararaman V: **The impact of diagnostic feedback to teachers on student learning: Experimental evidence from India.** *Econ. J.* 2010; **120**(546): F187–F203.
Publisher Full Text

Oakley G, King R, Scarparolo G: *An evaluation of ELLN Digital: Technology-supported teacher professional development on early language, literacy, and numeracy for K-3 teachers.* Foundation for Information Technology Education and Development; 2018.

OECD: *Global Teaching InSights: A Video Study of Teaching.* OECD Publishing; 2020.
Publisher Full Text

Piper B, Dubeck MM: **Responding to the learning crisis: Structured pedagogy in sub-Saharan Africa.** *Int. J. Educ. Dev.* 2024; **109**: 103095.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Piper B, Zuilkowski SS, Mugenda A: **Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative.** *Int. J. Educ. Dev.* 2014; **37**: 11–21.
Publisher Full Text

Piper B, Ralaingita W, Mejía J, *et al.*: **Science of teaching: Structured pedagogy how-to guides.** n.d.
Reference Source

Pithers RT, Soden R: **Scottish and Australian teacher stress and strain: A comparative study.** *Br. J. Educ. Psychol.* 1998; **68**: 269–279.
Publisher Full Text

Ramachandran V, Beteille T, Linden T, *et al.*: *Teachers in the Indian education system.* National University of Educational Planning and Administration (NUEPA); 2016.
Reference Source

Romero M, Bedoya J, Yanez-Pagans M, *et al.*: **Direct vs indirect management training: Experimental evidence from schools in Mexico.** *J. Dev. Econ.* 2022; **154**: 102779.
Publisher Full Text

Sabarwal S, Abu-Jawdeh M, Kapoor R: **Teacher beliefs: Why they matter and what they are.** *World Bank Res. Obs.* 2022; **37**(1): 73–106.
Publisher Full Text

Sankar D, Linden T: *How much and what kind of teaching is there in elementary education in India? Evidence from three states.* World Bank; 2014.
Reference Source

Sharma Kukreja S: *Policy and programmatic initiatives to strengthen and sustain efforts for achieving universal foundational literacy & numeracy (FLN) in India.* Isaac Centre for Public Policy (ICPP); 2024.
Reference Source

Shashidhara S, Mamidi P, Gupta G, *et al.*: **Improving parental engagement in foundational literacy and numeracy learning.** *Educ. Stud.* 2025; 1–19.
Publisher Full Text

Stern JMB, Piper B: *Resetting targets: Examining large effect sizes and disappointing benchmark progress (RTI Press Occasional Paper No. OP-0060-1904).* RTI Press; 2019.
Publisher Full Text

STiR & IDinsight: **Impact of STIR's programming on teacher motivation and student learning.IDinsight.** 2018.
Reference Source

Teotia AK, Kumar S: *Teacher Development Coordinator (TDC) program: An evaluative study [Report].* District Institute of Education & Training (N/E), State Council of Educational Research & Training; 2019.
Reference Source

The Behavioural Insights Team (BIT): *EAST: Four simple ways to apply behavioural insights.* London, UK: Behavioural Insights Team; 2014.

Travers CJ, Cooper CL: *Teachers under pressure: Stress in the teaching profession.* London: Routledge; 1996.

UNESCO: *Teaching and learning: Achieving quality for all. EFA Global Monitoring Report.* UNESCO; 2019.

Wolf S, Aber JL, Behrman JR, *et al.*: **Experimental impacts of the "Quality Preschool for Ghana" interventions on teacher professional well-being, classroom quality, and children's school readiness.** *J. Res. Educ. Effect.* 2019; **12**(1): 10–37.
Publisher Full Text

Wolfenden F, Cross S, Henry F: **MOOC adaptation and translation to improve equity in participation.** *Journal of Learning for Development.* 2017; **4**(2).
Publisher Full Text

World Bank: *World development report 2018: Learning to realize education's promise.* Washington, DC: World Bank; 2018.
Publisher Full Text