Countering Misinformation Early: Evidence from a Classroom-Based Field Experiment in India[*]

Priyadarshi Amar[†] Sumitra Badrinathan[‡] Simon Chauchard[§] Florian Sichart[¶]

Abstract

Misinformation poses serious risks for democratic governance, conflict, and health. This study evaluates whether sustained, classroom-based education against misinformation can equip schoolchildren to become more discerning consumers of information. Partnering with a state government agency in Bihar, India, we conducted a field experiment in 583 villages with 13,500 students, using a 4-month curriculum designed to build skills, shift norms, and enhance knowledge about health misinformation. Intent-to-treat estimates demonstrate that treated respondents were significantly better at discerning true from false information, altered their health preferences, relied more on science, and reduced their dependence on unreliable news sources. We resurveyed participants 4 months post-intervention and found that effects persisted, as well as extended to political misinformation. Finally, we observe within-household treatment diffusion, with parents of treated students becoming more adept at discerning information. As many countries seek long-term solutions to combat misinformation, these findings highlight the promise of sustained classroom-based education.

Keywords: Misinformation, Field Experiment, Education, India, Media Literacy

[*]Authors listed alphabetically. This study is funded by the Social Science Research Council. We are grateful to Jeevika and to IAS officers Rahul Kumar and Himanshu Sharma, as well as Pushpendra Singh Tiwari and Rakesh Kumar, for making the Bihar government's collaboration on this project possible. Our implementation partner is DataLeads in New Delhi, India. Our survey partner is Sunai Consultancy in Patna, India. We received IRB approval from American University and the project was reviewed by the UC3M's GDPR data protection officer. We thank Josiah Gottfried for outstanding research assistance, and Minati Chaklanavis, Pranav Chaudhary and Nishant Attri for on-field support. Pre-registration available at OSF. For comments and feedback we are grateful to Danny Choi, Austin Davis, Maria Hernandez de Benito, Andy Guess, Don Green, Erik Kramon, Horacio Larreguy, Gareth Nellis, Laura Paler, Casey Petroff, Laura Schechter, and seminar and workshop participants at George Washington, Penn, American, Rochester, Brown, Michigan, Sciences Po, UMontreal, Columbia, Humboldt, University of Oslo, IC3JM, and APSA and ICA conferences.

[†]Postdoctoral Research Fellow, University Carlos III Madrid & Instituto Carlos 3 - Juan March. priyadarshi.amar@uc3m.es

[‡]Assistant Professor, American University. sumitrab@american.edu. Corresponding Author

[§]Associate Professor, University Carlos III Madrid & Instituto Carlos 3 - Juan March. simon.chauchard@uc3m.es

[¶]PhD Candidate, Princeton University. fsichart@princeton.edu

Introduction

Around the world, educational programs have long been seen as potential catalysts for societal transformation. Political leaders have acknowledged the power of schooling as a key nation-building tool, using education to foster productive citizens, instill civic values, and prepare youth for national political and economic roles (Paglayan, 2024; Wiseman et al., 2011; Ramirez and Boli, 1987). Empirically, numerous studies have examined the causal effects of educational programs in reshaping outcomes that are often resistant to change. For instance, in India, Dhar, Jain, and Jayachandran (2022) showed that engaging adolescents in discussions about gender equality transformed entrenched gender attitudes. In Western Europe, Cavaille and Marshall (2019) demonstrate that an additional year of schooling reduced anti-immigration sentiments later in life. In China, Cantoni et al. (2017) found that school curriculum reforms fostered positive attitudes towards the nation. And in Mali, Gottlieb (2016) demonstrated that civic education resulted in more informed voting decisions among citizens. These studies offer compelling evidence that educational programs can shape and even sustain attitudinal and behavioral change, whether it concerns voting, immigration views, or gender norms – issues often seen as difficult to influence. The success of educational interventions in these areas suggests a promising avenue for addressing another pressing issue: misinformation. In this paper we ask: can sustained, classroom-based education on misinformation meaningfully improve students' knowledge, change norms, and equip them with the skills necessary to resist false information?

A substantial body of research has evaluated the effectiveness of misinformation countermeasures, including fact-checking and corrections (Porter and Wood, 2019; Bowles et al., 2023; Clayton et al., 2019), accuracy prompts (Pennycook and Rand, 2019), inoculation (Pereira et al., 2024; Roozenbeek et al., 2022), and tip-based information (Guess et al., 2020). While many of these strategies show promise, they are typically one-off, online interventions targeting digitally literate, urban populations and are rarely adapted for offline communities (Blair et al., 2023). In parallel, governments and NGOs have increasingly turned to classroom-based media and information literacy programs aimed at the youth, with a global uptick in such initiatives since 2016. For example, New Jersey has advanced mandatory K–12 media literacy education (Sitrin, 2020), echoing efforts in California, Estonia, and Finland. Theoretically, these programs share important design features: they emphasize repetition, peer-based learning, and delivery by authority figures who can help shape norms. Yet despite their growing adoption and comparatively high cost, there is a striking lack of causal evidence assessing their impact. To date, no study has rigorously evaluated the effects of sustained, classroom-based interventions on misinformation outcomes.[1] This reveals a significant disconnect between the academic literature on misinformation interventions and the types of programs being implemented in the real world.

To fill this gap, we conducted a field experiment in 583 villages in Bihar, one of India's least developed states, involving over 13,500 adolescents aged 13–18. India, where misinformation has led to health crises and to political violence (Siddiqui, 2020; Badrinathan, Chauchard, and Siddiqui, 2024), is a critical case for understanding how falsehoods spread and persist: it represents a combination of low state capacity, shrinking independent media, and elite-driven disinformation in a context of increasing polarization, making misinformation an issue as crucial as it is challenging to address. Our intervention targeted students in grades 8 through 12 and consisted of classroom-based sessions on misinformation. Over a 14-week period, students participated in four 90-minute sessions, held approximately every three weeks, with homework assignments between sessions. The curriculum, designed specifically for this context but building on principles of media and information literacy initiatives across the world, focused on health misinformation and aimed to (1) enhance scientific knowledge about health and counter health-related misinformation, (2) equip students with broad critical skills and practical tools to encourage a more responsible consumption of information, and (3) shift norms surrounding misinformation.

Crucially, we partnered with the Bihar state government – specifically, with the Bihar Rural Livelihoods Promotion Society (BRLPS), commonly known as Jeevika – to deliver the intervention as an official course offered through the government.[2] This helped enhance the legitimacy and reach of the intervention, reducing non-compliance, and simulating a real-world rollout of a government program. We employed a placebo-controlled design, with control villages receiving classes on conversational English, ensuring equivalent engagement with a long-term program

[1] A partial exception is Apuke, Omar, and Asude Tunca (2023) which reports positive effects from a six-week media literacy course in Nigeria, though concerns about sample size, spillovers, and compliance limit its internal validity.

[2] Jeevika is run autonomously by officers from the Indian Administrative Service under both the Bihar state government's Department of Rural Development and the Indian government's Ministry of Rural Development. See https://brlps.in/overview.

and only varying the content of instruction.

We hypothesized that the treatment would influence a range of attitudes and behaviors related to misinformation. Specifically, because our intervention included modules designed to strengthen these competencies, we expected it to increase students' awareness of the risks posed by misinformation, enhance the ability to distinguish true from false content, reduce likelihood of sharing false information, and improve capacity to assess source credibility and identify evidence-based health practices. Further, as the curriculum incorporated normative discussions and hands-on exercises focused on combating misinformation, we also anticipated that it would boost respondents' willingness to engage and participate in corrective actions and misinformation counter-measures.

Intent-to-treat estimates measured soon after the intervention ended indicate that it had a strong and significant impact on students' capacity to comprehend and process information, as well as to apply classroom teachings to real-life contexts. At the conclusion of the curriculum, treated respondents demonstrated heightened discernment in evaluating information and making decisions regarding the sharing of news items (0.32 SD), with effect sizes substantially larger than those previously identified. Notably, the intervention also brought about changes in health preferences (0.21 SD), diminishing reported reliance on alternative medical approaches to cure serious illnesses, as well as changes in ability to evaluate the credibility of sources (0.21 SD). Finally, while intent-to-treat estimates showed no overall effect of the treatment on behaviors regarding misinformation countermeasures, it did result in willingness to change costly behaviors among boys, suggesting that such changes may be more difficult in contexts where conservative gender norms act as barriers for girls.

Strikingly, we found that these effects persist over time. We resurveyed a random subsample of 2,059 participants 4 months after the intervention and detected significant effects on students' ability to discern true from false information (0.26 SD). Crucially, our second endline survey included a battery of political items that were not discussed in the classroom and not included in the first endline. We find that there are large effects on these entirely new items – respondents were better able to discern true from false political news 4 months after an intervention that focused solely on health misinformation (0.31 SD), demonstrating that they were able to learn from the treatment, retain its lessons, and apply it to new, and polarizing, domains. Finally, we also find that parents of treated students are better able to discern true from false information, demonstrating the ability of sustained educative interventions to have within-household treatment diffusion, and trickle-up socialization from children to parents (Carlos, 2021; Dahlgaard, 2018). Several of the outcomes we measure assess and require the application of skills rather than relying solely on recall. As a result, expressive responding and social desirability biases are less likely to have influenced these outcomes, as they emphasize critical thinking rather than recall-based responses.

This study has significant implications not only for the literature on countering misinformation but also for the creation of education policy and public health strategies, and for work on behavioral change in developing countries. Its findings contribute to several academic literatures: to work in American politics advancing knowledge on information and persuasion broadly (Huber and Arceneaux, 2007; Coppock, 2023); to experimental methods, focusing on theory and practical strategies for communicating scientific ideas (Andrews and Shapiro, 2021; Alsan and Eichmeyer, 2024); to comparative politics, especially research examining how public infrastructure can strengthen democratic outcomes (Green et al., 2024; Boas and Hidalgo, 2011; Gottlieb, Adida, and Moussa, 2022); and finally to work focusing on politics in South Asia, exploring effective informational and behavioral interventions to enhance governance and societal outcomes (Dhar, Jain, and Jayachandran, 2022; Ghosh et al., 2025; Cheema et al., 2023; Banerjee et al., 2014).

Sustained Classroom Education Against Misinformation

The global rise in misinformation has prompted intense academic and policy interest in the topic (Persily, Tucker, and Tucker, 2020), leading to a proliferation of studies and interventions to counter it. Among these, media and information literacy has emerged as a popular approach. In 2021, the United Nations General Assembly called on member states to develop policies and strategies to promote media and information literacy (MIL hereafter). UNESCO followed suit, rolling out 26 MIL programs across 59 countries with nearly $5 million in funding between 2022 and 2023. Governments have acted as well: New Jersey, for example, became the first U.S. state to mandate MIL education from kindergarten onwards, and Finland has long incorporated it into school curricula. While these initiatives include a diversity of theoretical and practical modules, they tend to share three features which distinguish them from other misinformation countermeasures: (1) instruction delivered in group or classroom settings, (2) guidance by a trusted authority figure, typically a teacher, and (3) repeated exposure over time to encourage retention and norm internalization.

But while these elements characterize policy-led initiatives, academic scholarship on media and information literacy looks starkly different. We show in Table 1 a list of experimental studies that describe their interventions as "media literacy" or related labels (e.g., digital or news literacy). Most are brief, one-off treatments: nudges, reminders, or short videos (Ali and Qazi, 2023; Gottlieb, Adida, and Moussa, 2022), typically only minutes long, with the longest being an hour (Badrinathan, 2021). They tend to lack the extended, classroom-based, and socially embedded components emphasized by policy initiatives. Taken together, these observations reveal a critical gap: the model of MIL training now increasingly adopted in the real world has never been causally evaluated in academic research. There is, thus, a large divergence between how policymakers define media and information literacy, and how academics tend to operationalize it. Despite the growing adoption of MIL initiatives, credible evaluations of their causal effects remain absent, without which we cannot rule out the possibility that media and information literacy programs might be ineffective or even counterproductive.

Our study addresses this gap. The intervention we design and evaluate in this paper is called BIMLI, the *Bihar Information and Media Literacy Initiative*, a long-term program focused on equipping students with tools to recognize and resist misinformation. BIMLI evaluates, by design, a fundamentally different model from those examined in the existing academic literature: a classroom-based intervention which actively mimics the initiatives policymakers are implementing across the world. We highlight a few design features of the intervention. First, in terms of mode of delivery, we administered the program face-to-face, fostering a peer-based, interactive environment, where respondents encountered key lessons repeatedly over multiple sessions delivered by an instructor. Research suggests that peer interactions in classroom settings can deepen understanding by exposing learners to diverse perspectives (Dhar, Jain, and Jayachandran, 2022), while repeated exposure allows for reinforcement of concepts (Fazio, Rand, and Pennycook, 2019), and authority figures promote norm-building (Tankard and Paluck, 2016). Second, to mimic the governmental and international organization support for such initiatives

Table 1: Examples of Media Literacy Interventions

Study	Intervention Label	Operationalization	Dosage (duration)
Gottlieb, Adida, and Moussa (2022)	”Digital Literacy”	A video of a French journalist from Fact News presenting tips	Four minute video
Badrinathan (2021)	”Media Literacy”	Training on two tools to verify information plus demonstrations from research personnel	One hour training
Ali and Qazi (2023)	”Digital Literacy”	Video with info. about fake news and tips + personalized enumerator feedback	Three minute video
Guess et al. (2020)	”Digital literacy”	One time exposure to six strategies that readers can use to identify false or misleading stories	A few minutes to read the strategies
Hameleers (2020)	”News Media Literacy”	Article with a misinformation stimulus, a written fact-check below it, followed by 3 tips to spot misinformation	Approx. 10 minutes to read treatment materials
Tully, Vraga, and Bode (2020)	”News Literacy”	Single tweet reminding people to evaluate and be critical news consumers	Approx. one minute to read the tweet

across the world, we secured partnership with an agency of the Bihar state government to roll out the program as an official government-offered course.

A key contribution of our work, therefore, is empirical: this study is the first (to our knowledge) to evaluate the causal effect of a sustained, classroom-based media literacy program. However, there are also strong theoretical reasons to implement this approach. The structure of our program, mirroring existing efforts in the real world, is grounded in how children best learn and retain complex information: via peer learning, in a legitimate setting, with instructors, and over time (Fazio, Rand, and Pennycook, 2019; Dhar, Jain, and Jayachandran, 2022).

Media literacy treatments in the academic literature have a mixed record, resulting in either weak or null findings (Blair et al., 2023). Studies in the Global North report modest positive effects, but every media literacy intervention conducted in the Global South to date has produced null results, leading some to question whether the optimism around media literacy is warranted (Blair et al., 2023). However, it is possible that limitations with existing work lie in how interventions have been designed and delivered, which do not sufficiently adapt to local constraints and information environments, or to best practices around how children learn and internalize information. Taking stock of this, we designed a treatment to mimic how education programs against misinformation are actually delivered on the ground: face-to-face, long-term, and integrated into existing school structures. In a context like India where information sharing is predominantly offline (Gadjanova, Lynch, and Saibu, 2022), this delivery model is not only practical but also necessary.

Our argument for media and information literacy thus owes partly to contextual fit. Digital interventions such as algorithmic labeling or online corrections and fact-checking are widely studied but largely irrelevant in our setting, where only one in ten respondents owns a personal mobile phone. We are therefore agnostic about their applicability in these low-connectivity environments. Lighter, critical thinking-based approaches like inoculation or nudges could theoretically be adapted for offline use, but existing evidence suggests limited success in similar low-literacy, low-access populations (Roozenbeek et al., 2022; Guess et al., 2020). For instance, Badrinathan and Chauchard (2023b) and Guess et al. (2020) find positive effects from tip-based and social correction interventions, but only among urban, internet-using, English-speaking Indians. In contrast, results are null when these interventions are deployed offline: Guess et al. (2020) find no effects from face-to-face tips; Harjani et al. (2023) report null results from inoculation adapted to offline settings; and Badrinathan (2021) documents null effects from a face-to-face digital literacy campaign.

Importantly, these interventions are all one-time, single-session treatments. These patterns suggest that beyond contextual fit, dosage may also matter. Thus, we adopt a different mode of delivery that allows for repeated exposure. Finally, we expand the content to go beyond simple nudges or reminders. Rather than assuming individuals already possess the necessary skills to counter misinformation, our approach actively provides these skills while simultaneously targeting normative change.

Consequently, our study also contributes to theoretical debates about misinformation. Given the mixed results of prior media literacy work, our intervention serves as a critical test. We evaluate a rigorous, contextually grounded program that closely mirrors real-world efforts – delivered in classrooms, over time, by credible authority figures. If such a comprehensive intervention fails, it would raise serious doubts about the efficacy of media literacy as a strategy. But if successful, it suggests that past null results may reflect weak implementation rather than theoretical limits. This shifts the theoretical conversation. Rather than attributing the failure of media literacy solely to motivated reasoning or psychological resistance (Taber and Lodge, 2006; Flynn, Nyhan, and Reifler, 2017), we highlight the importance of implementation, delivery mechanisms, and contextual fit. In doing so, we offer a more optimistic – but also more demanding – theoretical account of how and when corrective information can reduce belief in falsehoods.

The Politics of Misinformation in India

Health misinformation is widespread in India. For instance, from our own control group data, 55% of respondents reported believing that exorcism can cure snake bites. In other studies from similar contexts (Chauchard and Badrinathan, 2025), over 60% of respondents claimed that cow urine could cure covid-19. While this type of belief may seem harmless, it can have severe consequences by discouraging citizens from seeking actual medical solutions and leading to potentially fatal outcomes (Bridgman et al., 2020). The negative consequences of belief in misinformation may be particularly pronounced in regions with lower levels of state capacity and socio-economic development (Badrinathan and Chauchard, 2023a).

In India, such deeply entrenched beliefs are closely tied to social identities and are often exploited by political elites to gain electoral support. Traditional health remedies, many rooted in ancient Hindu culture, have been used to appeal to Hindu voters – particularly under the Hindu nationalist Bharatiya Janata Party (BJP), which currently leads the federal government and portrays itself as a defender of Hindu values (Jaffrelot, 2021). One striking example involved a member of parliament hosting a public event promoting cow urine as a COVID-19 cure, which resulted in several hospitalizations (Siddiqui, 2020). Research shows that misinformation tied to long-standing identities is especially resistant to correction (Nyhan, 2021), and India's enduring Hindu-Muslim cleavages make religious identity a particularly potent factor in belief formation (Brass, 2011; Chauchard and Badrinathan, 2025). When elites deliberately reinforce falsehoods to polarize, such misinformed beliefs can be especially persistent: evidence from India suggests that motivated reasoning can impede correction efforts (Taber and Lodge, 2006; Badrinathan, 2021). Bihar, our study site, is part of a larger northern Indian media ecosystem including neighboring states like Uttar Pradesh, where elite-driven disinformation has sometimes resulted in violence (Badrinathan, Chauchard, and Siddiqui, 2024).

For citizens in such contexts, finding ways out of the misinformation trap can be challenging. This is particularly true in Bihar, India's poorest state and home to 127 million people, where one-third live below the poverty line. The state's relative underdevelopment translates into a lack of essential services such as healthcare and education, alongside the failure of many public programs (Sharma, 2015). The population we study faces profound structural barriers to learning. Children in Bihar, especially girls, are significantly less likely to attend school compared to those in other states (Muralidharan and Prakash, 2017). Students often work for wages instead of attending school, teacher absenteeism is common, infrastructure is lacking (many classrooms lack electricity, seating, or basic materials), and learning suffers: only about half of Indian children enrolled in grade five can read a simple paragraph at the second-grade level (50.1 percent of children), or solve a two-digit subtraction problem (52.3 percent of children).[3] These alarming statistics have opened a serious debate on "what works" to improve learning in India, sparking a robust literature on education-based RCTs to which we contribute (de Barros et al., 2022).

Access to the internet is also limited: according to our baseline data, only 11.5% of respondents own a personal cellphone, and only 19% reported using the internet. With most interactions and information sources offline, children largely depend on their families for information. Yet adults may themselves be misinformed, and strong cultural norms of deference to elders make it difficult for children to question them (Malhotra and Pearce, 2022). Even in households with internet access, it is typically via a shared mobile phone, marking a sharp contrast with Western contexts, where access is individualized (Steenson and Donner, 2017). Limited connectivity is further exacerbated by a deteriorating informational environment. Independent media and dissenting voices are increasingly under threat, as state capture of institutions, including news outlets, grows (Mohan, 2021; Sen, 2023). These trends reflect broader patterns of democratic decline in India, where the space for credible information has narrowed significantly alongside eroding state capacity (Tudor, 2023).

While vulnerability to misinformation can be thought of as a country-wide problem, Bihar faces distinct structural challenges related to state capacity, compounded by a nexus of elite-backed disinformation, weak institutions, lack of credible media, and low socio-economic status.

[3] Data from the Annual Status of Education Report in India.

Experimental Design and Data Collection

We implemented a field experiment to test the efficacy of the BIMLI program with a sample of 583 villages across 32 districts of the state of Bihar. Treatment was assigned at the village level, with participants clustered within villages having the same treatment status. Respondents in treatment villages received classroom lessons about misinformation, and we included a placebo control condition for comparison (additional details below).

The Treatment

The BIMLI program featured four classroom sessions, each about 90 minutes long and approximately 2-3 weeks apart, as well as homework assignments between sessions. We created a custom curriculum and lesson plan for this study. In doing so, our educative curriculum, though bundled, focused on media and information literacy and critical thinking, with the goal of changing norms and providing knowledge and skills. In Table 2 we provide a summary of the treatment lesson plan, including a description of learning objectives, modules included in each session, key theoretical works on which curriculum design relied, and strategies to tailor the lesson to the local context.

The BIMLI curriculum aimed to achieve two core objectives: (a) enhancing knowledge through factual and skills-based learning, and (b) shifting social norms surrounding misinformation. We conceptualize knowledge in two dimensions. The first is recall: the ability to remember specific facts taught in class. The second is application: the ability to use general tools acquired in class to critically assess new information, such as evaluating emotional language, identifying unreliable sources, or pausing before sharing content. We also sought to influence norms, shaping what students perceived as acceptable to believe, share, or correct within their social circles. Curriculum modules explicitly addressed the dangers of misinformation, its societal relevance, and how individuals can intervene when others spread false claims. Because educational institutions often serve as powerful sources of normative influence (Tankard and Paluck, 2016), government-backed implementation and teacher-led delivery likely reinforced these messages (Paluck and Shepherd, 2012). In targeting both cognitive and normative dimensions, BIMLI aimed to foster durable shifts in attitudes and behaviors.

The curriculum emphasized interactive instruction, encouraging engagement between teachers and students as well as among students themselves, approaches notably lacking in many Indian classrooms where rote memorization and passive instruction dominate (Kumar, 1986; Bhattacharya, 2022). This approach aimed to cultivate analytical thinking and deep learning rather than relying solely on passive reception of information, representing a significant departure from the traditional structure of schooling in India (Kumar, 1986). For instance, in Session 4 the lesson plan incorporated role-playing exercises in the classroom. In one activity, a student took on the role of a child while another acted as a parent, with the child tasked with employing strategies to engage with a parent that shared misinformation at a family dinner. The scenario aimed to highlight the challenges of addressing health misinformation with adults, particularly when such discussions involve confronting deeply ingrained beliefs in settings where confrontation with adults is discouraged (Malhotra and Pearce, 2022).

Finally, a key instructional goal of this program was to focus on fostering critical thinking rather than offering prescriptive tips to spot misinformation. This approach was particularly suited to the Indian context, where much information is shared through friends and family, making source-specific advice (e.g., favoring one TV channel over another) ineffective. Given the decline in mass media credibility amid democratic backsliding (Mohan, 2021), we also avoided endorsing specific media outlets. Instead, we emphasized cues to critically assess information, such as recognizing emotional tone, not relying on shared ethnic identities as a cue to assess information, and identifying appropriate authorities as credible sources for specific topics – for instance, relying on community health workers employed by the government (called ASHA workers) for health-related information. Substantively, the curriculum relied on examples related to health misinformation.

We collaborated with DataLeads, a Delhi-based media literacy organization, along with local Bihar educators and Indian experts, to co-develop a standardized curriculum, including time-use lesson plans for instructors to ensure consistent classroom delivery.[4] To reinforce learning beyond the four in-person sessions, we assigned reflective homework – such as story writing, observations, and family discussions – and distributed concise take-home summary sheets after

[4] Appendix B provides an overview of the materials used in the treatment.

Table 2: The BIMLI Curriculum

	Learning Objectives and Content	Theoretical Works	Tailoring to India
Session 1: *"Understanding the Fundamental Elements of Media and Information Literacy"*	Objectives: 1. Introduce topic; 2. Define key terms; 3. Raise awareness of perils of misinformation. Modules: 1. Intro: the changing nature of information; 2. Definitions: what misinformation is and is not; 3. Where misinformation appears (examples); 4. Adverse consequences of misinformation on health, violence, etc.	Guess and Lyons (2020)'s definition of misinformation and several journalistic examples of recent misinformation and its effects.	Examples and illustrations all local: health misinformation leading to vaccine hesitancy in India, falsehoods and doctored images on Indian WhatsApp groups.
Session 2: *"Understanding Biases and Critical Thinking"*	Objectives: 1. Develop critical thinking skills; 2. Develop awareness of human biases in info. consumption; 3. Develop awareness of media biases in the production of information. Modules: 1. Recap of session 1; 2. Intro to human psychological biases like confirmation bias; 3. News and media system biases; 4. Critical thinking - definitions and strategies to enhance one's critical thinking.	Motivated reasoning from Taber and Lodge (2006), list of psychological biases adapted from Roozenbeek et al. (2022), list of media biases adapted from Ashley, Maksl, and Craft (2013), news framing effects from Druckman and Nelson (2003), fact-opinion discernment from Graham and Yair (2025).	Introduction to the news media environment in India + how biases manifest in the Indian context (for example, scapegoating minorities).
Session 3: *"Identifying Reliable Sources, Verifying and Sharing information"*	Objectives: Provide students with tools and tips to 1. Evaluate sources; 2. Evaluate the accuracy of information; 3. Decide what information is worth sharing. Modules: 1. Recap of sessions 1 and 2; 2. How to evaluate the reliability of sources; 3. How to evaluate veracity & verify information; 4. How to decide whether to share information.	Concrete examples of tips to spot misinformation (Guess, Nagler, and Tucker, 2019; Vraga, Bode, and Tully, 2022; Badrinathan, 2021), focus on sharing as different from belief (Brashier and Schacter, 2020).	Tailored Indian examples focused on WhatsApp such as reverse image search, looking for the 'forwarded many times' tag, introduction to Indian fact-checking websites.
Session 4: *"Vaccine Importance and Talking About Misinformation with Family"*	Objectives: 1. Highlight importance of correcting/combating misinformation; 2. Develop strategies to deal with friends and relatives who spread misinformation. Modules: 1. General recap of lessons 1-3; 2. Strategies to fight misinformation at home; 3. Role play and memory games.	Efficacy of social corrections (Bode and Vraga, 2018; Badrinathan and Chauchard, 2023b), talking to family and community about misinformation (Pearce and Malhotra, 2022).	Role-playing exercise and games adapted to context, for example how to talk to an elder Indian relative about misinformation.

each session to serve as reference guides.

Administering classes

To bolster the credibility of BIMLI, we signed a memorandum of understanding to secure official collaboration with an agency of the Bihar state government, the Bihar Rural Livelihoods Promotion Society (BRLPS, or as it is commonly known, Jeevika). Despite their governmental affiliation, Jeevika operates autonomously under the leadership of an Indian Administrative Services officer. To ensure its broad acceptance, Jeevika promoted the program as an official government-offered certified course, enhancing its credibility. This allowed us to reach remote rural populations often underrepresented in misinformation research.

In our study, participants were school students in grades 8 through 12, aged between 13 and 18 years old. To dispense the intervention classes, Jeevika made available to us 100 community libraries across 32 districts in Bihar.[5] We ran our classes in these libraries from November 2023 to March 2024, delivering classes after school hours. The libraries were equipped with essential infrastructure – seating, blackboards, and other class equipment – which offered a level of standardization we would not have easily achieved in public schools. These libraries were also relatively new constructions which allowed for conducive classroom settings that may have encouraged attendance, otherwise a major problem across the state's public schools.[6]

Recognizing that program success depends not only on student compliance but also on teacher attendance and quality, we recruited a separate pool of teachers rather than using existing public school staff. Meetings with government officials revealed that public school teachers in Bihar are often overburdened, with high rates of absenteeism among both teachers and students, making compliance difficult. Each recruited teacher visited a given classroom approximately once every two to three weeks.[7] The curriculum was designed to be taught fully offline, us-

[5]These 100 libraries were located in 100 distinct blocks across the 32 districts.

[6]Data from the ASER survey, the Annual Status of Education Report which provides data from annual surveys on children's schooling and learning levels in rural India, highlights some of these issues in public schools. For example, their 2022 report points out that on the days that ASER surveyed schools, only 50% of enrolled children were actually present in public schools in Bihar.

[7]DataLeads, our consulting partner, put out an ad to recruit teachers and received 400 applications; they selected 50 teachers through interviews and a two-day training. The final cohort included school teachers, journalists, professors, and fact-checkers. Each was assigned 6–9 classrooms across 2–3 libraries and remained with the same classrooms throughout.

ing face-to-face discussion, printed materials, and minimal digital tools—mirroring the typical learning environment of rural schoolchildren in India.[8]

These choices, by design, were aimed at maximizing the likelihood of detecting treatment effects, if they existed, by incentivizing enrollment and sustained participation. Bihar is India's poorest state, and our intervention required respondents to voluntarily attend additional, uncompensated sessions. In a context where time in class competes with income-generating work or caregiving, this posed a significant barrier. Compounding this, as mentioned earlier, students often read below grade level and many do not complete school. The broader literature on information provision reinforces the importance of our design choices to mitigate these structural issues. Randomized evaluations in similar settings show that information provision alone often fails to shift beliefs or behavior. Scholars note that constraints like low trust, limited resources, and weak incentives hinder treatment uptake unless interventions also generate salience and reinforce the perceived efficacy of action (Kosec and Wantchekon, 2020). Social dynamics matter as well: peer environments can alter receptivity (Lieberman, Posner, and Tsai, 2014). Implementation challenges also loom large: in many developing contexts, inadequate state capacity or lack of elite buy-in undermines program success (Rao, Ananthpur, and Malik, 2017). To address these issues, we designed the intervention to take place in a trusted classroom setting, partnered with government institutions to bolster legitimacy, emphasized peer learning, and maintained close oversight of implementation. We sought, ultimately, to minimize technical and implementation failures so that any null effects would more clearly reflect limitations of the underlying concept rather than execution and delivery (Karlan and Appel, 2016).

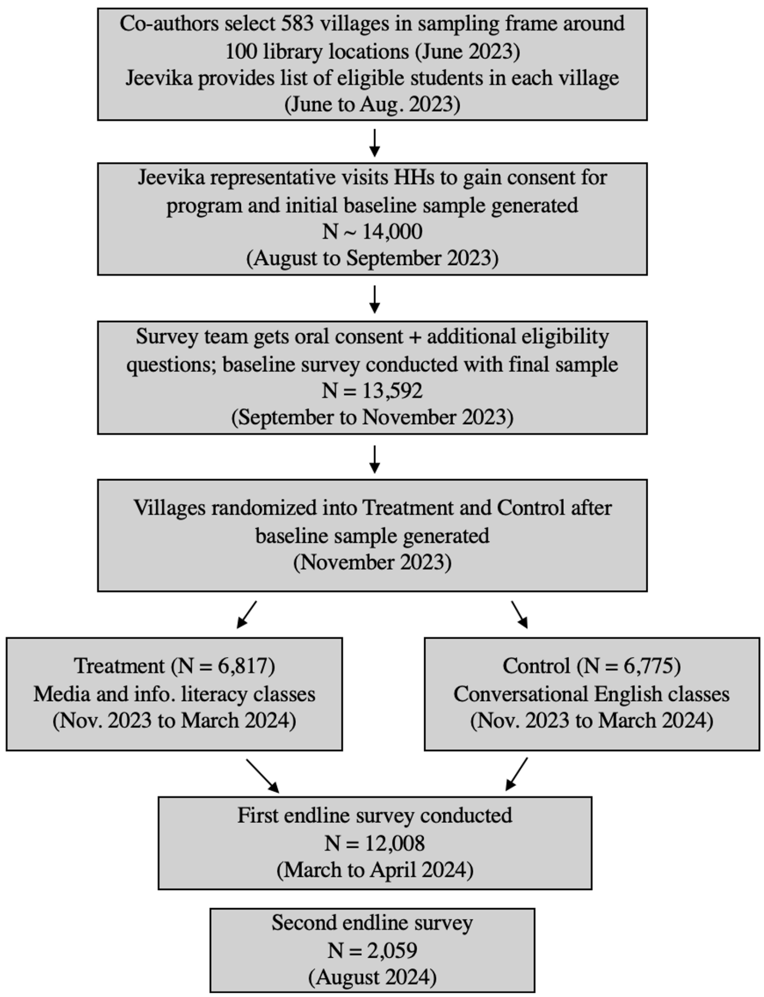
Sampling, enrollment and baseline data

Figure 1 outlines the timeline and flow of recruitment and roll-out of the study. We sampled $\approx$ 6 villages within a 3 km distance from each of the 100 libraries and randomized roughly 50% of these to receive our treatment; the remaining served as control villages. Our sampling procedures resulted in the selection of 583 villages across the state of Bihar. Before randomization, we categorized each village as having either high or low spillover potential, based on how many

[8]See Appendix B1 for a classroom session example.

sampled villages fell within the same Gram Panchayat (GP) – a local administrative unit encompassing multiple villages. Spillovers were expected to be higher within the same GP, as children from those villages are more likely to attend the same schools or classes. Therefore, GPs with multiple selected villages were classified as high-spillover, while those with only one selected village were considered low-spillover. We then randomly assigned to treatment and control within each library area and spillover strata (see Appendix A for details).

Figure 1: Study Flow and Timeline



In each of the 583 selected villages, Jeevika provided household lists based on enrollment in state programs. From these, households with children in grades 8–12 were identified as eligible for our program. Jeevika staff visited these homes to explain the program, producing a final list of 20–25 interested households per village. Our survey team then conducted an in-person baseline survey before randomization. Crucially, we note that randomization occurred *after* students opted in, avoiding issues with differential opt-in rates between treatment and control. Everyone involved in the study – including teachers, implementation partners, government officials, and coauthors – were blind to treatment status during recruitment and baseline data collection. During household visits, the recruitment pitch stated that students could participate in a free, government-endorsed certificate course with four sessions, designed to benefit their future careers. Students were unaware of their treatment assignment until the first session.[9]

The baseline survey collected demographic, household, and attitudinal data, including items on perceptions of the state, media usage, views on science and vaccines, and social ties. Our baseline sample included 13,592 respondents across 583 villages, with 49.9% assigned to treatment and 50.1% to control.[10] In Appendix A, we show balance tables confirming that respondents in treatment and control groups were balanced on key demographics, attitudes, and behaviors. The Appendix also shows that treatment and control villages themselves were balanced on key variables based on census parameters.

Control condition

Control group units participated in four modules of conversational English language classes, serving as a placebo rather than a pure control. This was done to achieve parity in effort exerted by students, since school attendance is a major problem in Bihar, and since our intervention lasted 4 months. We aimed to create comparable classroom dynamics and peer interactions, varying only the content of instruction. We additionally wanted control respondents to benefit from the program and hence chose a topic that fostered engagement without being related to misinformation outcomes. Subjects like math, science, and history were excluded due to overlap with standard curricula or national identity narratives, and non-academic topics like cooking were discarded due to expected gender biases in their uptake. We ultimately implemented a

[9] Appendix A shows locations of treatment and control villages across Bihar.

[10] The sample was 58% girls, with respondents ranging from grades 8 to 12 (median grade 10), and 96% enrolled in government schools. It was 91% Hindu and 69% OBC, on par with state census demographics. Language diversity included 43% Hindi-speaking households, 30% Bhojpuri, and 9% Magahi. Fathers' median education was grades 6-9, and mothers' median education was grades 1-5. Socio-economic indicators at the household level showed 15% owned a refrigerator, 3.6% a washing machine, and 19% had access to an internet-enabled mobile phone. Trust in media was high: 90% for newspapers, 84% for TV, and 61% for social media. While 77% were vaccinated for COVID-19, 87% believed in alternative medicine like Ayurveda and homeopathy.

curriculum of four sessions on basic conversational English given students had very limited prior exposure. The curriculum focused on spoken skills, covering self-introduction, naming objects, describing activities, and asking questions, using role-playing and group exercises similar to those in the treatment group (see Appendix B). Topics avoided media, technology, and politics, and the very basic instruction level was unlikely to enable control students to independently navigate new information sources.[11]

Endline data and compliance

Our first endline survey was conducted in-person in the weeks following the end of the fourth and last session. Because of the large sample, the endline took 5 weeks to complete, and we were able to re-contact 12,008 of the houses sampled at baseline, with an attrition rate of 11.3%. There is no significant difference in attrition between treatment and control, although we do find that attrition is lower among girls and higher among older respondents (see Appendix K1). Moreover, from fieldwork and interviews with enumerators, we note that houses that attrited at endline did so because we were unable to contact them after several tries (in most cases, this was because the respondent was not at home). Crucially, no household refused our survey team entry for the endline survey. We conducted a second endline survey about 4 months, on average, after the intervention, to assess if treatment effects persisted over time. This survey was conducted over the phone with a random subset of 2,059 students and, in each case, one parent or adult guardian.[12]

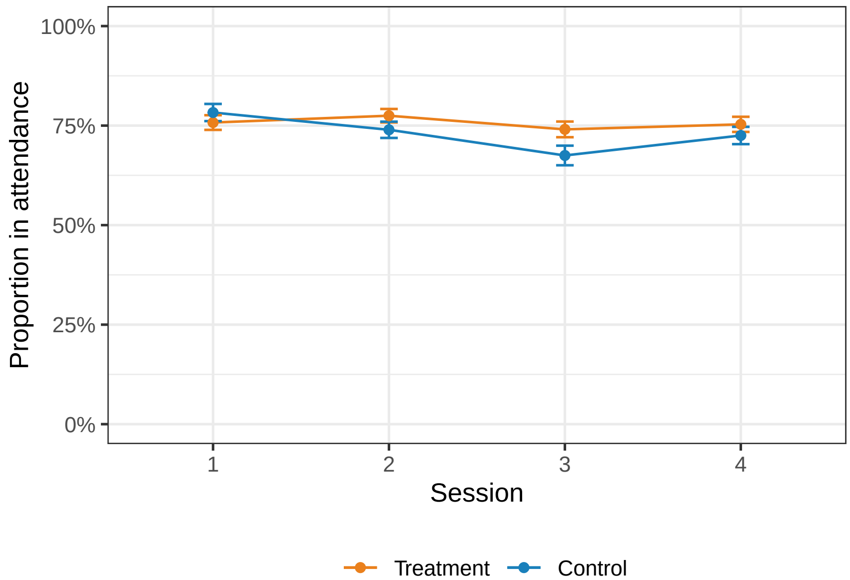
To boost compliance, we implemented a detailed monitoring system. Jeevika staff, women known locally as *didis*, regularly reminded households about upcoming classes. Students were motivated by the promise of a government-issued certificate upon completing the program. External monitors also made random visits to verify teacher presence and adherence to class sched-

[11] The teacher selection and training differed between the treatment and control groups. DataLeads recruited and trained treatment teachers, while English class teachers were recruited via a local Bihar consultant, resulting in variations in socio-economic characteristics and teaching experiences. Consequently, the treatment effects we measure are influenced by both the treatment content and the teachers' differing backgrounds, and we are unable to implement teacher fixed effects. Appendix E summarizes teacher demographics by group.

[12] The time gap between the first and second endline surveys varied across households because it took about 30 days to survey all homes in each round. For some respondents, the gap was around 3 months, while for others, it extended to 5-6 months. Therefore, we report an average gap of 4 months.

ule. Co-authors also visited during initial and final sessions.

Figure 2: Compliance data across treatment and control



Teachers were required to upload respondent-level attendance data after each session via an app. On average, students attended 2.97 classes and 52.7% of the sample attended all four classes across treatment and control. We detect no significant difference in attendance numbers across treatment and control, with similar proportions attending both sets of classes. However, we do see a significant drop off in attendance for control group respondents during session three, though we note that the difference is substantively small (67% in control group and 74% in treatment) and dissipates during session 4 (see Figure 2). Further, we find that girls were more likely to attend classes compared to boys.[13] Importantly, from an inference point of view, since our main specification estimates the ITT and not the ATE, (lack of) differential effects on attrition (Appendix K) are more crucial relative to the few differential effects on compliance that we detect (Appendix C).

[13]Girls' higher rates of compliance and lower rates of attrition may be attributed to Jeevika's women-led structure, which likely encourages their participation, and the library serving as a rare safe space for girls after school. Unlike boys, who have various options for public spaces like sports, girls have limited alternatives. Additionally, the initial sample consisted of 58% girls to begin with.

Outcome measures

We hypothesized that the intervention would influence a range of misinformation-related attitudes and behaviors. First, since each session highlighted the prevalence and dangers of misinformation, we expected students' awareness of the issue to increase. Second, given that the curriculum explained what misinformation is (session 1), how people process information in biased ways (session 2), and how to assess accuracy (session 3), we anticipated improvements in students' ability to distinguish true from false information. Third, by repeatedly emphasizing the harms of misinformation and providing concrete sharing strategies (especially in session 3), we expected the program to reduce students' likelihood of sharing false content. Fourth, through critical thinking exercises and practical tips for evaluating material, we hypothesized gains in students' ability to assess source credibility. Fifth, since all examples were health-related by design, we expected the program to increase students' knowledge of and trust in scientifically-vetted health strategies. Finally, because the curriculum integrated normative messaging and practical exercises (particularly in session 4) we expected greater willingness among students to take action against misinformation.

Building on these intuitions, we pre-specified and included seven distinct families of outcomes in the first endline survey: accuracy discernment, sharing discernment, health attitudes, trust in sources, engagement with misinformation countermeasures (attitudes), engagement with misinformation countermeasures (behaviors), and awareness of misinformation.[14] Each outcome family comprises multiple survey items. For the analysis, we construct inverse-covariance weighted (ICW) indices that aggregate and weight these items, standardized relative to the control-group mean and SD. Our primary analyses focus on these seven indices. Appendix D outlines the rationale for using ICW indices, their pre-specified construction, and correlations between outcome measures. In the second endline, we measured accuracy discernment for both the respondent and one parent or guardian.

[14]Our pre-analysis plan was posted to OSF before endline data collection in February 2024 and is available here.

Estimation and Results

Due to the possibility of non-compliance, our main specification estimates the intent-to-treat ITT_Y effect: the effect of being assigned to the treatment group. To test hypotheses about the overall effect of the treatment on average outcomes, we use the following two models:

$$Y_{ijk} = \beta_0 + \beta_1 T_{ijk} + \sum_{k=1}^{m-1} \gamma_k + \varepsilon_{ijk} \tag{0.1}$$

$$Y_{ijk} = \beta_0 + \beta_1 T_{ijk} + \sum_{c} \alpha_c X_{ci} + \sum_{k=1}^{m-1} \gamma_k + \varepsilon_{ijk} \tag{0.2}$$

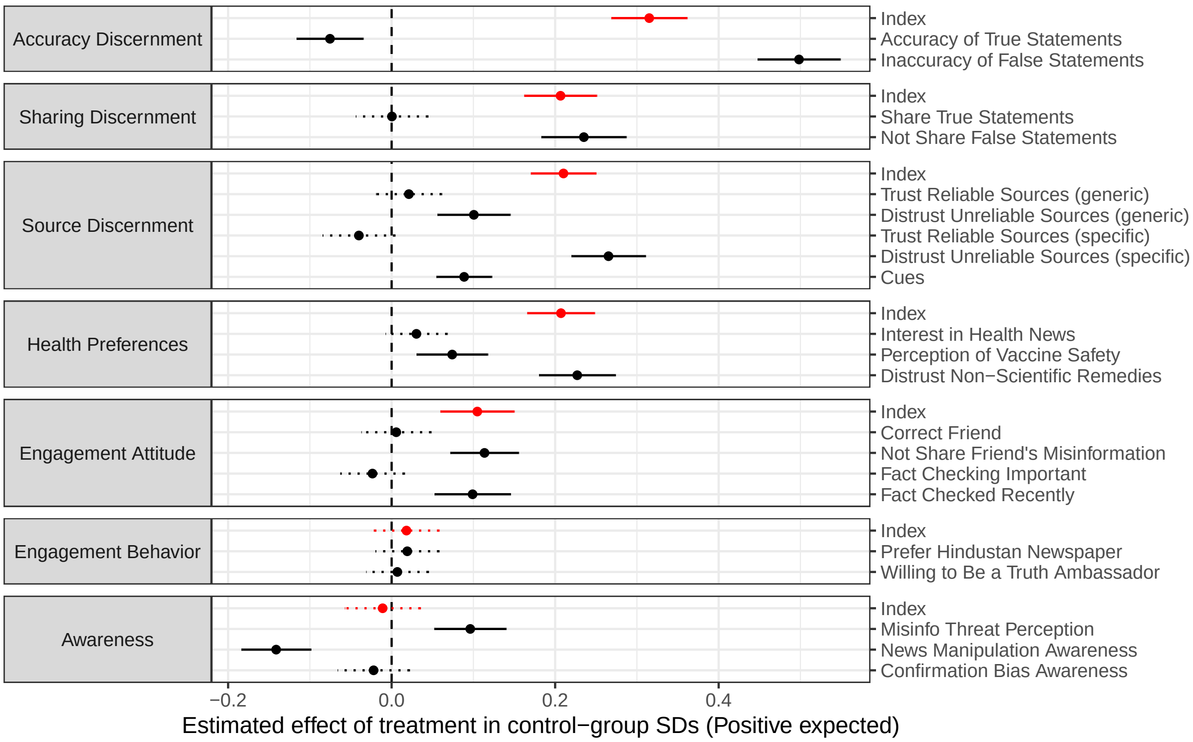
where Y_{ijk} is the primary outcome of interest Y for student i in classroom j and library-spillover strata $k \in \{1, ..., m\}$, β_0 is the intercept, T_{ijk} is a treatment indicator, α_c denotes the coefficient for the control variable X_c, γ_k denotes fixed effects for each library-spillover strata k, and ε_i denotes the random error term for individual i. β_1 denotes the estimated effect of treatment assignment (ITT) on outcome Y. To estimate this equation, we use linear regression with heteroskedasticity-robust standard errors, clustered at the village level. To complement the ITT analysis, we also estimate complier average causal effects (CACE).[15]

First Endline

This section examines the effect of BIMLI on outcomes from the first endline survey. Our main results are summarized in Figure 3, which shows the estimated effect of assignment to treatment on seven outcomes. The estimates of treatment effects we present in Figure 3 can be seen as conservative because of dilution due to partial non-compliance, so we additionally compute the causal effect among compliers (Appendix J). Next we show in Figure 4 an illustration of the distribution of each index across treatment and control groups. Finally we also compute the treatment effect in % of control-group means to offer a simplified summary of treatment effects across outcome domains, useful for descriptive reporting (Appendix G).

[15] See Appendix J for CACE specification.

Figure 3: Estimated effect of assignment to BIMLI treatment



Notes: This figure plots the estimated ITT effect of assignment to BIMLI for 7 outcome families. Each index is an ICW calculation of components within an outcome family. Each component is standardized relative to the control mean and SD. Confidence intervals are at the 95% level and are based on standard errors clustered at the village (classroom) level. Tabular results are in Appendix G.

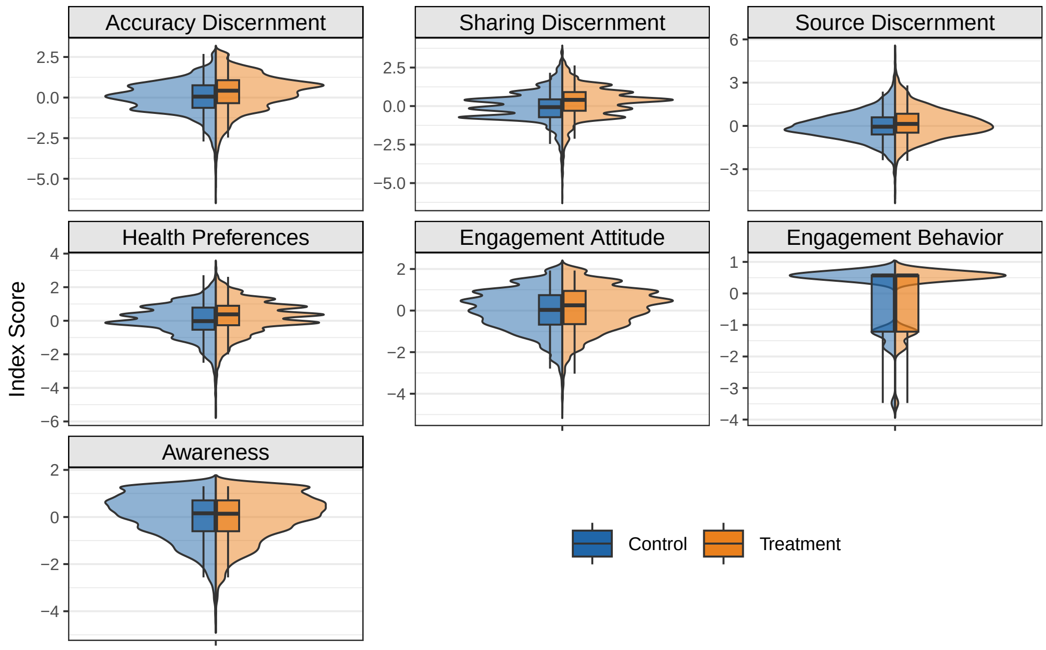
Accuracy and sharing discernment

Recent years have seen a growing consensus on testing the efficacy of misinformation interventions through measuring discernment between true and false information. This approach involves (1) rating a mix of true and false content and (2) analyzing ability to discern between them (Guay et al., 2023). Following this, we asked respondents to rate the perceived accuracy of 8 veracity-balanced news stories on a 4-point scale. Importantly, only 2 of these stories were discussed in class, while 6 were new, meaning that any discernment effects we detect reflect skill application rather than mere recall.[16] We also measured sharing intention using the same items.[17]

[16]We re-estimate effects dropping the 2 items discussed in class and find that results hold (Appendix J4).

[17]Since some previous work has shown that thinking about the accuracy of a story can affect intentions to share (Pennycook et al., 2021), we randomized the order of the sharing and accuracy discernment battery such that one half of the sample is asked each set of questions first.

Figure 4: Distribution of outcome indices, by treatment group



Notes: Each half-violin shows the distribution of standardized, inverse-covariance weighted outcome indices by treatment group. Scores are scaled in units of the control group standard deviation, and higher values reflect more desirable outcomes. Boxplots indicate the interquartile range and median within each group. Note that the Engagement Behavior index consists of only two items, and responses are heavily skewed, with the majority of participants selecting the maximum value, which explains the asymmetric distribution.

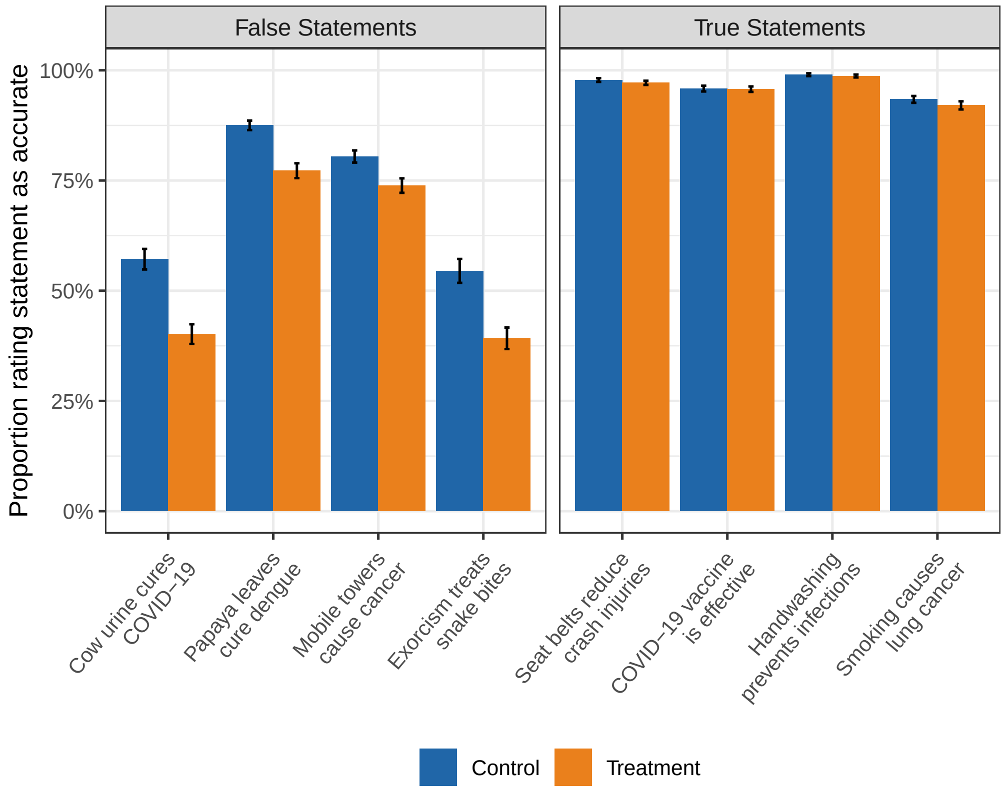
The selection for these stories was based on extensive fieldwork and piloting to identify the most commonly believed health-related myths, each debunked by at least one fact-checking service in India. Stories were presented to respondents in random order.

With respect to accuracy discernment, ITT estimates show that the treatment significantly helped respondents discern between true and false stories (Figure 3). The magnitude of this effect, a 0.32 SD increase in discernment relative to the control group, is substantively large compared to effects from comparable contexts.[18] Further, when we compare ITT to CACE estimates, we find that the effect on accuracy discernment is even larger among compliers.

We do see variation in the true and false components of the discernment measure. In

[18]For example, Guess et al. (2020) find that their digital literacy intervention in India led to a 0.11 SD increase in discernment, while Gottlieb, Adida, and Moussa (2022)'s intervention in Cote d'Ivoire produced effect sizes of 0.12 to 0.15 SD. We note that larger and more comparable effect sizes to ours tend to emerge from interventions that are more intensive and of longer duration (for example see Bowles et al. (2023) where the intervention was 6 months long) lending further support to our argument that sustained exposure and iterative learning are more effective in shifting outcomes.

Figure 5: Accuracy discernment by headline



Notes: The figure displays the average share of respondents in the treatment and control groups who rated each headline as either "very accurate" or "somewhat accurate" (coded as 1), as opposed to "not very accurate" or "not at all accurate" (coded as 0).

Figure 5 we graph perceived accuracy by individual headlines and find that large proportions of respondents believed falsehoods, and the treatment significantly decreased respondents' perceived accuracy of all 4 false stories, with effect sizes ranging from 0.44 SD (cow urine can cure covid) to 0.18 SD (mobile phone towers cause cancer). With respect to true stories, there is little variation in how treatment and control group respondents rated these stories; on average all respondents were better at discerning true stories relative to false.

We find that even if the overall discernment effect is a net positive, the treatment made respondents marginally more skeptical of all news. However, we do not view this as normatively problematic in this context. The baseline tendency among our sample is to trust nearly all information – true and false alike. Further, India has a media environment where misinformation is frequently disseminated by mainstream sources, not just fringe or anonymous actors, and so encouraging some level of critical scrutiny may be both necessary and desirable. To illustrate, the media coverage during the 2025 India–Pakistan conflict had several prominent Indian news outlets broadcast unverified or doctored footage. Videos from video games were aired as real combat footage, and fabricated stories about airstrikes and casualties emanated straight from reputed sources (Das and KB, 2025). In such an environment, the risk is not that people are too skeptical, it is that they are too trusting of information from sources that are not credible. In light of this, we believe that a slight increase in skepticism, even toward some true statements, is a reasonable tradeoff for improved overall discernment.

Empirically, we note that the apparent increase in skepticism on true items should be interpreted with caution: As Figure 5 shows, belief in true statements was already near ceiling at baseline. This limited variance inflates the standardized effect size, giving the impression of a stronger change than is actually the case. In absolute terms, the decrease is small. Finally, our second endline yields null effects on discernment measures for true information (Table 3). This suggests that while the intervention's positive effect on reducing belief in false information persists over time, the temporary decrease in belief in true information is no longer detectable in the follow-up.

With respect to sharing discernment, we find that the treatment has a large and significant effect (0.21 SD). Overall our results on discernment confirm that the treatment was successful at helping respondents prioritize accuracy when believing content as well as sharing it. That we are able to detect effects on stories that were not discussed in the classroom demonstrates a crucial learning component that treated respondents were able to glean from the program. Further, unlike previous studies on misinformation that measure outcomes immediately after treatment, or even as part of the same instrument, given the gap between classroom sessions and the endline survey we can be confident that recall or demand effects are not primarily driving this finding.

Trust in sources and source discernment

To complement accuracy discernment, we introduced measures to evaluate how respondents assess and trust news sources. Recognizing that individuals rarely encounter headlines without accompanying source cues, we incorporated three measures focusing on news sources, including both mediums of news (e.g., platforms, mass media) and the transmitters of news through these mediums. Our approach includes a novel focus on informal sources, such as word-of-mouth and local elites, which are heavily relied upon in our study context (Gadjanova, Lynch, and Saibu, 2022).

First, we measure general source discernment by asking respondents to rate their trust of transmitters (e.g., word of mouth), mediums (e.g., radio, Facebook), and institutions (e.g., the WHO). The index includes three sources we expect to increase trust in (MBBS doctors, healthcare workers, government health notices) and three we hope to decrease trust in (ayurvedic doctors, unqualified practitioners, and word of mouth/rumors). Next, we assess situation-specific trust by using a vignette where respondents seek emergency advice for a sick family member and could go to a number of sources. We provide three trustworthy sources (community health center, government materials, TV doctors) and three untrustworthy ones (family myths, WhatsApp forwards, TV interviews with ayurvedic doctors). This helps distinguish between general and situation-specific trust and separates transmitters from mediums. Finally, we explore which factors foster trust in specific pieces of information, examining whether reliance on signals like likes/shares online, shared ethnicity, as well as message tone and emotionality reduces due to the treatment. Our results show that BIMLI, overall, significantly changed how respondents interact with and trust sources for the better, with a notable shift in the index (SD = 0.21).

Health preferences

We measured health preferences through three components: interest in health news, vaccine safety perceptions, and reliance on alternative medicine. Respondents rated their interest in health news on a scale from very interested to not interested. For vaccine safety, they rated the safety of both the covid and measles vaccines. To assess reliance on alternative medicine, respondents were asked if they would visit traditional healers and unqualified practitioners, or use home remedies for serious illnesses, and whether they agreed that Ayurveda and homeopathy could cure serious diseases.

Despite the prevalence of health misinformation and reliance on alternative medicine in our context, we show that BIMLI was able to significantly alter respondents' health preferences

(0.21 SD). Item-wise results indicate that the treatment reduced vaccine hesitancy and stated reliance on alternative forms of medicine. This finding holds significance: traditional home remedies and the misinformation surrounding them have long existed in India, passed down through generations, suggesting that these beliefs may be deeply ingrained and therefore more resistant to change. Additionally, prior research has indicated that belief in medical misinformation in India is associated with social identities such as religion and partisanship, and given that these identities underpin enduring societal divisions (Chauchard and Badrinathan, 2025), motivated reasoning may impede the effectiveness of misinformation countermeasures (Taber and Lodge, 2006). Despite this, BIMLI had a significant impact on altering health preferences.

Engagement with misinformation countermeasures

We assessed engagement with misinformation countermeasures using attitudinal and behavioral measures. Attitudinally, we focused on shifting norms around misinformation through four self-reported measures: (1) likelihood of correcting a friend sharing misinformation, (2) likelihood of personally sharing misinformation from friends, (3) perceived importance of verifying information, and (4) frequency of verifying information in the past two months. The treatment significantly influenced respondents' attitudes on this index, but we observed variation across items. Treated respondents were more likely to abstain from sharing misinformation, even from close acquaintances, but were hesitant to correct it, reflecting cultural norms in India that may discourage direct confrontation (Malhotra and Pearce, 2022). While respondents hesitated to correct friends, the shift toward not sharing misinformation suggests that the treatment was effective in shifting norms in this context. We also see that there is no effect on perceived importance of fact-checking but a positive effect on frequency of fact-checking. This likely reflects ceiling effects: views on the importance of fact-checking were already extremely high in the control group (82% agree), leaving little room for upward movement. In contrast, the self-reported frequency of fact-checking measure exhibited far more variation across response options.

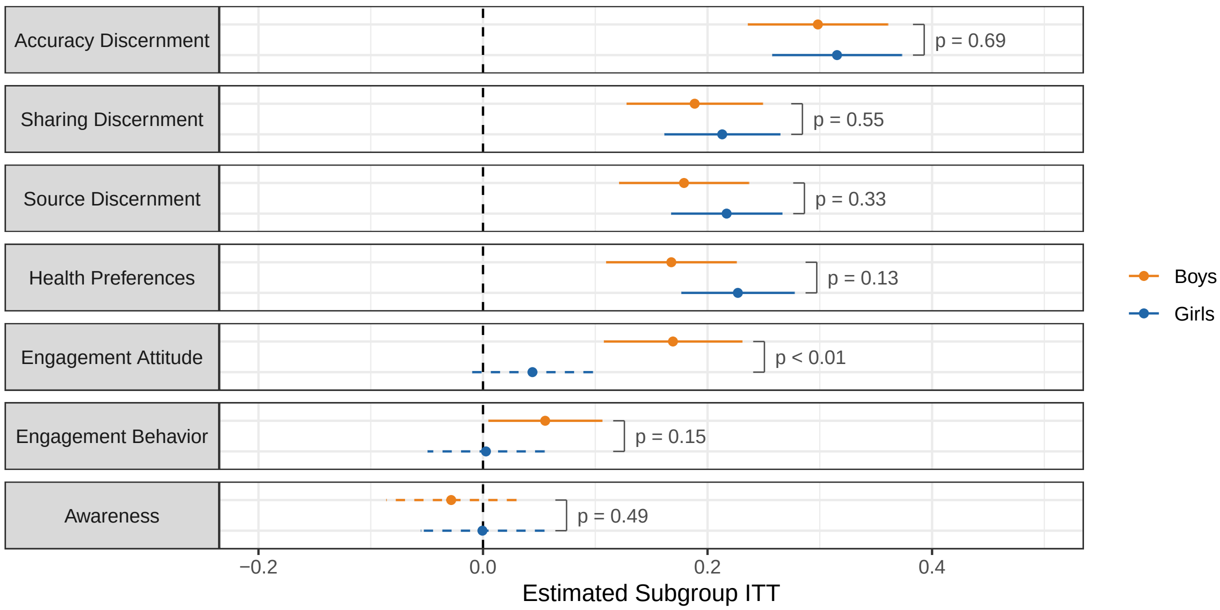
Children in India are accustomed to tests and often excel in educational settings. To ensure our findings were not solely driven by this familiarity, we incorporated two behavioral measures. First, respondents entered a lottery to choose between two subscriptions: a credible Hindi newspaper, *Hindustan*, or a popular entertainment magazine, *Manohar Kahaniyan*. We hypothesized greater demand for news among the treatment group. Second, we invited respondents to become "truth ambassadors," a community role described as supporting local government by dispelling misinformation during crises, described as costly in terms of time and effort. We expected higher willingness for this role in the treatment group. ITT results showed no significant impact on these behaviors, with the overall index a null effect.

However, the overall null effect on the ITT estimate masks significant gender variation. Analyzing ITT by gender subgroup shows differences in misinformation engagement measures, even though indices for other outcomes show no such variation (Figure 6). Boys are significantly more likely to report intentions to engage in misinformation countermeasures, both in attitudes and behaviors, while the treatment had no effect on girls. Breaking down this result further, control group means for boys are much higher than for girls for both indices. Although point estimates are positive for both groups, boys demonstrate a steeper increase, indicating that updating on these indices is concentrated among those already more amenable to such behaviors (Appendix H2). This result aligns with India's patriarchal context, where strong gender norms condition behavior (Brulé, 2020; Prillaman, 2023; Heinze, Brulé, and Chauchard, 2024). Our indices of behaviors and intentions reflect not only measures on misinformation but also the capacity and willingness to engage in community-based actions, which may require shifts in gender norms (e.g. permission for women to engage publicly) and public safety. For instance, correcting a friend's misinformation demands assertiveness and confrontation, traits not directly targeted by the intervention and particularly challenging to change for women in India. While both girls and boys improved equally in discernment, behavioral changes proved harder where cultural and gender norms created barriers. This suggests that while private preferences can be shifted for all, public behaviors improved only among boys. Achieving similar changes among girls may require interventions that address societal norms alongside misinformation.

Awareness

Overall we find a null effect on the awareness index. This index assessed awareness of misinformation and recall of classroom material through five items. The first measured perceptions of

Figure 6: Effect of assignment to BIMLI by gender subgroup



Notes: This figure plots the effect of BIMLI for 7 outcome families with ITT coefficients by gender subgroup. Each index is an ICW calculation of components within an outcome family. Each component is standardized relative to the control mean and SD. Confidence intervals at the 95% level are based on standard errors clustered at the village (classroom) level. P-values indicate the significance of the difference between boys and girls coefficients.

misinformation as a threat. While exposure to BIMLI significantly increased this perception, 78% of respondents were already self-reporting misinformation as a threat, limiting room for further change. Awareness of media and cognitive biases was measured using 4 items adapted to the Indian context from Ashley, Maksl, and Craft (2013). These items focused on defining theoretical classroom concepts, and we find no improvement for treated respondents compared to control (p = 0.64). This could be due to (1) the time gap between lessons and the survey: biases were introduced in session 2, at least two months before the endline, (2) the curriculum's focus on application rather than rote learning, and (3) the complexity of these theoretical concepts. Despite this, we underscore that the significant effects on discernment and other outcomes suggest respondents were able to successfully retain and apply skills learned in the classroom, even if they were unable to recall theoretical definitions of concepts.

Heterogeneous treatment effects

We look at heterogeneous effects analyses based on a number of variables. Most importantly, to proxy motivated reasoning, we examine interaction effects with partisan identity. While direct questions about party ID were not permitted in the baseline survey due to our collaboration with the government, we estimated household-level partisanship through additional surveys with village-level local elites. We surveyed 1,664 elites across 550 villages and asked questions on sub-caste category-wise party preferences in recent elections. Matching this data back to our baseline, we were able to estimate party ID at the household level.[19] We also analyzed heterogeneous effects for household mobile internet access as previous work indicates that prior exposure to media and the internet can influence how individuals interact with misinformation (Guess et al., 2023). Demographically, we examined socio-economic status, age, gender, caste, and religion. We also looked at basic science knowledge. The results, detailed in Appendix H, show no consistent patterns. Aside from the gender subgroup effects discussed earlier, we found no systematic interaction effects for any demographics, including partisan identity. This is notable, as past research suggests that partisanship often moderates the impact of misinformation interventions (Flynn, Nyhan, and Reifler, 2017). Our findings indicate that belief change in this context was driven by a model of learning and updating with no obvious pattern of motivated reasoning, consistent with conclusions from Coppock (2023). Finally, we looked at whether results are different as function of being in a high- or low-spillover village.[20] We find that for three outcomes, engagement attitudes, awareness of misinformation and source discernment, assignment to treatment in a low-spillover village positively affects respondents. This is notable especially with regard to the awareness index, as our main effect was a null result.

[19]See Appendix H.1 for notes on party ID estimation.

[20]We note that the number of low-spillover villages increased after randomization at the library level. This was because, during randomization, all villages within a GP occasionally fell into the same treatment group, reducing concerns about spillovers between treatment and control. We re-classified these as low-spillover. Since this post-randomization classification more accurately reflects spillover potential, we use it in heterogeneity models to evaluate whether spillovers affect results. However, since the post-randomization spillover classification is not reflective of the stratified randomization procedure outlined above, our main models use pre-randomization spillover-strata for the library-spillover FEs.

Robustness Checks

To test the robustness of our results, we undertake several analyses. First, we re-estimate the baseline model incorporating library fixed effects, district fixed effects, and district-spillover stratum fixed effects. The main results remain unchanged. Second, we run an adjusted model with pre-registered control variables, including demographics (age, gender, grade, caste, religion, language of schooling), household-level variables (asset index as a proxy for income and access to mobile internet), baseline covariates (reading skill and science knowledge indices), and village-level variables (development proxied by nighttime lights data, and partisanship measured by BJP vote share in the last assembly election). Results are robust to these controls. Following this, we apply multiple-hypothesis test corrections across indices, as pre-registered. Results on our main dependent variables remain significant. Next, to address concerns around parental presence prompting respondent answers, we conducted subgroup ITT analyses based on the number of individuals present during the interview and find that results hold regardless of parent/guardian presence. All these results are reported in Appendix J. Finally, to exclude the possibility that our results are driven by differential attrition between treatment and control based on unobservables, we undertake sensitivity analyses using a tipping point method, inverse probability weighting, and Lee bounds (Appendix K3).

Second Endline

We conducted a follow-up survey with a random subset of 2,059 respondents approximately four months after the intervention to assess its long-term effects.[21] The extended time gap is particularly relevant, as India's 2024 general elections occurred between our two endlines – a period when political and partisan attitudes typically become more salient (Michelitch and Utych, 2018). The follow-up had three main objectives: (1) to assess whether discernment capacity persisted over time, (2) to evaluate if respondents could apply this skill to political stories - a new and unrelated domain, as the intervention deliberately avoided political topics due to our collaboration with the government, and (3) to examine within-household treatment diffusion to

[21] Appendix I describes sampling for the second endline; attrition & compliance are discussed in Appendix K.

untreated members. To measure this, we interviewed one randomly selected parent or guardian for each of the follow-up households.[22]

Remarkably, our findings indicate that participants in the treatment group continued to exhibit an improved ability to discern truth from falsehood (0.26 SD), as shown in Table 3. Moreover, treated respondents exhibited a significantly higher capacity to accurately assess the veracity of political stories (0.31 SD). This result is striking given that the intervention focused solely on health content and did not address political claims. The political stories were entirely new narratives that went viral during the 2024 election, and were introduced only in the second endline. Yet treated respondents showed improved ability to distinguish true from false political information. This suggests they were not just recalling content but applying learned principles across domains. The findings highlight that even when narrowly focused on a specific topic (such as health), educational interventions can yield transferable benefits across other domains.[23]

Finally, we find that parents/guardians of treated students were significantly better at discerning true from false health information (0.27 SD), as demonstrated in Table 4. This result is particularly notable as it highlights the potential for "trickle-up" socialization, where children's learning influences their parents (Dahlgaard, 2018).[24] It also suggests that sustained learning may generate valuable within-network diffusion effects. One mechanism for this effect may have been the homework assignments and handouts given to students. Both treatment and control groups received written materials summarizing classroom lessons to take home (see Appendix B). Students worked on assignments at home and had physical copies of handouts and fliers that family members could potentially view or discuss with them. We view this finding as noteworthy, underscoring that educative interventions can have effects that transfer to other important members of networks, thereby adding to a literature that identifies change in adults that stem from children's behaviors (Carlos, 2021; McDevitt and Chaffee, 2002; Washington, 2008).

[22] See Appendix D for survey items on political discernment.

[23] We note that we observe very limited differences between the random follow-up sample that we recontacted versus those who eventually answered, implying that the persistence we observe likely generalizes to the whole sample (see Appendix K).

[24] We note that we are only able to robustly detect diffusion of treatment effects to guardians for health accuracy discernment outcomes. Please see Appendix I for results on other outcomes from the guardian survey.

Table 3: Effect of assignment to BIMLI treatment on 4-month follow-up

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**
Accuracy Discernment (health)	Index	1,945	0.26***	0.048	<0.001
Accuracy of True Statements	Sub-index	2,053	-0.06	0.040	0.14
Inaccuracy of False Statements	Sub-index	1,962	0.33***	0.041	<0.001
Accuracy Discernment (politics)	Index	1,863	0.31***	0.049	<0.001
Accuracy of True Statements	Sub-index	1,992	-0.01	0.041	0.88
Inaccuracy of False Statements	Sub-index	1,887	0.31***	0.043	<0.001
Source Discernment	Index	2,028	0.10*	0.044	0.03
Trust Reliable Sources	Sub-index	2,040	-0.07	0.042	0.08
Distrust Unreliable Sources	Sub-index	2,056	0.14***	0.041	<0.001

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

Table 4: Effect of assignment to BIMLI on treatment group parents/guardians

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**
Accuracy Discernment Index	Index	1,786	0.27***	0.054	<0.001
Accuracy of True Statements	Sub-index	2,020	-0.01	0.047	0.88
Inaccuracy of False Statements	Sub-index	1,804	0.28***	0.049	<0.001

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

Discussion and Conclusion

In this study, we evaluated the impact of a large-scale, classroom-based intervention aimed at combating misinformation, implemented among over 13,500 adolescents in Bihar, India. In collaboration with a state government agency, we developed a curriculum of sustained education against misinformation that spanned 4 months. ITT estimates showed significant improvements on several outcomes. By the program's end, treated respondents demonstrated better discernment in evaluating and sharing information, shifted health preferences away from alternative medicine, and enhanced source credibility assessments. We also detected effects on behavioral measures among boys. These effects persisted among a sub-sample interviewed 4 months later. Importantly, follow-up surveys showed that students were able to accurately discern true from false political news, a topic not covered in the program, demonstrating the transferability of the acquired skills. Finally, we found that parents/guardians of treated students were significantly better at discernment, indicating that such educational interventions can have additional effects within social networks, with knowledge trickling upwards through socialization. Several of the outcomes we measure evaluate the acquisition of skills rather than mere recall, reducing the possibility that expressive responding or social desirability alone drove responses.

These findings are significant given the mixed or null results typically seen in media literacy interventions (Blair et al., 2023). In contrast, our program produced measurable effects in a particularly challenging environment. Bihar, where the study was conducted, has low educational prioritization and a 42% dropout rate before 10th grade (Muralidharan and Prakash, 2017). Session compliance in our study reached roughly 70%, a respectable figure given the region's limited state capacity and consistent underperformance on public service delivery (Desai, 2019; Rasul and Sharma, 2014; Mathew and Moore, 2011; Jha, 2023). Thus, it was not obvious that a media literacy curriculum like BIMLI would yield positive effects; to our knowledge, this is the first intervention in this context to produce significant belief change in misinformation outcomes. These results suggest that more intensive strategies, featuring peer learning, norm-setting, and repeated exposure, may be essential for meaningfully shifting entrenched beliefs, especially where one-off informational interventions have failed.

Despite these encouraging findings, we acknowledge several limitations of the study. First, the intervention was delivered as a bundled, high-dosage program with multiple components, making it difficult to isolate which elements (content, dosage, or delivery format) were most effective, or to tease out mechanisms. Session-wise attendance is not a reliable proxy for variation, as session topics are confounded with peer effects; students attending earlier sessions may form social networks that generate endogenous downstream effects. Moreover, the curriculum involved substantial repetition, with each session revisiting earlier material, further complicating efforts to identify topic-specific impact. Our goal was to design a comprehensive intervention to address the limited success of prior media literacy programs, but future research could unbundle the curriculum to assess which elements drive results. A second limitation concerns cost and scalability: implementing such a sustained program required substantial resources. Due to budget and power constraints, we were unable to experimentally vary treatment dosage, but future work could test the minimum intensity required to produce effects.[25] Another important dimension for future exploration involves delivery format. Our treatment combined

[25]We can, in theory, look at subgroup IIT by session attendance. When we do this we find that it takes at least 2 sessions to produce any effects and for most outcomes 3 sessions, but note that this analysis is biased because attendance beyond the first session is non-random and downstream from the treatment.

content, peer learning, and instruction from teachers trained to incorporate interaction and discussion. It is unclear whether the same syllabus, delivered via online modules without peer interaction or a teacher, would yield similar results. Disentangling the roles of content, authority, and peer dynamics will be critical for informing scalable and effective policy design in the future. Lastly, we acknowledge some design limitations. First, we were only able to implement survey-based behavioral measures and behavioral intentions, rather than tracking actual behaviors, due to several logistical constraints. One meaningful outcome we would have liked to track post-treatment is actual vaccine uptake. However, challenges in accessing administrative data and tracking respondents over time made this infeasible. Second, since we had to hire teachers for treatment and control from separate pools, we were unable to implement teacher-level fixed effects to determine if outcomes changed due to teacher quality.

Finally, we reflect on the generalizability of our results. As noted earlier, our study took place in a low-capacity setting with limited access to credible news and low socioeconomic status. To ensure the intervention's success in this context, we made deliberate design choices such as bringing in external teachers and partnering with a trusted government agency. The program may have been effective in part because it stood out in this context: a rare, high-quality educational opportunity delivered in an engaging style. Supporting this, over 95% of surveyed parents - across both treatment and control groups - said they would enroll their children again in such a program. Among them, over a quarter emphasized their trust in Jeevika as being a reason for interest (Appendix I.2). We thus caution against assuming straightforward generalizability to other contexts that may share surface similarities with Bihar, such as low state capacity, offline information sharing, or low socioeconomic status. While Bihar exhibits these features, we deliberately incorporated design elements to mitigate their impact on learning outcomes, including intensive teacher training focused on interactive pedagogy and incentives to encourage attendance. Without such supports, it is unclear whether similar results would hold in public school systems elsewhere in India or across the Global South. On the other hand, in contexts with similarly engaging educational environments and high institutional trust – such as many settings in the Global North – we see no reason that such an intervention would not work. Moreover, our data show minimal heterogeneity in treatment effects across a number of pre-treatment characteristics, including income, socioeconomic status, religion, caste, and political affiliation, suggesting the intervention could have similar impacts across a range of diverse populations (Appendix H).

Despite these limitations, our positive findings offer valuable insights for both academic research on misinformation and policy development. Following the 2016 surge in media literacy initiatives, many were implemented without evidence of their causal effects. To our knowledge, this is the first randomized controlled trial testing the efficacy of such an intervention. The implications are broad: we believe policy-makers and researchers alike should prioritize sustained, iterative treatments. In many settings, these may be the only viable solutions, especially where populations lack internet access, making platform-based solutions like fact-checking unfeasible. From a policy perspective, modules like ours could be integrated seamlessly into school curricula, particularly in contexts with high educational quality. Finally, we note that after undertaking cost-effectiveness calculations under several assumptions, we find that our intervention can be delivered for approximately $4.84 per student under a full-cost model, and for under $1 per student when using existing public school teachers and excluding one-time startup expenses such as curriculum development costs. Overall, we estimate that the program was successful in shifting the median student from the 50th to the 61st percentile of the control-group distribution, highlighting its scalability and cost-effectiveness despite its dosage intensity (Appendix M).

We attribute these hopeful findings to the setting in which we fielded the study: classrooms and schools have consistently been identified as pivotal sites for knowledge acquisition beyond the household, and public education systems play a crucial role as agents of socialization, especially in contexts where information spread takes place offline. Therefore, our study not only contributes to the literature on persuasion and information processing but also examines the enduring impacts of education and learning. This aligns with existing work exploring the transformative potential of education within schools, investigating education to reshape gender attitudes in India (Dhar, Jain, and Jayachandran, 2022) and foster nation-building efforts (Bandiera et al., 2019), along with the potential of interaction with the state via education to shape economic views (Davies, 2023). Further, scholars have explored the efficacy of educational tools such as textbooks in persuasion and attitude change (Cantoni et al., 2017), as well as their role in shaping perceptions of representation and marginalization (Haas and Lindstam, 2023). By situating our study within the broader context of educational interventions, we contribute to scholarly understanding of the multifaceted impacts of schooling on attitudes and behaviors.

Data Availability Statement: Research documentation and data that support the findings of this study are openly available at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/G2BK1G.

Acknowledgments and Notes: Authors listed alphabetically. We are grateful to Jeevika and to IAS officers Rahul Kumar and Himanshu Sharma, as well as Pushpendra Singh Tiwari and Rakesh Kumar, for making the Bihar government's collaboration on this project possible. Our implementation partner is DataLeads in New Delhi, India. Our survey partner is Sunai Consultancy in Patna, India. We received IRB approval from American University and the project was reviewed by the UC3M's GDPR data protection officer. We thank Josiah Gottfried for outstanding research assistance, and Minati Chaklanavis, Pranav Chaudhary and Nishant Attri for on-field support. Pre-registration available at OSF. For comments and feedback we are grateful to Danny Choi, Austin Davis, Maria Hernandez de Benito, Andy Guess, Don Green, Erik Kramon, Horacio Larreguy, Gareth Nellis, Laura Paler, Casey Petroff, Laura Schechter, and seminar and workshop participants at George Washington, Penn, American, Rochester, Brown, Michigan, Sciences Po, UMontreal, Columbia, Humboldt, University of Oslo, IC3JM, and APSA and ICA conferences.

Funding Statement: This research was funded by the Social Science Research Council under their Mercury Project.

Conflict of Interest: The authors declare no ethical issues or conflicts of interest in this research.

Ethical Standards: The authors declare the human subjects research in this article was reviewed and approved by American University (Protocol IRB-2024-5) and by the UC3M's GDPR data protection officer and certificates are provided in the Dataverse. The authors affirm that this article adheres to the principles concerning research with human participants laid out in APSA's Principles and Guidance on Human Subject Research (2020).

References

Ali, Ayesha, and Ihsan Ayyub Qazi. 2023. "Countering misinformation on social media through educational interventions: Evidence from a randomized experiment in Pakistan." *Journal of Development Economics* 163: 103108.

Alsan, Marcella, and Sarah Eichmeyer. 2024. "Experimental evidence on the effectiveness of nonexperts for improving vaccine demand." *American Economic Journal: Economic Policy* 16 (1): 394–414.

Andrews, Isaiah, and Jesse M Shapiro. 2021. "A model of scientific communication." *Econometrica* 89 (5): 2117–2142.

Apuke, Oberiri Destiny, Bahiyah Omar, and Elif Asude Tunca. 2023. "Literacy concepts as an intervention strategy for improving fake news knowledge, detection skills, and curtailing the tendency to share fake news in Nigeria." *Child & Youth Services* 44 (1): 88–103.

Ashley, Seth, Adam Maksl, and Stephanie Craft. 2013. "Developing a news media literacy scale." *Journalism & Mass Communication Educator* 68 (1): 7–21.

Badrinathan, Sumitra. 2021. "Educative Interventions to Combat Misinformation: Evidence From a Field Experiment in India." *American Political Science Review* 115 (4): 1325–1341.

Badrinathan, Sumitra, and Simon Chauchard. 2023a. "Researching and countering misinformation in the Global South." *Current Opinion in Psychology*: 101733.

Badrinathan, Sumitra, and Simon Chauchard. 2023b. ""I Don't Think That's True, Bro!" Social Corrections of Misinformation in India." *The International Journal of Press/Politics*: 1–16.

Badrinathan, Sumitra, Simon Chauchard, and Niloufer Siddiqui. 2024. "Misinformation and support for vigilantism: An experiment in India and Pakistan." *American Political Science Review*: 1–19.

Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo. 2019. "Nation-building through compulsory schooling during the age of mass migration." *The Economic Journal* 129 (617): 62–109.

Banerjee, Abhijit, Donald P Green, Jeffery McManus, and Rohini Pande. 2014. "Are poor voters indifferent to whether elected leaders are criminal or corrupt? A vignette experiment in rural India." *Political Communication* 31 (3): 391–407.

Bhattacharya, Usree. 2022. ""I Am a Parrot": Literacy Ideologies and Rote Learning." *Journal of Literacy Research* 54 (2): 113–136.

Blair, Robert, Jessica Gottlieb, Brendan Nyhan, Laura Paler, Pablo Argote, and Charlene Stainfield. 2023. "Interventions to Counter Misinformation: Lessons from the Global North and Applications to the Global South." *USAID Report*.

Boas, Taylor C, and F Daniel Hidalgo. 2011. "Controlling the airwaves: Incumbency advantage and community radio in Brazil." *American Journal of Political Science* 55 (4): 869–885.

Bode, Leticia, and Emily K Vraga. 2018. "See something, say something: Correction of global health misinformation on social media." *Health Communication* 33 (9): 1131–1140.

Bowles, Jeremy, Kevin Croke, Horacio Larreguy, Shelley Liu, and John Marshall. 2023. "Sustaining exposure to fact-checks: Misinformation discernment, media consumption, and its political implications." *American Political Science Review*: 1–24.

Brashier, Nadia M, and Daniel L Schacter. 2020. "Aging in an era of fake news." *Current directions in psychological science* 29 (3): 316–323.

Brass, Paul R. 2011. *The production of Hindu-Muslim violence in contemporary India*. Seattle, WA: University of Washington Press.

Bridgman, Aengus, Eric Merkley, Peter John Loewen, Taylor Owen, Derek Ruths, Lisa Teichmann, and Oleg Zhilin. 2020. "The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media." *Harvard Misinformation Review*.

Brulé, Rachel E. 2020. *Women, Power, & Property*. Cambridge: Cambridge University Press.

Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2017. "Curriculum and ideology." *Journal of Political Economy* 125 (2): 338–392.

Carlos, Roberto F. 2021. "The politics of the mundane." *American Political Science Review* 115 (3): 775–789.

Cavaille, Charlotte, and John Marshall. 2019. "Education and anti-immigration attitudes: Evidence from compulsory schooling reforms across Western Europe." *American Political Science Review* 113 (1): 254–263.

Chauchard, Simon, and Sumitra Badrinathan. 2025. "The Religious Roots of Belief in Misinformation: Experimental Evidence from India." *Forthcoming, British Journal of Political Science*.

Cheema, Ali, Sarah Khan, Asad Liaqat, and Shandana Khan Mohmand. 2023. "Canvassing the gatekeepers: A field experiment to increase women voters' turnout in Pakistan." *American Political Science Review* 117 (1): 1–21.

Clayton, Katherine, Spencer Blair, Jonathan A Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, Akhila Kovvuri, Jonathan Martin, Evan Morgan et al. 2019. "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media." *Political Behavior*: 1–23.

Coppock, Alexander. 2023. *Persuasion in parallel: How information changes minds about politics*. Chicago: University of Chicago Press.

Dahlgaard, Jens Olav. 2018. "Trickle-up political socialization: The causal effect on turnout of parenting a newly enfranchised voter." *American Political Science Review* 112 (3): 698–705.

Das, Anupreeta, and Pragati KB. 2025. "How the Indian Media Amplified Falsehoods in the Drumbeat of War." *The New York Times*. May 17, 2025. https://www.nytimes.com/2025/05/17/world/asia/india-news-media-misinformation.html.

Davies, Emmerich. 2023. "The Lessons Private Schools Teach: Using a Field Experiment To Understand the Effects of Private Services on Political Behavior." *Comparative Political Studies* 56 (6): 824–861.

de Barros, Andreas, Johanna Fajardo-Gonzalez, Paul Glewwe, and Ashwini Sankar. 2022. "Large-scale efforts to improve teaching and child learning: Experimental evidence from India." *Journal of Development Economics*.

Desai, Sonalde et al. 2019. *Human Development in India: Challenges for a Society in Transition*. Oxford: Oxford University Press.

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2022. "Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India." *American Economic Review* 112 (3): 899–927.

Druckman, James N, and Kjersten R Nelson. 2003. "Framing and deliberation: How citizens' conversations limit elite influence." *American Journal of Political Science* 47 (4): 729–745.

Fazio, Lisa K, David G Rand, and Gordon Pennycook. 2019. "Repetition increases perceived truth equally for plausible and implausible statements." *Psychonomic Bulletin & Review* 26: 1705–1710.

Flynn, D.J., Brendan Nyhan, and Jason Reifler. 2017. "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics." *Political Psychology* 38: 127–150.

Gadjanova, Elena, Gabrielle Lynch, and Ghadafi Saibu. 2022. "Misinformation across digital divides: Theory and evidence from northern Ghana." *African Affairs* 121 (483): 161–195.

Ghosh, Arkadev, Prerna Kundu, Matt Lowe, and Gareth Nellis. 2025. "Creating cohesive communities: A youth camp experiment in india." *Review of Economic Studies*: 1–26.

Gottlieb, Jessica. 2016. "Greater expectations: A field experiment to improve accountability in Mali." *American Journal of Political Science* 60 (1): 143–157.

Gottlieb, Jessica, Claire L Adida, and Richard Moussa. 2022. "Reducing Misinformation in a Polarized Context: Experimental Evidence from Côte d'Ivoire." *Working Paper*.

Graham, Matthew H, and Omer Yair. 2025. "Less partisan but no more competent: Expressive responding and fact-opinion discernment." *Public Opinion Quarterly* 89 (1): 7–30.

Green, Donald P, Dylan W Groves, Constantine Manda, Beatrice Montano, and Bardia Rahmani. 2024. "The Effects of Independent Local Radio on Tanzanian Public Opinion: Evidence from a Planned Natural Experiment." *The Journal of Politics* 86 (1): 231–240.

Guay, Brian, Adam J Berinsky, Gordon Pennycook, and David Rand. 2023. "How to think about whether misinformation interventions work." *Nature Human Behaviour* 7 (8): 1231–1233.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less than you think: Prevalence and predictors of fake news dissemination on Facebook." *Science Advances* 5 (1).

Guess, Andrew M, and Benjamin A Lyons. 2020. "Misinformation, disinformation, and online propaganda." *Social Media and Democracy* 10.

Guess, Andrew M, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 2020. "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India." *Proceedings of the National Academy of Sciences* 117 (27): 15536–15545.

Guess, Andrew M, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow et al. 2023. "How do social media feed algorithms affect attitudes and behavior in an election campaign?" *Science* 381 (6656): 398–404.

Haas, Nicholas, and Emmy Lindstam. 2023. "My History or Our History? Historical Revisionism and Entitlement to Lead." *American Political Science Review*: 1–25.

Hameleers, Michael. 2020. "Separating truth from lies: comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands." *Information, Communication & Society* 37 (2): 1–17.

Harjani, Trisha, Melisa-Sinem Basol, Jon Roozenbeek, and Sander van der Linden. 2023. "Gamified inoculation against misinformation in India: A randomized control trial." *Journal of Trial and Error* 3 (1): 14–56.

Heinze, Alyssa, Rachel Brulé, and Simon Chauchard. 2024. "Who Actually Governs? Gender Inequality and Political Representation in Rural India." *Forthcoming, The Journal of Politics*.

Huber, Gregory A, and Kevin Arceneaux. 2007. "Identifying the persuasive effects of presidential advertising." *American Journal of Political Science* 51 (4): 957–977.

Jaffrelot, Christophe. 2021. *Modi's India: Hindu nationalism and the rise of ethnic democracy*. Princeton: Princeton University Press.

Jha, Himanshu. 2023. "Pathways to develop state capacity in a weak state: the sub-national state of Bihar in India." *Commonwealth & Comparative Politics* 61 (4): 427–450.

Karlan, Dean, and Jacob Appel. 2016. *Failing in the field: What we can learn when field research goes wrong*. Princeton: Princeton University Press.

Kosec, Katrina, and Leonard Wantchekon. 2020. "Can information improve rural governance and service delivery?" *World Development* 125: 104376.

Kumar, Krishna. 1986. "Textbooks and educational culture." *Economic and Political Weekly*: 1309–1311.

Lieberman, Evan S, Daniel N Posner, and Lily L Tsai. 2014. "Does information lead to more active citizenship? Evidence from an education intervention in rural Kenya." *World Development* 60: 69–83.

Malhotra, Pranav, and Katy Pearce. 2022. "Facing falsehoods: strategies for polite misinformation correction." *International Journal of Communication* 16: 22.

Mathew, Athakattu Santhosh, and M Moore. 2011. "State Incapacity by Design." *Institute of Development Studies Working Paper*.

McDevitt, Michael, and Steven Chaffee. 2002. "From top-down to trickle-up influence: Revisiting assumptions about the family in political socialization." *Political Communication* 19 (3): 281–301.

Michelitch, Kristin, and Stephen Utych. 2018. "Electoral Cycle Fluctuations in Partisanship: Global Evidence from Eighty-Six Countries." *The Journal of Politics* 80 (2): 412–427.

Mohan, Janani. 2021. "Media Bias and Democracy in India." *The Stimson Center*. June 28, 2021. https://www.stimson.org/2021/media-bias-and-democracy-in-india/.

Muralidharan, Karthik, and Nishith Prakash. 2017. "Cycling to school: Increasing secondary school enrollment for girls in India." *American Economic Journal: Applied Economics* 9 (3): 321–350.

Nyhan, Brendan. 2021. "Why the backfire effect does not explain the durability of political misperceptions." *Proceedings of the National Academy of Sciences* 118 (15).

Paglayan, Agustina. 2024. *Raised to Obey: The Rise and Spread of Mass Education*. Princeton: Princeton University Press.

Paluck, Elizabeth Levy, and Hana Shepherd. 2012. "The salience of social referents: a field experiment on collective norms and harassment behavior in a school social network." *Journal of Personality and Social Psychology* 103 (6): 899.

Pearce, Katy E, and Pranav Malhotra. 2022. "Inaccuracies and Izzat: channel affordances for the consideration of face in misinformation correction." *Journal of Computer-Mediated Communication* 27 (2).

Pennycook, Gordon, and David G Rand. 2019. "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning." *Cognition* 188: 39–50.

Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. 2021. "Shifting attention to accuracy can reduce misinformation online." *Nature* 592 (7855): 590–595.

Pereira, Frederico Batista, Natália S Bueno, Felipe Nunes, and Nara Pavão. 2024. "Inoculation reduces misinformation: experimental evidence from multidimensional interventions in brazil." *Journal of Experimental Political Science* 11 (3): 239–250.

Persily, Nathaniel, Joshua A Tucker, and Joshua Aaron Tucker. 2020. *Social media and democracy: The state of the field, prospects for reform*. Cambridge University Press.

Porter, Ethan, and Thomas J Wood. 2019. *False Alarm: The Truth About Political Mistruths in the Trump Era*. Cambridge University Press.

Prillaman, Soledad Artiz. 2023. *The Patriarchal Political Order: The Making and Unraveling of the Gendered Participation Gap in India*. Cambridge University Press.

Ramirez, Francisco O, and John Boli. 1987. "The political construction of mass schooling: European origins and worldwide institutionalization." *Sociology of Education*: 2–17.

Rao, Vijayendra, Kripa Ananthpur, and Kabir Malik. 2017. "The anatomy of failure: An ethnography of a randomized trial to deepen democracy in rural India." *World Development* 99: 481–497.

Rasul, Golam, and Eklabya Sharma. 2014. "Understanding the poor economic performance of Bihar and Uttar Pradesh, India: a macro-perspective." *Regional Studies, Regional Science* 1 (1): 221–239.

Roozenbeek, Jon, Sander Van Der Linden, Beth Goldberg, Steve Rathje, and Stephan Lewandowsky. 2022. "Psychological inoculation improves resilience against misinformation on social media." *Science Advances* 8 (34): eabo6254.

Sen, Somdeep. 2023. "Big money is choking India's free press — and its democracy." *Al Jazeera*. January 6, 2023. https://www.aljazeera.com/opinions/2023/1/6/big-money-is-choking-indias-free-press.

Sharma, Dinesh C. 2015. "India still struggles with rural doctor shortages." *The Lancet* 386 (10011): 2381–2382.

Siddiqui, Danish. 2020. "Hindu group offers cow urine in a bid to ward off coronavirus." *Reuters*. March 14, 2020. https://www.reuters.com/article/us-health-coronavirus-india-cow-urine-pa/hindu-group-offers-cow-urine-in-a-bid-to-ward-off-coronavirus-idUSKBN2110D5.

Sitrin, Carly. 2020. "New Jersey becomes first state to mandate K-12 students learn information literacy." *Politico*. January 1, 2023 https://www.politico.com/news/2023/01/05/new-jersey-is-the-first-state-to-mandate-k-12-students-learn-information-literacy-00076352.

Steenson, Molly Wright, and Jonathan Donner. 2017. "Beyond the personal and private: Modes of mobile phone sharing in urban India." In *The reconstruction of space and time*. Routledge.

Taber, Charles S, and Milton Lodge. 2006. "Motivated skepticism in the evaluation of political beliefs." *American Journal of Political Science* 50 (3): 755–769.

Tankard, Margaret E, and Elizabeth Levy Paluck. 2016. "Norm perception as a vehicle for social change." *Social Issues and Policy Review* 10 (1): 181–211.

Tudor, Maya. 2023. "Why India's Democracy Is Dying." *Journal of Democracy* 34 (3): 121–132.

Tully, Melissa, Emily K Vraga, and Leticia Bode. 2020. "Designing and testing news literacy messages for social media." *Mass Communication and Society* 23 (1): 22–46.

Vraga, Emily K, Leticia Bode, and Melissa Tully. 2022. "The effects of a news literacy video and real-time corrections to video misinformation related to sunscreen and skin cancer." *Health Communication* 37 (13): 1622–1630.

Washington, Ebonya L. 2008. "Female socialization: How daughters affect their legislator fathers' voting on women's issues." *American Economic Review* 98 (1): 311–332.

Wiseman, Alexander W, M Fernanda Astiz, Rodrigo Fabrega, and David P Baker. 2011. "Making citizens of the world: The political socialization of youth in formal mass education systems." *Compare: A Journal of Comparative and International Education* 41 (5): 561–577.

Online Appendix

Countering Misinformation Early:

Evidence from a Classroom-Based Field Experiment in India

(*Amar et al.*)

Contents

A Sampling

To dispense the classes in both treatment and control groups, the Bihar Rural Livelihoods Promotion Society (BRLPS), or Jeevika, made available to us 100 community libraries across 32 districts. Each library is located in a block, with several blocks/libraries in a given district. These libraries are owned and maintained by the BRLPS. Because the locations of classes (libraries) are not movable, we drew our sample from households in villages around libraries. Our unit of randomization was the village. Treatment was administered at the village level (N = 583), with participants clustered within villages having the same treatment status.

A.1 Village selection

To sample villages, we proceeded as follows. Using data collected from BRLPS and triangulating it with other publicly available village-level data, we first identified all villages within a radius of 3 kilometers of a library location. We restricted attention to villages in close proximity to libraries so that respondents could walk or cycle to classes. We then aimed to select 6 villages around each library (we settled on 6 villages around each library for power concerns, see below). To do so, we proceeded in several steps:

1. **First round**: First we selected one village from each gram panchayat (GP) – the administrative unit governing villages (a gram panchayat counts several villages on average) – that lies within a 3 kilometer radius around each library. Within each GP, we first excluded villages with < 100 households. From the remainder, we selected the largest village. Depending on the number of GPs within this radius, this procedure allowed us to select 1-6 villages around a library.

2. **Second round**: If the first round led to the selection of < 6 villages around each library (this was almost always the case), we used Round 2 to select additional villages. In this round, we tried to minimize the number of GPs we drew the rest of the villages from. To do so, we selected the remaining villages from as few different GPs as possible, and ideally from a single GP. Our intuition in doing so was to preserve as many of the villages we sampled in the first round — by definition, only one village per GP — from potential spillovers, noting that spillover potential is higher within a single GP and lower across GPs. As we selected villages for the second round, we first targeted the GP that counted the largest number of selectable villages (within the 3km radius, > 100 households) after the first round, and selected within it villages by population size (starting with the largest). If this did not allow us to complete our sampling of 6 villages around each library, we targeted the second GP that counted the largest number of selectable villages after the first round, and so on.

If there were several GPs with the same number of villages, we randomly selected one of these GPs. If there were two villages with the same total number of households within a GP, we randomly selected one of them. In case there were fewer than 6 villages with these "selectable" characteristics around a library, we selected as many as we were able.[1] Proceeding in this manner, we were ultimately able to select 583 villages around 100 libraries.

Relying on this sample, we were able to create two pre-randomization strata around each library: one with low potential for spillover (villages that were the only sampled village in a GP), and the other with some spillover potential (villages in which we resampled in the second round).[2]

A.2 Randomization

To ensure that the baseline characteristics of villages were similar between treatment arms, we randomly assigned to treatment or control within each library area. Additionally, to ensure that similar numbers of villages were assigned to treatment and control within each spillover category, we stratified by spillover strata. In cases where there was an even number of villages within one library-spillover stratum, we used complete randomization to assign exactly half of the villages to treatment and half to control. For example, in a given library area if there were 4 low-spillover villages, we assigned 2 to treatment and 2 to control using complete randomization. In case of an uneven number of villages within a spillover stratum, the last village was randomly assigned to treatment or control. Due to this, we end up with 5 libraries where the number of treatment villages is either one more ore one less than the number of control villages.

In libraries where there was an uneven number of villages within *both* spillover strata, high and low (this was the case for 25 libraries), we randomly selected one village from the low-spillover stratum and one from the high-spillover stratum and then again used complete randomization to assign one of them to treatment and one of them to control. For example, in libraries with 3 low-spillover and 3 high-spillover villages, we randomly selected one low and one high-spillover village, and then used complete randomization to assign one of these to treatment and one to control. With the remainder (4, two in each stratum), we again use complete randomization by stratum. Hence, in the 25 libraries where this was the case, we technically had a third, 'mixed' randomization stratum that consisted of one low-spillover village and one high-spillover village. Since the exact villages that were part of this third stratum were randomly drawn, we had theoretical reasons to expect the potential outcomes to be more correlated with other villages within the same spillover category, rather than between the two villages that were randomly selected to be part of the mixed randomization stratum. Hence, our models include library-spillover fixed effects (FEs). In Appendix J we show that our results are robust to a range of alternative fixed-effects specifications.

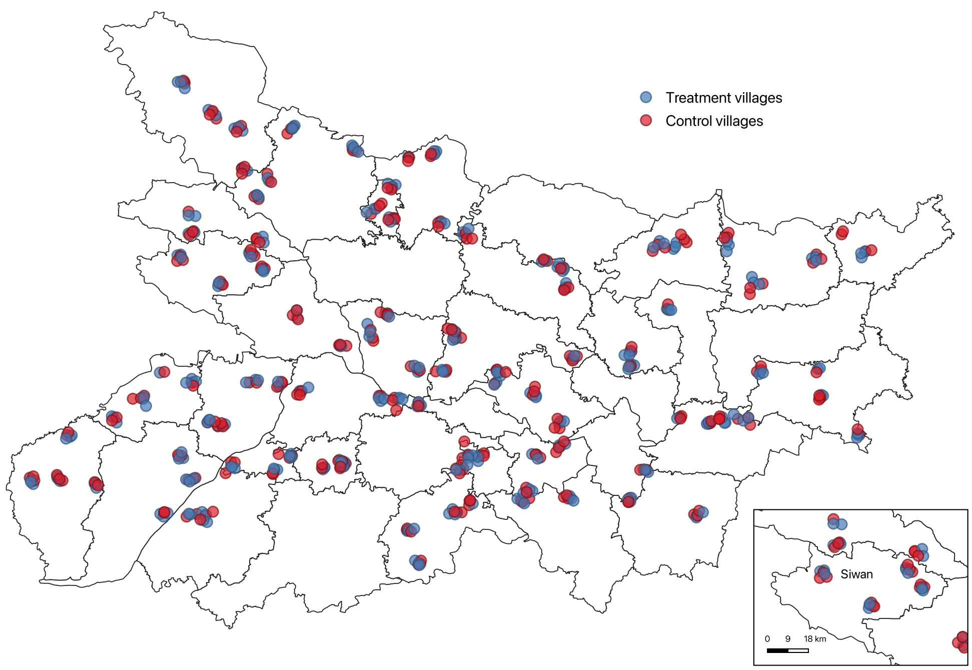
[1] This was the case for 7 of the 100 libraries. We selected, respectively, 5, 5, 4, 3, 3, 3 and 2 villages around these libraries.

[2] We anticipated no spillover *across* GPs, as they generally consist of around 2000 households, and students usually attend the nearest school within their GP. A risk of spillover on the other hand exists between villages *within* the same GP, as students may frequent the same school, but we expect it to be low since students are split into multiple classes per grade. However, we note that the risk of spillover cannot be fully eliminated, as there is still a possibility that a few participants from different villages might end up in the same class in secondary school. Estimating heterogeneous treatment effects by spillover strata allows us to test for spillover between treatment and control (see Table H18).

Overall, these procedures lead us to assign roughly the same number of units to treatment (294) and control (289). The procedure also allowed us to select a large group of units (both treatment and control) that have a low-spillover potential. Our sample counts 189 (out of 583) low-spillover villages pre-randomization. The number of low-spillover villages increased to 250 after randomization at the library level. This was because, during randomization, all villages within a GP occasionally fell into the same treatment group, reducing concerns about spillovers between treatment and control units. In such cases, these villages were re-classified as low-spillover according to our definition. Since this post-randomization classification more accurately reflects spillover potential, we use it in our heterogeneity models to evaluate whether spillovers affect results. However, since the post-randomization spillover classification is not reflective of the stratified randomization procedure outlined above, our main models use pre-randomization spillover-strata for the library-spillover FEs.

Figure A1 shows the location of treatment and control villages across Bihar.

Figure A1: Map of treatment and control villages



A.3 Sampling households

Within each of the 583 selected villages, we then relied on Jeevika to provide a list of students eligible for the study, based on existing household list data that the government has from voter rolls and enrollment in government programs. Initial criteria for eligibility in the study included households with children enrolled in government schools in grades 8 to 12.

Once a long list of such households was generated, a representative from Jeevika visited these households to (a) confirm that an age-eligible and school-going child was indeed present, and (b) if so, to ask whether the student as well as a parent or guardian present were interested in the program. Once students and parents agreed after this initial pitch, a shorter list of (20 to 24) eligible and interested students was generated within each selected village by Jeevika staff and given to us.

Crucially, we note that randomization at the village level took place *after* students opted in, so that the issue of differential opt-in rates between treatment and control was avoided. On this point, everyone involved in this study – including teachers, our implementation partners, government officials, and coauthors – was blind to treatment assignment at the time students were recruited. During the initial household visit, students and their parents/guardians were merely told that they would be given a free opportunity to go through a government-offered certificate course with 4 sessions, the content of which has benefits for their future and careers as they go off to college or on the job market.

After we received the short list of students in each village from Jeevika, our survey team visited these households. Enumerators visited each house in person for the baseline survey, which included three additional eligibility criteria. First, enumerators obtained official oral consent from both children as well as one parent or guardian present to conduct the baseline as well as return for an endline survey after the classroom sessions. Second, we included a one-item measure of students' basic (third-grade) reading comprehension in Hindi. If students failed this item, the household in question was replaced in our sampling frame. Third, we asked both students and parents to verbally affirm that the student would be able to attend four classroom sessions during the stipulated weeks and timings; if either said no we dropped the household from the sample and replaced it.

After students and their parents/guardians opted in, the baseline survey was completed, and the final sample was generated, we randomized and assigned half of the included villages to treatment based on the criteria described above. As such, students did not know whether they were receiving treatment or or control classes until they showed up to the first day of the sessions.

We recognize that opting in is not random, meaning our sample is unlikely to represent all potentially eligible households. While we lack systematic data on the factors influencing participation, field notes suggest that trust in the government representative who visited the household played a significant role. This likely reflects patterns seen in the actual uptake of government programs if they were to be rolled out in a similar manner.

A.4 Power calculations

We aimed to included one classroom size of students (20 to 24 students) per village in the study, which means that, overall, up to $600 * 24 = 14400$ students were targeted to be part of this study initially (with roughly half of these in the treatment group). Our choice for sample size and number of students in each group was based on evidence suggesting that our intervention could produce significant effects.

Specifically, this sample size would allow us to detect a treatment effect with at least

80% power. For effect sizes, we relied on estimates from recent work on misinformation in South Asia that finds that corrective interventions against misinformation lead to about a 4-8% decrease in belief in misinformation (Ali and Qazi, 2023; Chauchard and Badrinathan, 2025). While we designed our study to detect an effect of a similar size, we note that the BIMLI treatment lacks a valid direct comparison. It was intentionally more intensive than most existing, light-touch interventions. The closest comparison would have been Badrinathan (2021)'s media literacy intervention in the same region, which produced a null effect. Although comparisons with media literacy type studies in other contexts were possible, significant contextual differences made it more appropriate to reference a slightly different intervention within the same country. Notably, no intervention, even in other contexts, closely parallels BIMLI. In addition, due to the nature of our large-scale and in-person endline survey, respondents were surveyed 1-5 weeks after the intervention ended; most work on misinformation suggests that effects decay rapidly without an immediate post-test (Capewell et al., 2024). Thus, regardless of the reference point, direct comparisons for effect sizes were unavailable.

Under the following assumptions (significance level = 0.05; intra-cluster correlation (ICC) = 0.20; number of clusters per experimental group $\approx$ 300), we calculated that we would need to sample 14,000 students in about 600 clusters in order to achieve 80% power.

A.5 Balance tables

Table A1 examines whether treatment and control villages are balanced on demographic and developmental characteristics using census data from 2011, the last year that the Indian government published a census. We have data for 546 villages. We could not collect data on all 583 villages for three reasons: 1) the census data is from 2011, while our survey is current, 2) census data is unavailable for several villages in our sample, and 3) some villages could not be matched to the census due to incomplete, colloquial, or generic names, which have also changed over time. We show that treatment and control villages are not significantly different on a number of key characteristics.

Table A2 examines the balance between treatment and control participants across demographic and household variables, as well as pre-test covariates. We find that these groups are generally balanced, with the exception of age (and similarly grade), where treatment participants are slightly older (14.9 years vs. 14.8 years). However, these differences are substantively minor and disappear when p-values are adjusted for multiple comparisons. Similarly, for trust in social media, the treatment group mean is slightly higher, but this difference also dissipates after adjustment.

Finally, Table A3 compares village-level characteristics between villages included in our sample (those near libraries) and those slightly further away that were not included. Using census data, we find that sampled villages are socio-economically better off on several development indicators (again noting that Indian census data is from 13 years ago).

Table A1: Balance between treatment and control villages

	N	Treatment	Control	Diff.	SE	p-value
No. of households	546	1630.541	1830.761	-200.22	484.32	0.679
Total population	546	8767.474	9722.783	-955.31	2491.27	0.702
Share SC population	546	0.167	0.170	-0.00	0.01	0.780
Share ST population	546	0.009	0.008	0.00	0.00	0.602
Share literate	546	0.505	0.516	-0.01	0.01	0.165
Total area (km)	546	477.478	505.170	-27.69	55.30	0.617
No. of primary schools	546	2.926	3.004	-0.08	0.27	0.771
No. of middle schools	546	1.478	1.598	-0.12	0.15	0.417
No. of secondary schools	546	0.474	0.370	0.10	0.06	0.104
Pucca road	546	0.837	0.801	0.04	0.03	0.272
Power supply (domestic)	546	0.804	0.772	0.03	0.04	0.362
Power supply (agricultural)	546	0.444	0.435	0.01	0.04	0.820

*p<0.05; **p<0.01; ***p<0.001

Table A2: Balance between treatment and control participants

	N	Treatment	Control	Diff.	SE	p	p (FDR)
Gender - Girls	13,591	0.590	0.580	0.01	0.012	0.64	0.74
Grade	13,590	9.700	9.620	0.08	0.034	0.01*	0.13
Age	13,591	14.960	14.850	0.09	0.040	0.02*	0.13
Religion - Hindu	13,591	0.900	0.910	-0.01	0.013	0.52	0.72
Language - Hindi	13,591	0.420	0.440	-0.02	0.014	0.12	0.31
Caste - GEN	13,393	0.080	0.090	-0.01	0.010	0.26	0.54
Caste - OBC/EBC	13,393	0.690	0.700	-0.01	0.019	0.48	0.72
Caste - SC	13,393	0.210	0.190	0.03	0.017	0.13	0.31
Caste - ST	13,393	0.020	0.020	0.00	0.004	>0.9	>0.9
Asset Index	13,591	-0.010	0.010	-0.05	0.028	0.07	0.3
Father's Education	12,891	6.920	6.920	-0.04	0.117	0.76	0.84
Mother's Education	12,950	4.070	4.160	-0.13	0.122	0.3	0.57
Government School	13,591	0.960	0.960	0.01	0.004	0.1	0.3
Science Knowledge	13,591	4.290	4.290	-0.02	0.038	0.57	0.74
Mobile Internet	13,591	0.180	0.190	0.00	0.008	>0.9	>0.9
Trust Newspapers	13,591	0.910	0.900	0.01	0.006	0.39	0.63
Trust Social Media	13,591	0.620	0.590	0.03	0.011	0.02*	0.13
Trust TV	13,591	0.850	0.830	0.01	0.008	0.08	0.3
Trust Friends and Family	13,591	0.970	0.970	0.00	0.004	0.35	0.61
Trust Vaccinated	13,591	0.780	0.760	0.02	0.011	0.08	0.3
Trust Ayurveda	13,591	0.870	0.870	0.00	0.007	0.6	0.74

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover FEs.

Last column reports p-values adjusted for the False Discovery Rate (FDR).

Table A3: Comparison between sample and non-sample villages

	Sample	**Non-Sample**
No. of households	1731.751	439.499
Total population	9250.377	2404.270
Share SC population	1180.842	397.444
Share ST population	154.178	33.847
Share literate	4934.445	1174.652
Total area (km)	491.476	220.990
No. of primary schools	2.965	1.389
No. of middle schools	1.538	0.677
No. of secondary schools	0.421	0.158
Pucca road	0.819	0.637
Power supply (domestic)	0.788	0.614
Power supply (agricultural)	0.440	0.325
N	546	38646

B Teaching materials

B.1 Treatment group

For treatment group classes, we produced and distributed a wealth of materials relevant to each session in collaboration with our implementing partner DataLeads, to both students and teachers. We describe these and provide examples below:

1. An extensive **slide deck for teachers** was used during our training sessions with teachers; these provided an overview of the entire curriculum for BIMLI. The deck, broken up into the 4 modules corresponding to each lesson, discussed in detail the specific content of the curriculum, including definitions, materials for teachers to learn curriculum concepts, and examples. We conducted a comprehensive 2-day teacher training where facilitators taught and explained the curriculum to selected teachers using these slides. Following this, we printed and provided bound copies of the 100-page slide deck to each teacher for revision after the training, which doubled as a textbook outlining the curriculum. Below we provide examples of slides from the deck. In Figure B2 the slide discusses some common reasons why people are vulnerable to misinformation; this was included in the first module. In Figure B3 also from the first module, teachers were told to use the example of how viral misinformation in India has previously resulted in violent consequences such lynchings and deaths, to underscore a point about the consequences that misinformation can have. Figure B4 shows slides used to conduct the last module, a component of which focused on how children should address and deal with conversations where adults they know shared misinformation. The first slide discussed some tips when confronted with this situation, and the next outlined a role playing activity that students did in pairs in the classroom.

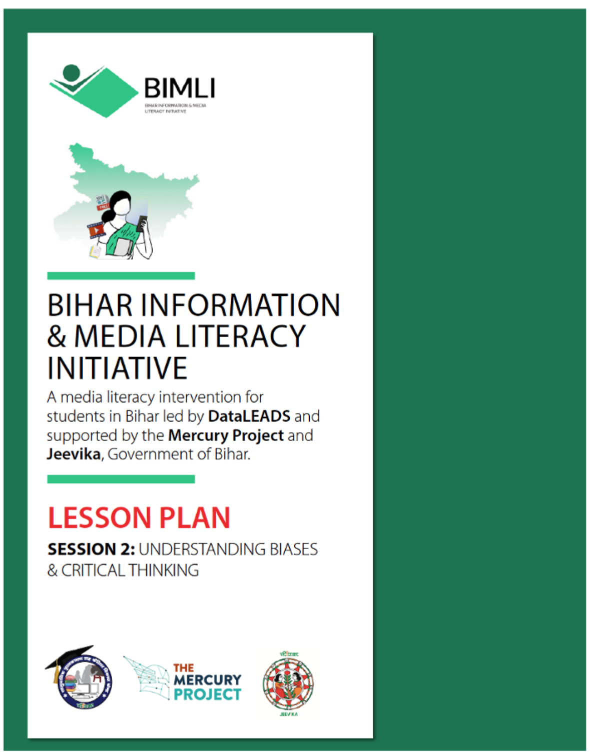
Figure B2: Slide deck example outlining reasons for vulnerability to misinformation

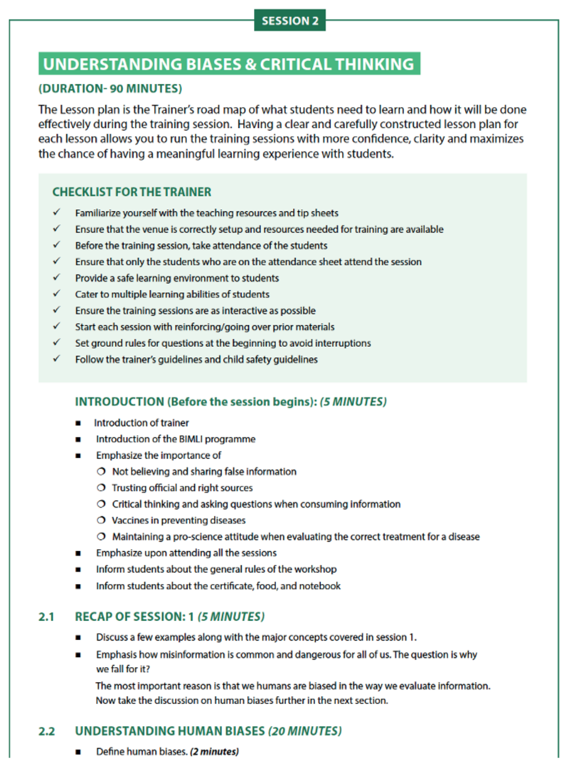
Figure B3: Slide deck example of a real-life consequence of misinformation in India

Figure B4: Slide deck example of class activity on talking to relatives

2. Next, we created detailed and comprehensive **lesson plans for teachers** relevant to each of the four sessions. These lesson plans were meant to provide teachers road maps translating the slide deck content into class time, to ensure a highly standardized delivery of the intervention across locations. Each plan contained (a) a procedural checklist for teachers of what to do to ensure the session goes well, followed by (b) roughly timed sub-modules, including discussion questions to ask students at each juncture, in a pre-specified order. Each module typically focused on a concept, and specified how to explain and illustrate it with as much participation and input from the class as possible. Starting with session 2, the first module following the introduction to the session was a reminder of previous sessions. Moreover, within each module, we specified activities and discussion topics to ensure that the sessions were lively and participatory. The modules also listed some examples, though teachers were also instructed to draw on other materials we provided them with (and to a lesser extent on class-generated valid alternative examples). The last module typically was a summary of the whole session. We provide an example of the lesson plan below in Figure B5 that was used for the second session.

Figure B5: Session 2 lesson plan example





SESSION 2

UNDERSTANDING BIASES & CRITICAL THINKING

(DURATION- 90 MINUTES)

The Lesson plan is the Trainer's road map of what students need to learn and how it will be done effectively during the training session. Having a clear and carefully constructed lesson plan for each lesson allows you to run the training sessions with more confidence, clarity and maximizes the chance of having a meaningful learning experience with students.

CHECKLIST FOR THE TRAINER

- ✓ Familiarize yourself with the teaching resources and tip sheets
- ✓ Ensure that the venue is correctly setup and resources needed for training are available
- ✓ Before the training session, take attendance of the students
- ✓ Ensure that only the students who are on the attendance sheet attend the session
- ✓ Provide a safe learning environment to students
- ✓ Cater to multiple learning abilities of students
- ✓ Ensure the training sessions are as interactive as possible
- ✓ Start each session with reinforcing/going over prior materials
- ✓ Set ground rules for questions at the beginning to avoid interruptions
- ✓ Follow the trainer's guidelines and child safety guidelines

INTRODUCTION (Before the session begins): *(5 MINUTES)*

- Introduction of trainer
- Introduction of the BIMLI programme
- Emphasize the importance of
  - Not believing and sharing false information
  - Trusting official and right sources
  - Critical thinking and asking questions when consuming information
  - Vaccines in preventing diseases
  - Maintaining a pro-science attitude when evaluating the correct treatment for a disease
- Emphasize upon attending all the sessions
- Inform students about the general rules of the workshop
- Inform students about the certificate, food, and notebook

2.1 RECAP OF SESSION: 1 *(5 MINUTES)*

- Discuss a few examples along with the major concepts covered in session 1.
- Emphasis how misinformation is common and dangerous for all of us. The question is why we fall for it?

  The most important reason is that we humans are biased in the way we evaluate information. Now take the discussion on human biases further in the next section.

2.2 UNDERSTANDING HUMAN BIASES *(20 MINUTES)*

- Define human biases. ***(2 minutes)***

3. **A set of A3 posters/charts** was additionally provided to each teacher to use over the course of their lectures/sessions as illustrative materials and to act as flip charts. Anticipating that in some cases classrooms may not have adequate resources (such

as blackboards) and given that the curriculum was designed to be taught offline, we generated certain teaching materials in the form of large posters to be circulated to students in class hour/stuck on walls as instructional material. These did not typically contain any new content but were aimed at switching up the oral lecture to incorporate visual elements. An example is provided below.

Figure B6: A3 flipchart posters that discuss confirmation bias using fables as examples

4. **A set of tipsheets aimed at students** that summarized some of the most salient points in each lesson and acted as take-home reminders of class content. Importantly, because these needed to remain compact, they did not contain all teaching points, but a few crucial take-aways. Figure B7 provides an example.

Figure B7: Lesson summary tip-sheets distributed to treatment students

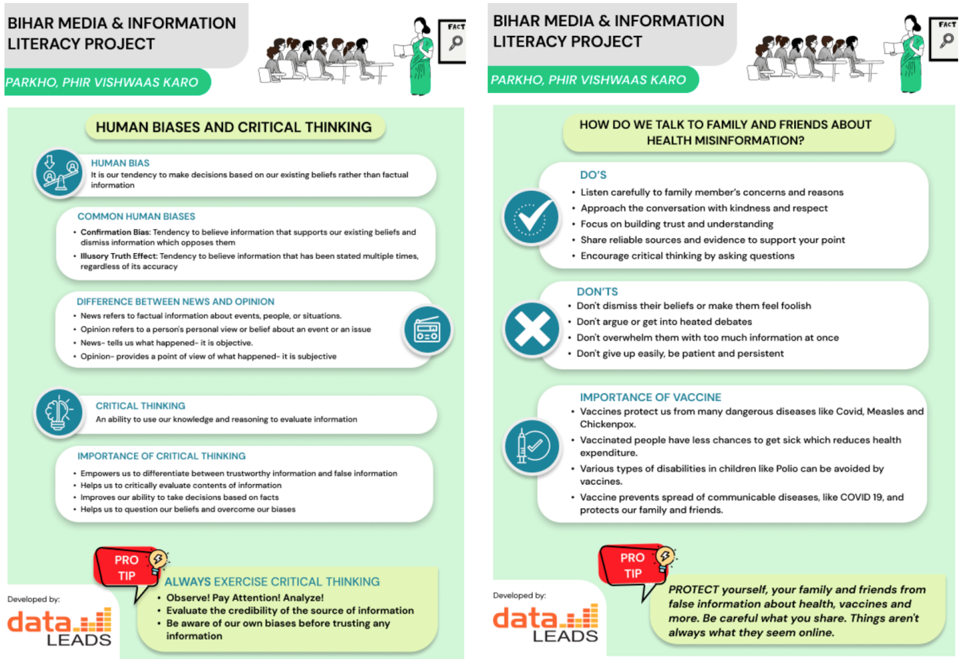


Figure B8: Example of a treatment classroom in session (Vaishali, Bihar)

5. A set of 3 **homework assignments for students** that were distributed at the end of classes 1, 2 and 3, and that students were asked to complete/reflect on between classes at home. Figure B9 provides an example of a take-home homework assignment.

Figure B9: Sample homework assignment for Session 3

HOMEWORK FOR SESSION 3

Student Name ____________________________

BIMLI

PLEASE ANSWER THE FOLLOWING AND BRING IT TO CLASS. WE WILL DISCUSS THESE QUESTIONS NEXT TIME.

1. **As we saw during our last class, it is very important to rely on reliable sources for information, especially information about matters of health. Imagine you have contracted the common cold and want to find out more about potential ways of treating it. List three reliable sources that you personally consult from home or within your village to find this information:**

   1. ____________________________
   2. ____________________________
   3. ____________________________

2. **Imagine a friend tells you about drinking lemon water every morning as a cure for chronic fatigue syndrome (constantly feeling tired and low energy). You want to know whether this information is true. Given this eample, go through the following three exercises to verify or debunk this information:**

   a. What is the source of information?

   b. In Question 1, you listed three reliable sources of information that you may consult. What do these sources say about drinking lemon water as a cure for chronic fatigue syndrome?

   c. Overall, what is the evidence for or against this information? Based on what you found, is this information accurate? Would you share it with a family member? Why? Why not?

   ____

3. **In the next class, we will learn about how to talk to parents and other relatives about misinformation. To prepare you for next class, go through the following exercises with a member of your family:**

   a. Explain to your relative what you've been learning about misinformation, its prevalence, and its potential consequences.

   b. Ask if they recall a time when they believed something to be true that they later found out to be wrong. It could be a piece of news, health advice, or even a misleading message on social media. Note down any examples they have.

   c. Discuss why you both think misinformation spreads so quickly.

   d. Brainstorm potential ways to verify a piece of information you've recently received. Note down a few ideas you discussed:

   ____

   ____

4. **On the back of this sheet, reflect on how the conversation with your relative went. Did you find it difficult to talk about misinformation with them? Did anything surprise you? Do you think they understood how important fact-checking and relying on verified information is?**

6. We also distributed a BIMLI-branded **notebook for students** to allow them to take notes about the content of each session. The notebooks included a number of headers corresponding to each module taught by the teachers, followed by blank pages allowing students to write notes, and ending with a reminder of the main points taught in each lesson. Teachers also encouraged the students to stick or staple the tipsheets within the notebooks in order to ensure they did not get lost.

B.2 Control group

Control group classes were entitled “The Basics of Communicative English”. As with the treatment, we hired external facilitators to deliver the lessons, and custom created a curriculum suited to the local context with the help of area experts. The primary objective of these sessions was to equip students with foundational language skills aimed at enhancing their prospects in future career endeavors and job interviews. Notably, these modules deliberately omitted written and reading components, with the instruction solely concentrated on spoken phrases and sentences. We highlight that the instructional content was tailored to cater to students with limited or no prior exposure to the English language, particularly those primarily educated in non-English medium schools. Consequently, the modules were designed to align with the proficiency level of first-time learners. As such, there is no concern that students would have acquired advanced English skills from these modules to navigate the internet or influence misinformation outcomes. The classes remained highly basic in their content and delivery. Importantly, similar to treatment we provided teachers with instructional materials and a structured time-use lesson plan. These materials included the same types of activities as treatment – paired exercises, classroom discussions, and role-playing. The level of engagement and the quality of discussion were deliberately held constant across both the treatment and control groups. Below we describe the content of each of the 4 modules:

1. **Introducing oneself and the importance of English**. This lesson motivated students by emphasizing the importance of learning English, similar to how Misinformation Lesson 1 raised awareness about misinformation and its dangers. Teachers discussed the benefits of English for careers, higher education, and global communication. The lesson then introduced self-introduction phrases, such as “My name is...,” “My father’s name is...,” and “My mother’s name is...,” along with vocabulary for family members and expressing where one lives. Students also learned to describe hobbies and interests using -ing verbs like singing and dancing. A role-playing activity had students practice introducing themselves to one another.

2. **Things around us:** This session introduced students to describing objects and things around them. This included things in the house, in the classroom, in the village, common objects, and food. Students were encouraged to look around them and name the things they see; these words were then translated and taught to the class. Next they were introduced vocabulary describing what these things were placed on; this part introduced them to “what” and “where” questions. The session ended with a think-pair-share activity.

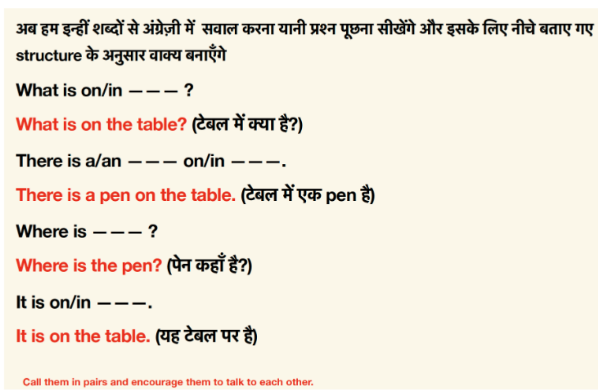
3. **Things we do everyday:** This session introduced action verbs to students with the aim of allowing them to describe their everyday actions. Vocabulary included verbs like learn, study, play, eat, drink, sleep, shower, etc. In the second part of the lesson we introduced adverbs and they were taught to say “I play sometimes” or “I study often”

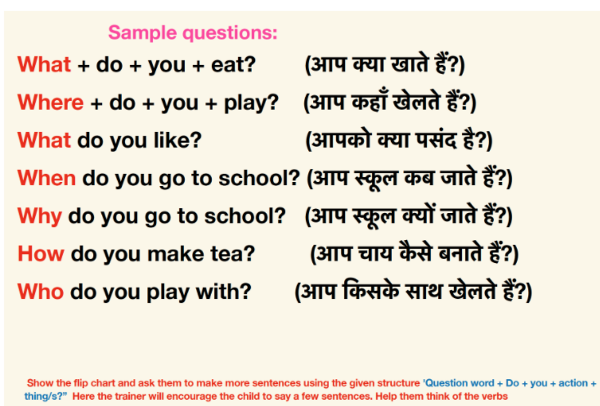
or “I never drink coffee”. Finally, a group activity had them ask each other questions about how frequently they do certain activities.

4. **Question words and tenses:** In the first half of this session students learned to ask questions and make sentences starting with what, who, where, when, why, and how. In the second half, they recapped the actions verbs from last time and learned their past tense. Some new verbs were also introduced in this round. Finally, they were taught one form of expressing future actions (e.g., I will go to school tomorrow). The concluding activity of the 4 sessions was that each student had the opportunity to introduce themselves in full to the entire class, using sentences and words they had learned throughout the course.

As with the treatment group, we held a training session for prospective teachers to determine if they were able to master the curriculum and had a dynamic approach in the classroom. We used instructional slides that backed up as lesson plans for these sessions, and ultimately provided selected teachers with hard and soft copies of these materials that acted as a detailed guide to each lesson. In addition, they were provided with large sized printed copies of flip charts to be used for certain activities. Figure B10 provides examples.

Figure B10: Flip chart examples from English classes





C Compliance

We implemented a number of steps to maximize treatment uptake and ensure continued engagement with all 4 sessions. First, because of our partnership with Jeevika, we were able to count on their local staff members (*didis*) to mobilize and remind students of upcoming classes. While we may anticipate variations in the degree of motivation of the *didis* at the local level, we note that these actors face strong incentives within the organization to comply; in addition, they were compensated financially for the extra work that the organization of the program generated. Students themselves had strong incentives to attend: upon enrolling, they were periodically reminded by *didis* to attend; second, if they attended they would receive a certificate specifying that they completed the course directly with Jeevika; third, they were provided with a variety of materials they could hold on to for reference, in addition to notebooks and stationery; finally they were provided with snacks and refreshments at the end of class.

Despite this, we anticipated the possibility that students would be unable or unwilling to attend all sessions, for a variety of reasons, making non-compliance an issue. In this study, we define compliance as a binary measure signifying minimal compliance, where 1 indicates that the particular student attended *at least* one of the four sessions, and 0 indicates that the student attended no session – we measure this with respondent-level attendance data gathered by teachers during each session. Table C6 shows the distribution of attendance by number of sessions attended for treatment and control.

Crucially, our definition of attrition does not hinge on compliance with the treatment protocol. Even participants who are (partially) non-compliant with the treatment but continue to engage with the study by completing the endline survey would not be classified as having attrited. Potential causes of non-compliance include the long intervention period and the difficult local (rural) context. For example, it might not be feasible for children to attend all four sessions, possibly due to travel constraints or conflicting commitments. Non-compliance may occur due to random factors that are unrelated to any observed or unobserved characteristics of the individuals.

Table C4 displays the baseline predictors of compliance, operationalized as a binary measure as described above. It highlights that female respondents, respondents whose mothers have higher levels of education, and those with higher science knowledge attended sessions at higher rates. On the other hand, older students, those who are in higher school grades/classes attended at lower rates. Table C5 highlights differential compliance between treatment and control groups. While the coefficient on gender and social media is significant (indicating that among individuals assigned to treatment, social media usage and being a girl has a significantly larger effect on compliance), these effects are substantively small and barely significant. Moreover this table estimates multiple coefficients; with a 95% CI, we would expect some estimates to be significant under the null by chance alone.

Table C4: Compliance predictors

Predictor	**N**	**Estimate**	**SE**	**p-value**
Gender - Girl	13,591	0.05***	0.007	<0.001
Grade	13,590	-0.02***	0.002	<0.001
Age	13,591	-0.02***	0.002	<0.001
Religion - Hindu	13,591	0.00	0.012	0.78
Language - Hindi	13,591	0.00	0.008	0.55
Asset Index	13,591	0.00	0.003	>0.9
Father's Education	12,891	0.00	0.001	0.23
Mother's Education	12,950	0.00**	0.001	0.002
Government School	13,591	0.00	0.015	>0.9
Science Knowledge	13,591	0.01**	0.002	0.002
Mobile Internet	13,591	0.00	0.008	0.65
Newspapers	13,591	0.00	0.009	0.69
Social Media	13,591	0.00	0.006	0.88
TV	13,591	0.01	0.007	0.46
Friends and Family	13,591	0.00	0.015	0.85
Vaccinated	13,591	0.00	0.007	>0.9
Ayurveda	13,591	0.00	0.008	0.65

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

Table C5: Compliance predictors (*treatment)

Predictor	**N**	**Estimate**	**SE**	**p-value**
Gender - Girl * T	13,591	0.03*	0.013	0.04
Grade * T	13,590	-0.01	0.005	0.26
Age * T	13,591	-0.01	0.004	0.14
Religion - Hindu * T	13,591	0.00	0.022	>0.9
Language - Hindi * T	13,591	0.00	0.013	0.75
Asset Index * T	13,591	0.00	0.006	0.45
Father's Education * T	12,891	0.00	0.001	>0.9
Mother's Education * T	12,950	0.00	0.001	>0.9
Government School * T	13,591	0.06	0.030	0.07
Science Knowledge * T	13,591	0.00	0.004	>0.9
Mobile Internet * T	13,591	-0.01	0.015	0.44
Newspapers * T	13,591	0.03	0.017	0.08
Social Media * T	13,591	0.03*	0.011	0.02
TV * T	13,591	0.02	0.014	0.08
Friends and Family * T	13,591	0.04	0.028	0.15
Vaccinated * T	13,591	0.00	0.014	>0.9
Ayurveda * T	13,591	0.01	0.015	0.44

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

Table C6: Number of sessions attended

Sessions attended	N Control (%)	N Treatment (%)
0	565 (8.34%)	798 (11.71%)
1	617 (9.11%)	431 (6.32%)
2	858 (12.67%)	465 (6.82%)
3	1473 (21.74%)	1224 (17.96%)
4	3261 (48.14%)	3899 (57.20%)

D Outcomes

D.1 Outcome construction

To estimate the effect of BIMLI we use endline survey data to compare treated individuals to the control group. Our endline survey covers seven main families of outcomes. For each family of outcomes, we compute inverse covariance weighted (ICW) indices that are standardized relative to the control group. In our primary hypotheses (pre-specified in PAP), the main comparison of interest is every respondent assigned to treatment relative to every respondent assigned to control. Apart from measuring each outcome as an index, we also break down the index into its components to understand and visualize which items in the index play a more or less influential role. We opted to use inverse-covariance weighted (ICW) indices because they provide a data-driven method for combining multiple outcomes into a single overall index. This approach has three main advantages. First, it addresses multiple comparison concerns by allowing us to use one outcome measure to test each main hypothesis. Second, weighting by the inverse of the covariance, as opposed to creating a simple additive index, minimizes arbitrary decisions by researchers regarding the number of individual survey items used to measure a concept. According to the logic of ICW, if multiple questions measure the same latent tendency, the overall index down-weights each item proportionally, effectively treating them as one. Third, the weighting method is more intuitive and less arbitrary than applying factor analysis. This strikes a good balance between letting the data speak without simply letting an opaque algorithm do the thinking for us. The creation of the ICWs require multiple steps and iterations of standardization, which we detail below using the example of the *Accuracy Discernment* measure:

1. Standardize each true and false statement with respect to the control group;
2. Calculate the ICW index of true information and the ICW index of false information;
3. Standardize both ICW subindices with respect to the control group
4. Calculate the difference between the ICW_{TRUE} and ICW_{FALSE} indices to get a measure of discernment; and lastly,
5. Standardize the resulting discernment measure relative to the control group so that any treatment effects are in terms of standard deviations of the final discernment measure relative to the control group.

Table D11 provides a summary of the different components that comprise each family of outcomes, the survey items measuring each component, and the the method of index construction for each outcome family. The full instrument and codebook is available with our dataset in the Harvard Dataverse.

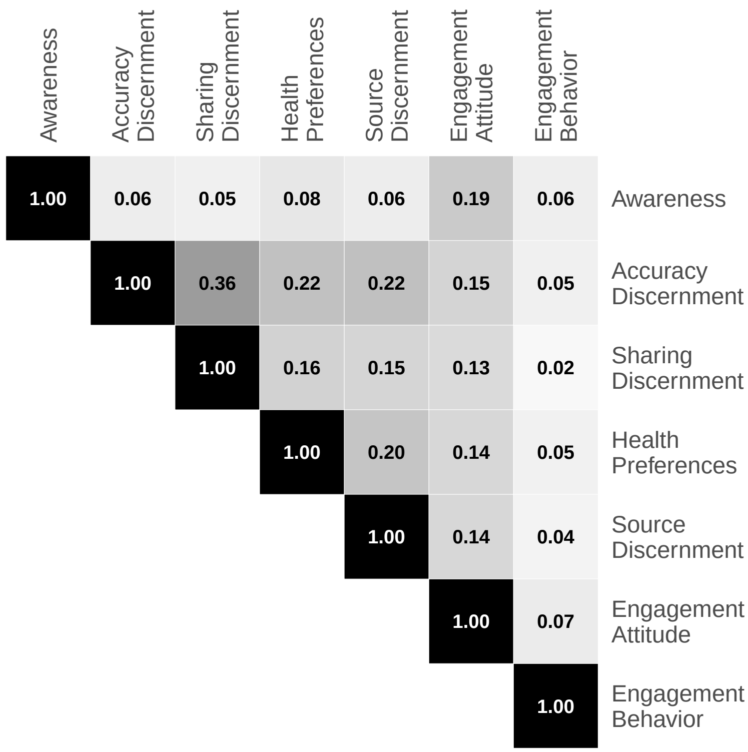
Figure D11: Index Calculation for Outcomes

Outcome Family	Component	Items	Component Index/Measure Construction	Combined Index Construction
1. Awareness	Dangers of misinformation	misinfo_threat1	single-item z-score, standardized relative control group	ICW index using all component questions, standardized relative to control (lower values indicate better awareness)
	Awareness of media biases	news_manipulation1; news_manipulation2	simple average of z-scores relative to control (lower values indicate higher awareness)	
	Awareness of pyschological biases	bias2; bias3	simple average of z-scores relative to control (lower values indicate higher awareness)	
2. Accuracy discernment	Accuracy of false statements	discernment1; discernment2; discernment4; discernment5	ICW index, standardized relative to control group (lower values indicate higher levels of believed accuracy)	ICW index score (true) – ICW index score (false) (lower values indicate higher accuracy discernment)
	Accuracy of true statements	discernment7; discernment8; discernment9; discernment10	ICW index, standardized relative to control group (lower values indicate higher levels of believed accuracy)	
3. Sharing discernment	Sharing intentions of false statements	sharing1; sharing2; sharing4; sharing5	ICW index, standardized relative to control group (lower values indicate higher levels of sharing intentions)	ICW index score (true) – ICW index score (false) (lower values indicate higher sharing discernment)
	Sharing intentions of true statements	sharing7; sharing8; sharing9; sharing10	ICW index, standardized relative to control group (lower values indicate higher levels of sharing intentions)	
4. Health preferences	Interest in health news	news_interest_health	single-item z-score, standardized relative control group (lower values indicate higher interest)	ICW index of component measures, standardized relative to control (lower values indicate better health preferences)
	Vaccine perception	vaccine_safety1; vaccine_safety2	ICW index, standardized relative to control group (lower values indicate higher levels of perceived vaccine safety)	
	Reliance on alternative medicine	illness_response (go to traditional healer and treat at home); ayurveda (reverse coded)	simple average of z-scores relative to control (lower values indicate less reliance on alternative medicine)	
5. Source discernment	General source discernment	source_discern_generic	standardized ICW index of good sources (MBBS doctors; health workers such as ASHA; government health pamphlets) — standardized ICW index of bad sources (word of mouth; ayurvedic doctors; jholachhap doctors) (lower values indicate less reliance on bad sources)	ICW index of component measures, standardized relative to control group (lower values indicate better source discernment)
	Situation-specific source discernment	source_discern_specific	standardized ICW index of good sources (local health worker or community health center; government-issued health pamphlets or posters; TV interview with doctor from AIIMS) — standardized ICW index of bad sources (Stories or remedies passed down in your family; messages and videos shared as WA forwards; TV interview with an ayurvedic doctor) (lower values indicate less reliance on bad sources)	
	Cues/heuristics	cues	standardized index of good cues (reputable outlet; sensational tone) — standardized index of bad sources (number of likes; whether sender is from same community) (lower values indicate less reliance on cues)	
6. Demand for fact-checking	Reaction to misinformation	engagement_attitude1	standardized measure of good response (correcting friend) — standardized measure of bad response (share with others) (lower values indicate better responses)	ICW index of component measures, standardized relative to control group (lower values indicate higher demand)
	Importance of verification	engagement_attitude2	single-item z-score, standardized relative to control (lower values indicate better engagement)	
	Frequency of verification	engagement_attitude3	single-item z-score, standardized relative to control (lower values indicate better engagement)	
7. Engagement with misinformation counter efforts	Sign up for reputable newspaper	engagement_behavior1	single-item z-score, standardized relative control group (lower values indicate higher propensity)	Simple average of component measures (lower values indicate higher engagement)
	Truth ambassador	engagement_behavior2	single-item z-score, standardized relative control group (lower values indicate higher propensity)	

D.2 Correlation matrix

To explore whether each outcome family captures a distinct construct, we generated a correlation matrix of the individual outcomes (Figure D12). Correlations across items are generally low, with the highest Pearson's r being 0.36 – between sharing discernment and accuracy discernment – which is both expected and reasonable.

Figure D12: Correlations between outcome measures



D.3 Political discernment

Table D7: Political Discernment Stories (Second Endline)

Story	**Veracity**
The Congress Party regularly receives campaign funds from foreign countries.	False
In the recent Indian general election, the Election Commission found evidence of manipulation/tampering in EVM machines.	False
The BJP has received more money in campaign funds in the 2024 elections than any other party in India	True
Prime Minister Modi lost his Lok Sabha seat in Varanasi in 2024	True

E Teacher data

In Table E8 we tabulate summary statistics of teachers by treatment condition. We note that the recruitment and selection process differed for treatment and control, likely leading to some differences in the pool of teachers across conditions. For example, control group classes recruited largely existing local school teachers, hence this group was likely to have lower levels of education and reside in rural areas; on the contrary, treatment group classes, because of their specialized nature, ended with a pool of teachers educated in Patna, with on average more years of education. It is possible that this selection of a more urban population for treatment classes also led to the inclusion of more women. Finally, religious differences in the sample pool could reflect different networks of recruiting: for treatment classes we relied on an external consultant, DataLeads, and for control classes we relied on Jeevika.

In Table E9, we examine whether teachers' demographic characteristics influenced outcomes, by interacting three demographic variables with the treatment—religion (an indicator for Muslim), gender (an indicator for female), and caste (an indicator for general caste) in separate models. Note that in these models we use only district fixed effects, since in most cases the same teacher taught all classes at one library location. We find no significant interaction effects for most teacher covariates, with the exception for health preferences, where there is a statistically significant positive interaction effect with the teacher being female. We also note that teacher gender was not randomly assigned but we do not have any reason to expect it would correlate with potential outcomes.

Table E8: Summary statistics of teachers, by treatment condition

	Treatment	Control
Statistic	Mean (SD)/Prop.	Mean (SD)/Prop.
Age	34.422 (8.41)	31.211 (8.57)
Gender		
Female	0.333	0.158
Male	0.644	0.842
Caste Category		
ST	0.000	0.000
SC	0.022	0.053
General / Upper	0.533	0.386
OBC	0.289	0.526
Prefer not to say	0.089	0.035
Religion		
Hindu	0.711	0.912
Muslim	0.178	0.088
Other	0.111	0.000
Education - School Type		
Government	0.644	0.667
Private	0.333	0.281
Other	0.022	0.053
N	45	57

Table E9: Heterogeneity results by teacher characteristics

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
T x Female teacher	0.109	0.006	0.174*	0.107	-0.026	0.020	0.022
	(0.075)	(0.085)	(0.086)	(0.070)	(0.096)	(0.068)	(0.103)
N	10551	10693	10734	10837	11234	11364	11153
T x Muslim teacher	-0.024	0.032	0.110	0.033	-0.086	-0.034	0.003
	(0.101)	(0.098)	(0.115)	(0.081)	(0.090)	(0.090)	(0.120)
N	10551	10693	10734	10837	11234	11364	11153
T x General caste teacher	0.025	-0.019	-0.013	-0.044	0.111	0.002	0.036
	(0.066)	(0.064)	(0.061)	(0.061)	(0.067)	(0.060)	(0.078)
N	10551	10693	10734	10837	11234	11364	11153
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*p<0.05; **p<0.01; ***p<0.001. Models include district FEs.

F Ethics

The study was approved by IRB (expedited review) at American University and the project was reviewed by the UC3M's GDPR data protection officer (approval certificates in the Dataverse). Ethical considerations were paramount in the design of our study in several ways, given that our respondents were below the age of 18.

In regard to the baseline and endline surveys, we first secured independent consent from both parents/guardians and children for participation in both the surveys and associated classes. Furthermore, conducting the surveys within respondents' homes facilitated parental or adult supervision during the survey process. Enumerators were strictly instructed not to initiate any interviews without the presence of an adult, although in most instances, adults chose only to briefly be present or to not be there at all. Additionally, our survey partner, Sunai Consultancy, is known for its extensive experience in conducting surveys related to education on behalf of governmental and non-governmental organizations, such as Pratham. Consequently, the enumerators possessed considerable expertise in interviewing children as young as second graders. Despite this, we undertook a rigorous and comprehensive enumerator training, facilitated by the co-authors in the field. This training entailed observation of enumerators conducting simulated interviews, with a recurring emphasis on demonstrating empathy while posing survey questions, attentiveness to students' requirements, and the importance of halting or pausing interviews upon request.

Concerning teachers within the classroom setting, an elaborate module during our teacher training sessions extensively addressed issues of child safety and ethical conduct. Teachers were reminded of safety protocols, including refraining from unsupervised interactions with students outside of class hours, avoiding one-on-one meetings, and fostering an inclusive and non-discriminatory environment. A dedicated session addressed strategies for managing conflicts within the classroom, should any arise, such as instances of student-teacher or student-student disagreements. We also included detailed discussions on data privacy encompassing directives against soliciting personal information from students, refraining from photographing students for any purposes, refraining from putting up any information on social media, and establishing protocols for attendance data collection. Finally, teachers were required to sign a consent form acknowledging their participation in the training and were given physical copies of these guidelines (see Figure F13).

Figure F13: Ethics guidelines for teachers

GUIDELINES 2023-24

CHILD SAFETY GUIDELINES: BIHAR INFORMATION & MEDIA LITERACY INITIATIVE

Everything you need to know about working with children as part of this project is available in this document. The scope of this project defines anyone under the age of 18 as children. Trainers will conduct workshops for students in grades 8-12, in Hindi only.

- ➔ All communication regarding the training under this project will take place only with the parents/ legal guardian / nominated POC, not the children directly.
- ➔ The parents/legal guardians will be required to sign a mandatory consent form before collecting any personal information about the children.
- ➔ Only essential information (name and contact information) required for training participation will be collected.
- ➔ The program will ensure that the personal data (name and contact information) collected from the students and their parents are handled securely and in compliance with data protection regulations.
- ➔ The personal data collected will be retained only during the training period and will be deleted after the completion of training.
- ➔ A designated organizing team member will co-facilitate with trainers to meet child safety requirements.

DURING THE TRAINING

- ➔ There should always be a trainer/POC present acting as a responsible adult for the children during the training workshop.
- ➔ There will never be any unsupervised contact between trainers and school children without an organizing team member or parent's presence.
- ➔ Trainers will promote an inclusive and non-discriminatory environment during the training sessions. Ensure that all students feel respected and valued regardless of their socio-economic background or abilities.
- ➔ Ensure the maintenance of appropriate behavior and professionalism while interacting with the students.
- ➔ No personal contact details (email or mobile number) shall be collected from children in any form during the training sessions.
- ➔ Trainers and stakeholders to respect the privacy and confidentiality of the students. Personal information and any disclosure made during the sessions should be handled discreetly and shared only with relevant authorities when necessary.
- ➔ Students should be encouraged to not take photographs of other students, trainers, team members and must not share them on any platform.

SOCIAL MEDIA RULES

- ➔ Students are not encouraged to click any screenshot/pictures/videos during the training session and post them on social media

AFTER TRAINING

- ➔ The program will strictly not collect children's personal contact details (email or mobile number) in any form

In case of any violation of the above-mentioned rules, the aggrieved party can report the issue to the organizing team member present at the training location or directly reach out to below-mentioned authority.

Nishant Attri
Project Lead, DataLEADS
nishant@dataleads.co.in

Surabhi Jajodia
Outreach Lead, DataLEADS
sjajodia@dataleads.co.in

G Main effects tabular results

Table G10: Accuracy Discernment

Outcome	Type	N	Estimate	SE	p-value	Ctrl. Mean	Ctrl. SD
Index	—	11,121	0.32***	0.02	<0.001	0.00	1.00
Accuracy of True Statements	Sub-index	11,932	-0.08***	0.02	<0.001	0.00	1.00
Inaccuracy of False Statements	Sub-index	11,158	0.50***	0.03	<0.001	0.00	1.00
Inaccuracy: False Statement1	Question	11,570	0.44***	0.03	<0.001	2.49	1.21
Inaccuracy: False Statement2	Question	11,615	0.30***	0.02	<0.001	1.58	0.88
Inaccuracy: False Statement3	Question	11,814	0.18***	0.02	<0.001	1.81	1.02
Inaccuracy: False Statement4	Question	11,971	0.40***	0.03	<0.001	2.50	1.29
Accuracy: True Statement1	Question	11,995	-0.03**	0.01	0.002	-1.15	0.47
Accuracy: True Statement2	Question	11,972	0.00	0.01	0.701	-1.26	0.61
Accuracy: True Statement3	Question	11,994	-0.03***	0.01	<0.001	-1.10	0.35
Accuracy: True Statement4	Question	11,984	-0.04*	0.01	0.012	-1.26	0.73

Note: This table reports treatment effects on accuracy discernment, including the index as well as sub-index items. ICW was used to construct indices. Estimated using OLS regressions with library-spillover strata fixed effects and standard-errors clustered at the classroom level. All indices are standardized with respect to the control group mean and standard deviation. *p<0.05; **p<0.01; ***p<0.001.

Table G11: Sharing Discernment

Outcome	Type	N	Estimate	SE	p-value	Ctrl. Mean	Ctrl. SD
Index	—	11,263	0.21***	0.02	<0.001	0.00	1.00
Share True Statements	Sub-index	11,894	0.00	0.02	>0.9	-0.00	1.00
Not Share False Statements	Sub-index	11,319	0.24***	0.03	<0.001	-0.00	1.00
Not Share False Statement1	Question	11,650	0.11***	0.02	<0.001	1.71	0.69
Not Share False Statement2	Question	11,709	0.11***	0.01	<0.001	1.34	0.66
Not Share False Statement3	Question	11,861	0.08***	0.02	<0.001	1.44	0.69
Not Share False Statement4	Question	11,954	0.10***	0.01	<0.001	1.64	0.64
Share True Statement1	Question	11,990	-0.01	0.01	0.103	-1.10	0.42
Share True Statement2	Question	11,948	0.00	0.01	0.768	-1.17	0.52
Share True Statement3	Question	11,989	0.00	0.01	>0.9	-1.10	0.41
Share True Statement4	Question	11,970	0.01	0.01	0.266	-1.22	0.52

Note: This table reports treatment effects on sharing discernment, including the index as well as sub-index items. ICW was used to construct indices. Estimated using OLS regressions with library-spillover strata fixed effects and standard-errors clustered at the classroom level. All indices are standardized with respect to the control group mean and standard deviation. *p<0.05; **p<0.01; ***p<0.001.

Table G12: Source Discernment

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**	**Ctrl. Mean**	**Ctrl. SD**
Index	—	11,452	0.21***	0.02	<0.001	-0.00	1.00
Trust Reliable Sources (generic)	Sub-index	11,922	0.02	0.02	0.304	0.00	1.00
Distrust Unreliable Sources (generic)	Sub-index	11,893	0.10***	0.02	<0.001	-0.00	1.00
Generic Trust: MBBS Doctors	Question	11,976	0.00	0.01	0.806	-1.27	0.56
Generic Trust: Health Workers	Question	11,972	0.03*	0.01	0.015	-1.65	0.71
Generic Trust: Unqualified Practitioner	Question	11,960	0.04*	0.02	0.048	2.32	0.94
Generic Trust: Word of Mouth	Question	11,960	0.08***	0.02	<0.001	1.85	0.79
Generic Trust: Government pamphlets	Question	11,973	-0.01	0.02	0.659	-1.67	0.75
Generic Trust: Ayurvedic Doctors	Question	11,970	0.04**	0.01	0.005	1.61	0.69
Trust Reliable Sources (specific)	Sub-index	11,892	-0.04	0.02	0.077	-0.00	1.00
Distrust Unreliable Sources (specific)	Sub-index	11,899	0.27***	0.02	<0.001	0.00	1.00
Specific Trust: Health Workers	Question	11,950	0.02	0.01	0.132	-1.60	0.69
Specific Trust: Gov Pamphlets	Question	11,969	-0.05**	0.01	0.001	-1.70	0.72
Specific Trust: AIIMS Doctors	Question	11,968	-0.04*	0.02	0.012	-1.62	0.73
Specific Trust: Traditional Remedies	Question	11,967	0.19***	0.02	<0.001	1.68	0.79
Specific Trust: WhatsApp	Question	11,961	0.22***	0.02	<0.001	2.02	0.87
Specific Trust: Ayurvedic Doctors	Question	11,971	0.06***	0.02	<0.001	1.81	0.76
Cues: Reputable Source	Question	11,880	-0.08***	0.01	<0.001	-1.60	0.60
Cues: Sensational Tone	Question	11,858	-0.09***	0.01	<0.001	-1.73	0.63
Cues: Likes and Shares	Question	11,902	0.14***	0.01	<0.001	1.67	0.63
Cues: Same Community	Question	11,897	0.13***	0.01	<0.001	1.61	0.61
Cues	Sub-index	11,656	0.09***	0.02	<0.001	0.00	1.00

Note: This table reports treatment effects on source discernment, including the index as well as sub-index items. ICW was used to construct indices. Estimated using OLS regressions with library-spillover strata fixed effects and standard-errors clustered at the classroom level. All indices are standardized with respect to the control group mean and standard deviation. *p<0.05; **p<0.01; ***p<0.001.

Table G13: Health Preferences

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**	**Ctrl. Mean**	**Ctrl. SD**
Index	—	11,309	0.21***	0.02	<0.001	0.00	1.00
Interest in Health News	Question	12,007	0.03	0.02	0.117	0.00	1.00
Perception of Vaccine Safety	Sub-index	11,471	0.07***	0.02	<0.001	0.00	1.00
COVID Vaccine Safety	Question	11,943	0.02**	0.01	0.001	-1.12	0.37
Chicken Pox Vaccine Safety	Question	11,513	0.03*	0.01	0.028	-1.31	0.58
Distrust Non-Scientific Remedies	Sub-index	11,798	0.23***	0.02	<0.001	-0.00	1.00
Not Use Traditional Remedies	Question	11,921	0.14***	0.01	<0.001	1.61	0.59
Not Go to Traditional Healer	Question	11,949	0.05***	0.02	<0.001	1.69	0.67
Ayurveda/Homeopathy Ineffective	Question	11,924	0.11***	0.02	<0.001	1.61	0.78

Note: This table reports treatment effects on health preferences, including the index as well as sub-index items. ICW was used to construct indices. Estimated using OLS regressions with library-spillover strata fixed effects and standard-errors clustered at the classroom level. All indices are standardized with respect to the control group mean and standard deviation. *p<0.05; **p<0.01; ***p<0.001.

Table G14: Engagement Attitude

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**	**Ctrl. Mean**	**Ctrl. SD**
Index	—	11,868	0.10***	0.02	<0.001	-0.00	1.00
Correct Friend	Question	11,947	0.01	0.02	0.796	-0.00	1.00
Not Share Friend's Misinformation	Question	11,917	0.11***	0.02	<0.001	0.00	1.00
Fact Checking Important	Question	12,007	-0.02	0.02	0.243	0.00	1.00
Fact Checked Recently	Question	12,007	0.10***	0.02	<0.001	0.00	1.00

Note: This table reports treatment effects on engagement attitude, including the index as well as sub-index items. ICW was used to construct indices. Estimated using OLS regressions with library-spillover strata fixed effects and standard-errors clustered at the classroom level. All indices are standardized with respect to the control group mean and standard deviation. *p<0.05; **p<0.01; ***p<0.001.

Table G15: Engagement Behavior

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**	**Ctrl. Mean**	**Ctrl. SD**
Index	—	12,007	0.02	0.02	0.374	0.00	1.00
Prefer Hindustan Newspaper	Question	12,007	0.02	0.02	0.337	-0.00	1.00
Willing to Be a Truth Ambassador	Question	12,007	0.01	0.02	0.719	-0.00	1.00

Note: This table reports treatment effects on engagement behavior, including the index as well as sub-index items. ICW was used to construct indices. Estimated using OLS regressions with library-spillover strata fixed effects and standard-errors clustered at the classroom level. All indices are standardized with respect to the control group mean and standard deviation. *p<0.05; **p<0.01; ***p<0.001.

Table G16: Awareness

Outcome	Type	N	Estimate	SE	p-value	Ctrl. Mean	Ctrl. SD
Index	—	11,781	-0.01	0.02	0.641	-0.00	1.00
Misinfo Threat Perception	Question	11,947	0.10***	0.02	<0.001	-0.00	1.00
News Manipulation Awareness	Sub-index	11,909	-0.14***	0.02	<0.001	-0.00	1.00
News Manipulation Awareness1	Question	11,958	-0.12***	0.02	<0.001	0.00	1.00
News Manipulation Awareness2	Question	11,947	-0.11***	0.02	<0.001	-0.00	1.00
Confirmation Bias Awareness	Sub-index	11,905	-0.02	0.02	0.326	0.00	1.00
Confirmation Bias Awareness1	Question	11,955	0.00	0.02	0.866	-1.91	1.04
Confirmation Bias Awareness2	Question	11,950	-0.03	0.02	0.145	-1.69	0.88

Note: This table reports treatment effects on awareness, including the index as well as sub-index items. ICW was used to construct indices. Estimated using OLS regressions with library-spillover strata fixed effects and standard-errors clustered at the classroom level. All indices are standardized with respect to the control group mean and standard deviation. *p<0.05; **p<0.01; ***p<0.001.

G.1 Treatment effect in control-group means

Table G17: Effects in % of Control Group Means

Outcome	Control Mean	Treatment Mean	% Treatment Effect
Accuracy Discernment	1.01	1.36	34.61
Sharing Discernment	0.64	0.81	27.32
Source Discernment	0.51	0.72	42.05
Health Preferences	0.88	1.01	15.11
Engagement Attitude	-1.05	-1.00	4.71
Engagement Behavior	-2.03	-2.01	1.01
Awareness	-1.06	-1.11	-5.16

The treatment effect is expressed as a percentage change relative to the control group mean. For each index, we first rescale all contributing items to range from 0 to 1 based on the observed minimum and maximum in the data. Composite indices are then computed by summing the relevant rescaled items. The percentage effect is calculated as: $\frac{\text{Treatment Mean}-\text{Control Mean}}{|\text{Control Mean}|} * 100$.

In Table G17, we calculate the treatment effect in % of control-group means. Due to the complex nature of how we construct outcome indices (as per our pre-analysis plan), we proceed as follows. First, to account for the fact that variables are measured on different scales, we rescale all variables that constitute part of an outcome measure to range from 0 to 1 based on the observed minima and maxima in the data. We recode variables so that higher values always indicate 'better' outcomes. Next, for each individual respondent and per outcome family, we sum all variables before taking the control- and treatment-group means. We then calculate the percentage effect as $\frac{\text{Treatment Mean}-\text{Control Mean}}{|\text{Control Mean}|} * 100$.

This analysis offers a simplified summary of treatment effects across outcome domains and can be useful for descriptive reporting and for communicating results to non-technical audiences. However, it abstracts away from the more rigorous procedures used in our main outcome indices, which account for variation in item importance, measurement reliability, and the covariance structure of items within each outcome family. Specifically, the main indices use inverse-covariance weighting and standardization based on the control group distribution to ensure statistical comparability and robustness. The simplified method presented here does not adjust for differences in item scale reliability or inter-item correlations, and thus may overweight or underweight certain items. As such, this analysis should be viewed as a complementary descriptive summary rather than a substitute for the main results.

H Heterogeneous treatment effects

Table H18 presents results from the following specification:

$$Y_{ijk} = \beta_0 + \beta_1 T_{ijk} + \beta_2 Z_{ijk} + \beta_3 (T_{ijk} \times Z_{ijk}) + \sum_{k=1}^{m-1} \gamma_k + \varepsilon_{ijk} \tag{1}$$

where Z is the pre-treatment covariate hypothesized to moderate the treatment effect. β_3 represents the coefficient of the interaction term, and β_2 captures the estimated direct effect of Z on the outcome Y.

In Table H18, we present the coefficients for the interaction term across multiple pre-registered pre-treatment covariates. These include individual characteristics such as age, grade, gender, caste and religion, access to and use of media (measured using an index comprising mobile ownership, internet usage on mobile phones, and exposure to media outlets), prior attachment to non-scientific belief systems (measured using an index of responses regarding the effectiveness of Ayurveda and preferences for non-scientific treatments), asset index (measured through household asset ownership), an indicator for low-spillover status, and finally party affiliation estimated using an indicator for non-BJP party affiliation (see below for more details on party ID estimation). The table does not show any systematic patterns. For most outcomes and pre-treatment covariates, we do not observe statistically significant effects. However, the effect on engagement attitude is lower for girls, non-BJP affiliation is associated with a higher effect on awareness, and three outcomes are associated with a higher effect in the low-spillover stratum.

In Figure H14 we look at subgroup ITT effects by student grade, ranging from grades 8 to 12, to see if effects are concentrated within certain grades.

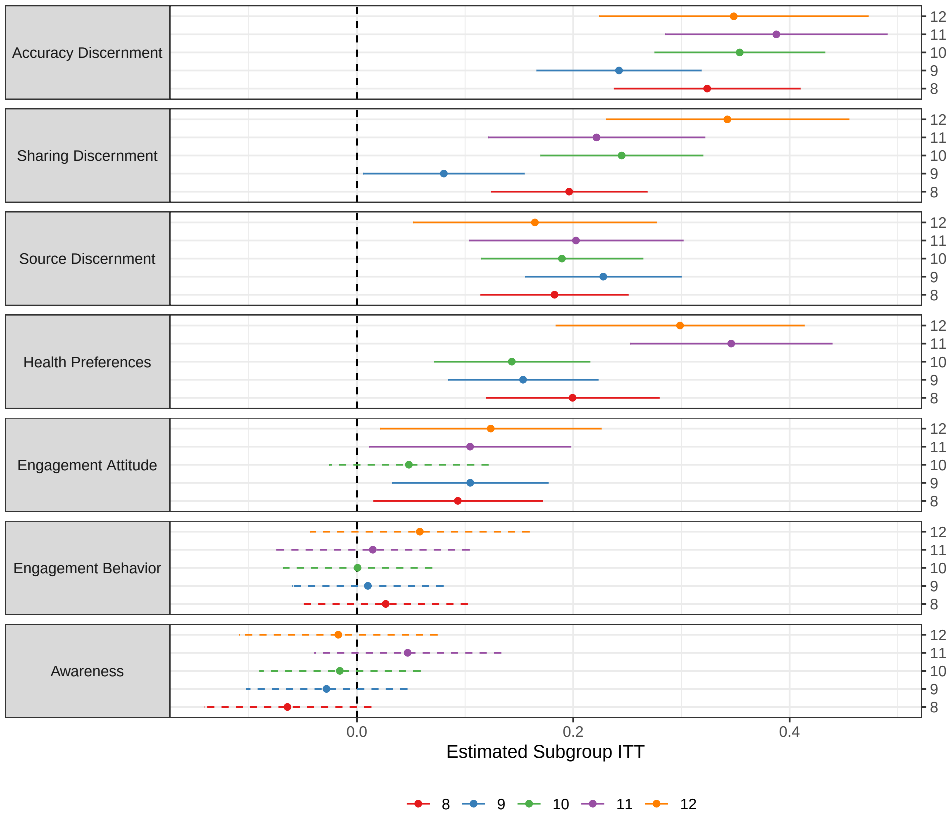
H.1 Estimating party ID

Due to our collaboration with the government, we were unable to include questions about party identification in our baseline survey. Therefore, we estimate party identification at the household level. To do this, we conducted a survey with local elites to identify the party affiliations of prominent sub-castes and communities (*jatis*) in each village. Previous work on India shows that voters from a *jati* within a village often coordinate support for the same party (Jaffrelot, 2013, 2023; Biswas, 2023; Blair, 1972). Thus, under the assumption that voters from the same *jati* typically back the same party, we estimate the party affiliation of each household in our sample.

We conducted this local elite survey independently of the household-level baseline survey. Overall, we identified and surveyed 1,664 elites across 550 villages in our sample. In each village, we interviewed at least three local elites, each belonging to a different caste category. Respondents were required to be residents of the village and could not belong to households already surveyed in our study.

Enumerators were instructed to survey informal leaders using a snowball sampling method. These informal leaders were village residents and could have one of a number of specific

Figure H14: Subgroup ITT by student grade



roles or distinctions: local government employees (*gram sevak*), grassroots implementation/facilitation workers (*vikas mitra*), child care center (*anganwadi*) workers, healthcare workers (*arogya sevika*), public distribution system shopkeepers, self-help group leaders, village revenue officers (*talathi*), police officers (*kotwal*), government employees not affiliated with the gram panchayat, and other informal leaders (e.g., religious leaders or caste panchayat leaders). We began by asking each local elite about their party preference, followed by a request to identify the largest *jatis* in the village and to report the voting preferences of these communities in the most recent state and national elections. Typically, respondents listed three jatis and their associated party preferences.

To integrate the data, we merged the party affiliation data with our baseline survey using the following steps:

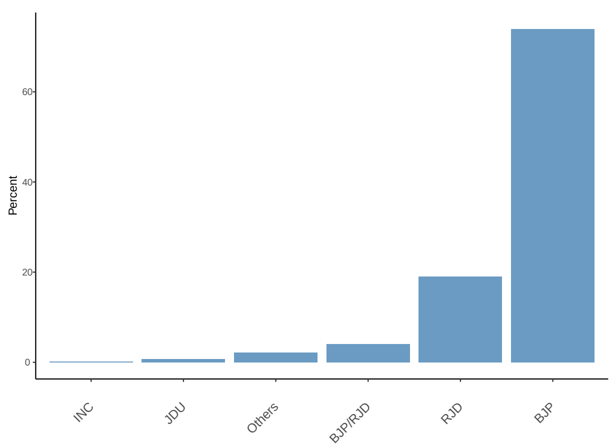
- **Standardizing *jati* names:** We standardized the *jati* names across both the elite survey and the baseline dataset. Some observations were lost during this process due to incomplete or unstandardized *jati* names; this issue was more prevalent in the baseline data than in the elite survey.
- **Assigning party ID:** The elite survey asked three local elites about the party preferences of *jatis* in the most recent state and national elections. For example, if all three

elites (A, B, and C) agreed that *jati* X voted for the BJP in the state elections, then we assigned BJP as the party ID for *jati* X in that village. If two elites mentioned BJP and one mentioned Congress, BJP was still assigned since the majority of elites choose BJP. In cases where all three elites provided different answers (e.g., BJP, Congress, and RJD), we assigned a mixed party ID, such as "BJP/Congress/RJD."

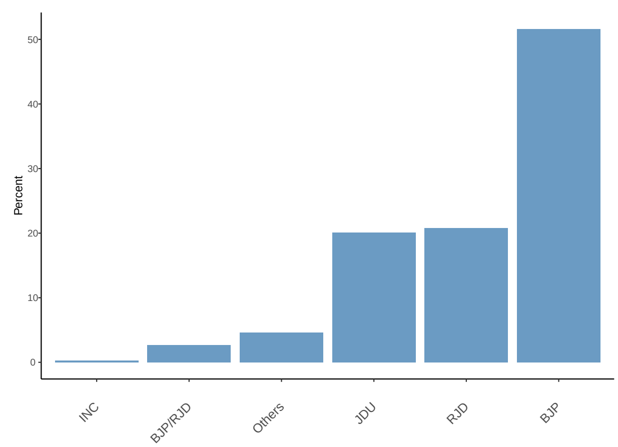
- **Village-*jati* dataframe:** Finally, we created a village-*jati* level dataframe with the assigned party IDs, which was then merged with our baseline data using village and *jati* identifiers to estimate the party preference of households in our sample.

Figure H15 illustrates the distribution of party preferences in the sample. It shows that respondents in most villages tended to favor a single party, with the BJP being the most preferred (especially in national elections), followed by the RJD (in national elections) and both the RJD and JDU (in state elections). A few village-jatis exhibited mixed party identification, such as support for both the BJP and RJD, but this was uncommon.

Figure H15: Party preference distribution



(a) 2019 National Elections



(b) 2020 State Elections

Figure H16 illustrates the party preferences of major *jatis* in our sample villages. The results align with our expectations for Bihar, confirming the widely held understanding that the RJD is perceived as an ethnic party, with Yadavs forming its primary vote bank. Figure H17 shows the distribution of preferences for the BJP and RJD (the major parties) across 100 library villages, with each bar representing a separate library location.

Figure H16: Party preference of major Jatis

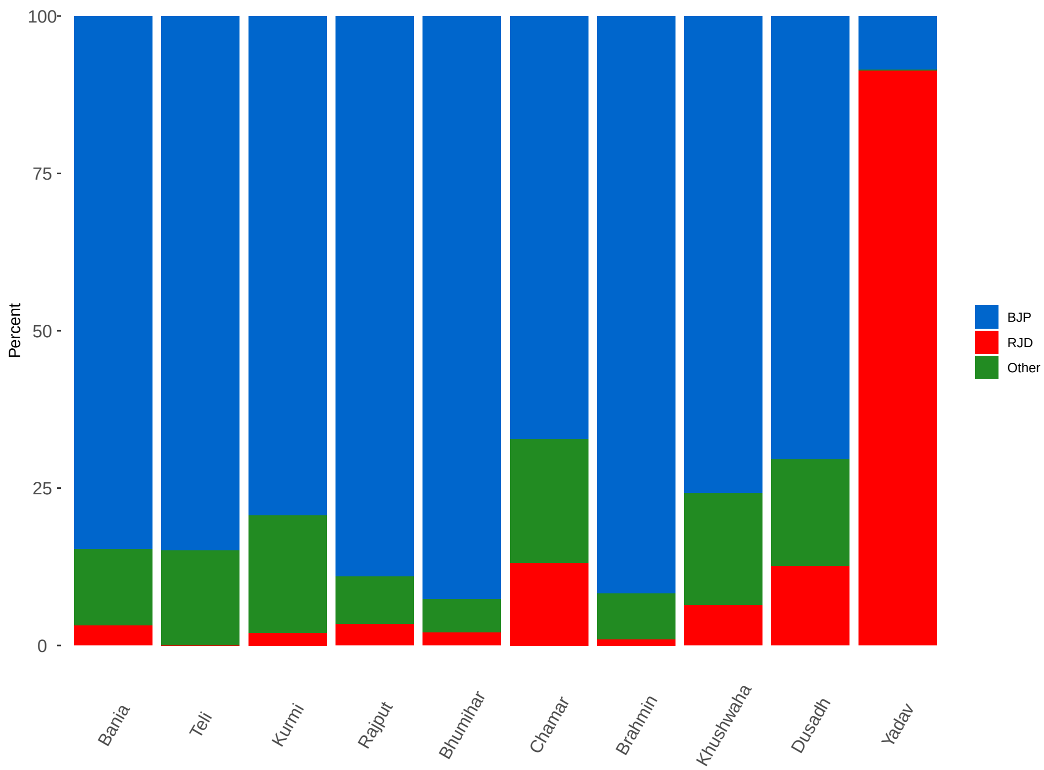


Figure H17: Party preference by library

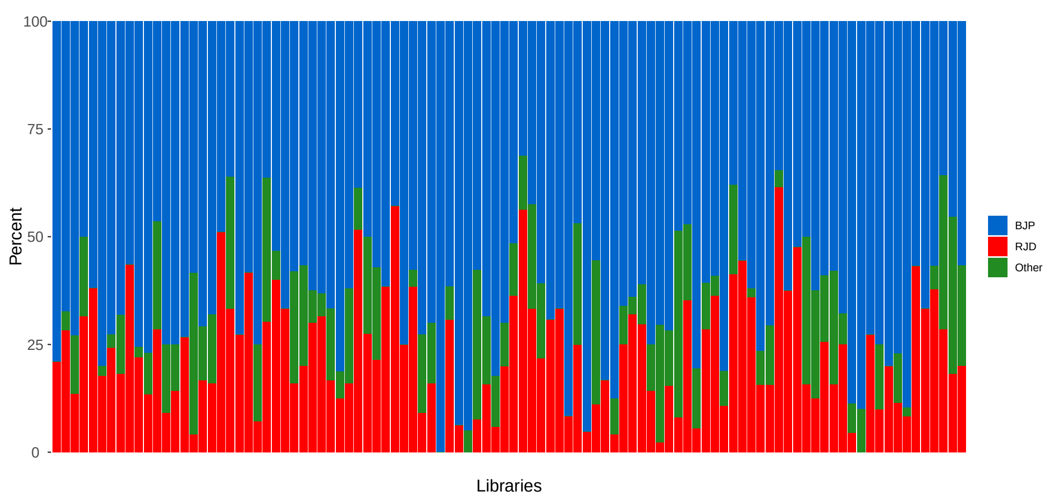


Table H18: Heterogeneity results

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
T x Age	0.014	0.012	0.009	0.006	-0.009	-0.003	0.007
	(0.014)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)
N	11121	11263	11309	11452	11868	12007	11781
T x Grade	0.012	0.026	0.020	0.008	-0.002	-0.008	0.011
	(0.016)	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)	(0.013)
N	11120	11262	11309	11451	11867	12006	11781
T x Girls	0.016	0.011	0.058	0.023	-0.135***	-0.054	0.024
	(0.044)	(0.041)	(0.038)	(0.042)	(0.040)	(0.039)	(0.036)
N	11121	11263	11309	11452	11868	12007	11781
T x Non-scientific beliefs	-0.017	-0.010	-0.001	-0.000	0.016	0.015	0.004
	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)	(0.014)
N	11016	11154	11198	11340	11750	11887	11664
T x Prior media exposure	0.019	0.005	-0.015	0.010	0.034	0.012	0.002
	(0.021)	(0.021)	(0.018)	(0.019)	(0.018)	(0.019)	(0.019)
N	11044	11188	11238	11377	11788	11927	11704
T x Asset Index	0.027	-0.022	0.014	-0.014	0.036	0.024	-0.013
	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.020)
N	11121	11263	11309	11452	11868	12007	11781
T x Party ID (non-BJP)	-0.028	-0.094	-0.010	0.059	0.021	-0.014	0.109*
	(0.065)	(0.068)	(0.060)	(0.060)	(0.063)	(0.057)	(0.055)
N	6944	7038	7026	7135	7421	7522	7373
T x Hindu	0.062	0.044	-0.000	0.031	-0.031	0.044	-0.075
	(0.078)	(0.075)	(0.081)	(0.065)	(0.070)	(0.070)	(0.074)
N	11121	11263	11309	11452	11868	12007	11781
T x Gen. caste	-0.020	0.019	-0.141	-0.058	-0.045	-0.008	-0.052
	(0.083)	(0.073)	(0.076)	(0.082)	(0.072)	(0.068)	(0.071)
N	11121	11263	11309	11452	11868	12007	11781
T x Low-spillover	0.011	-0.042	0.015	0.097*	0.101*	0.035	0.123*
	(0.047)	(0.046)	(0.042)	(0.041)	(0.047)	(0.039)	(0.048)
N	11121	11263	11309	11452	11868	12007	11781
Library-Spillover FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

H.2 Gender engagement outcomes

Table H19: Engagement indices means by gender (95% CIs in parentheses)

Variable	Group	Boys (mean)	Girls (mean)
Engagement Attitude	Control	0.014 (-0.047, 0.075)	-0.010 (-0.070, 0.051)
Engagement Attitude	Treatment	0.200 (0.138, 0.261)	0.041 (-0.019, 0.100)
Engagement Behavior	Control	0.070 (0.024, 0.116)	-0.049 (-0.107, 0.009)
Engagement Behavior	Treatment	0.135 (0.095, 0.176)	-0.030 (-0.080, 0.020)

I Second endline

I.1 Sampling and descriptive results

To sample for the the follow-up survey, we targeted respondents following a randomized order assigned within each village in the sample. Specifically, each respondent was randomly assigned a number corresponding to their calling order. Enumerators were instructed to strictly follow this order, beginning with the respondent assigned number 1 in each village and proceeding sequentially. To reach our final target sample of 2000, we randomized whether 3 or 4 subjects were to be called within each village.

The surveyors were instructed to proceed as follows. If a given village was randomly selected to contribute 3 subjects to the sample, the survey firm tried to reach the first 3 students in the list, based on the randomized calling order. If one of those 3 students was not reachable, the surveyors were instructed to try once more. If still unable to reach the target sample size, the surveyors moved down the list, calling students 4, 5, 6, etc. In villages with a target sample of 4, the first 4 students were called twice if not reachable initially, all subsequent students were only called once. The final follow-up sample was N=2059.

In the second endline survey, we included several questions aimed at understanding mechanisms and self-reported reasons for participating in the program. For parents, we inquired about their primary reasons for enrolling their child in the program (see Table I20). For respondents themselves, we sought to understand their self-reported reactions toward individuals sharing misinformation. Specifically, we asked: If someone you knew told you a piece of information that you know to be false/untrue, what would be your primary reaction to them? The results are presented in Table I21.

Table I20: Parents' reasons for sending child to classes

Reason	Percent
Wanted child to learn	70.68
Trust Jeevika	26.65
Cost free	1.59
Wanted child out of house	1.08

Table I21: Reactions to misinformation

Reaction	Control	Treatment	statistic	p-value
Emphasize that it is false	0.35	0.17	-11.43	< 0.001
Teach strategy to verify	0.22	0.30	4.88	< 0.001
Admonish for spreading false info	0.13	0.13	0.01	1
Emphasize not sharing	0.30	0.40	5.49	< 0.001

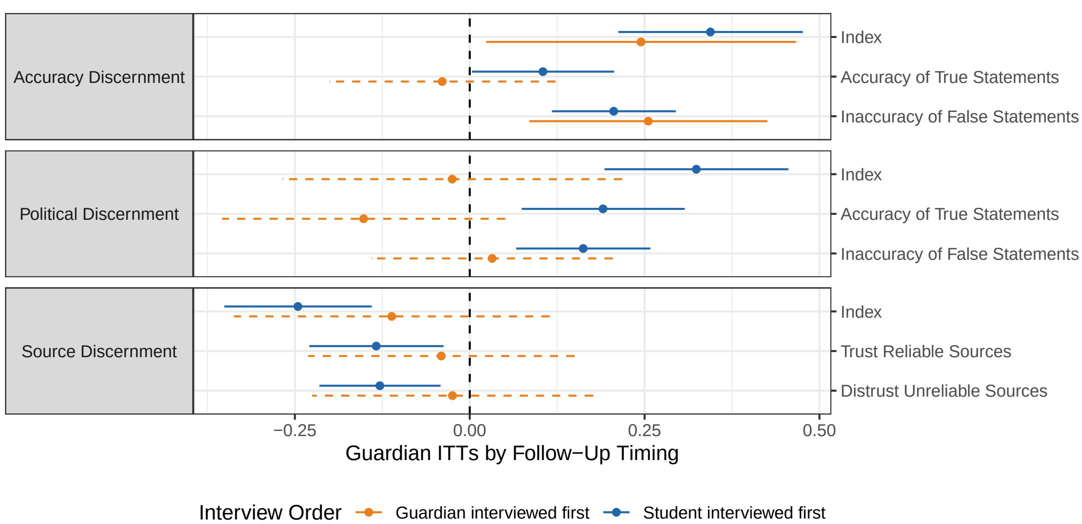
I.2 Spillovers to Guardians: Additional Results

In Figure I18, we present ITT results for guardians in the second endline survey, subset by whether the guardian or student was interviewed first. We do so because we were concerned that any treatment diffusion from students to guardians we observe might be caused by guardians repeating answers overheard from student respondents.

These results have to be interpreted with caution for a number of reasons. First, we did not randomly assign interview order, so we cannot be confident that whether the student or guardian was interviewed first did not correlate with respondents' potential outcomes. Second, the subset of the sample among which guardians were interviewed first is significantly smaller ($N = 592$) than the subset for which students were interviewed first ($N = 1397$).

Regardless, Figure I18 provides suggestive evidence that only the result on 'accuracy discernment' (health stories) is robust to accounting for potential repeating, which is why we only report this finding in the main paper. There is some evidence of a positive effect on political discernment and a negative effect on source discernment among the subset of guardians that were interviewed after students, but for the reasons above, we urge caution in over-interpreting these results.

Figure I18: Subgroup guardian ITT by survey timing



Notes: Figure displays ITTs for guardians in the second endline survey, subset by whether guardians or students were interviewed first.

J Robustness checks

J.1 Complier average causal effects

In Table J22, we present results for the 2SLS model in the CACE framework, which isolates the local average treatment effect (LATE) for compliers by using assignment as a source of variation in treatment. We use a minimalist definition of compliance that takes the value 1 if the student was assigned to treatment and attended at least one session, and 0 otherwise (measured with respondent-level attendance data gathered by teachers during each session).

- First stage (predicting treatment using the IV): The first-stage outcome measures treatment uptake (attending $\geq$1 session) among those assigned to treatment, with all control units coded as zero.
- Second stage (estimating the CACE): Predicted values from the first stage are used in place of the actual treatment variable in the outcome regression. This step estimates the effect of the treatment on the outcome for compliers only, estimating the expected change in the outcomes when attending any versus 0 sessions.

The model specification for the two stages are below:

$$D_{ijk} = \alpha_0 + \alpha_1 T_{ijk} + \sum_{k=1}^{m-1} \gamma_k + \varepsilon_{ijk} \tag{2}$$

where D is an indicator for whether at least one treatment session was attended and T is the treatment indicator.

$$Y_{ijk} = \beta_0 + \beta_1 \hat{D}_{ijk} + \sum_{k=1}^{m-1} \gamma_k + \upsilon_{ijk} \tag{3}$$

where $\hat{D}$ is the predicted share of sessions from the first stage and β_1 is the CACE estimate.

Table J22: Complier Average Causal Effects (CACE)

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.343***	0.225***	0.226***	0.229***	0.115***	0.020	-0.012
	(0.0255)	(0.0245)	(0.0229)	(0.0221)	(0.0251)	(0.0223)	(0.0257)
N	11121	11263	11309	11452	11868	12007	11781
R2	0.158	0.100	0.087	0.091	0.146	0.079	0.147
Library-Spillover FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*p<0.05; **p<0.01; ***p<0.001. SEs clustered at the village level in parentheses.

Further in Table J23 we also compute compliance as the estimated effect of assignment to treatment on outcomes conditional on attendance at the first session. This is reasonable in our context, as students were unaware of their treatment status or the content of the sessions prior to attending the first class. As a result, the proportion of never-takers should not differ systematically across treatment and control groups for first session attendees.

Table J23: ITT on Session 1 Attendees

Outcome	**N**	**Estimate**	**SE**	**p-value**
Accuracy Discernment	8,891	0.38***	0.025	<0.001
Awareness	9,393	0.01	0.025	0.82
Engagement Attitude	9,457	0.13***	0.024	<0.001
Engagement Behavior	9,562	0.03	0.022	0.19
Health Preferences	9,037	0.25***	0.023	<0.001
Sharing Discernment	9,009	0.23***	0.025	<0.001
Source Discernment	9,154	0.24***	0.023	<0.001

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

J.2 Alternative specifications

Here we present the main results using alternate specifications as robustness checks. First, Table J24 shows results from a specification that includes library fixed effects to control for characteristics at the time of the intervention, such as library infrastructure and staff cooperation. The results show that the main findings are robust to this alternate specification.

Table J24: Main results with library FE

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.314***	0.221***	0.206***	0.209***	0.103**	0.036	-0.008
	(0.0390)	(0.0325)	(0.0303)	(0.0310)	(0.0378)	(0.0296)	(0.0382)
N	11121	11263	11309	11452	11868	12007	11781
R2	0.028	0.015	0.010	0.010	0.005	0.001	0.004
Library FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*p<0.05; **p<0.01; ***p<0.001. SEs clustered at the village level in parentheses.

Next, Tables J25 and J26 report results from specifications that include district fixed effects and district-spillover stratum fixed effects, respectively, to account for differences between districts and differences between district-spillover strata. Our main results are robust to these specifications.

Lastly, Table J27 presents the results for the main ITT effects including a range of control variables we pre-specified. The model includes the following baseline controls: individual

Table J25: Main results with district FE

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.311***	0.217***	0.205***	0.208***	0.104***	0.032	-0.008
	(0.0284)	(0.0276)	(0.0254)	(0.0252)	(0.0289)	(0.0248)	(0.0309)
N	11121	11263	11309	11452	11868	12007	11781
R2	0.113	0.054	0.049	0.053	0.088	0.041	0.076
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*p<0.05; **p<0.01; ***p<0.001. SEs clustered at the village level in parentheses.

Table J26: Main results with district-spillover FE

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.315***	0.217***	0.207***	0.201***	0.105***	0.021	-0.009
	(0.0282)	(0.0271)	(0.0249)	(0.0248)	(0.0288)	(0.0237)	(0.0302)
N	11121	11263	11309	11452	11868	12007	11781
R2	0.117	0.058	0.054	0.056	0.092	0.047	0.084
District-Spillover FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*p<0.05; **p<0.01; ***p<0.001. SEs clustered at the village level in parentheses.

characteristics such as age, gender, medium of education, grade in school, reading skill and science skill indices, and prior exposure to mobile internet; household characteristics such as asset ownership, religion, party ID and caste category; and village-level characteristics such as development (proxied by nighttime lights data) and BJP vote share at the assembly constituency level.

Several control variables, including age, an indicator for gender (girls), an indicator for Hindi as the medium of education, grade, reading skill index, science skill index, prior exposure to mobile internet, and household characteristics like asset index, an indicator for Hindus, and caste category fixed effects (with the general category as the omitted category), are constructed using responses from the baseline survey. Among these, we define the asset index as a continuous measure ranging from 0 to 1, where 1 indicates that the household owns all 16 assets we asked about (a car, scooter, air-conditioner, computer, phone, WiFi connection, electric fan, washing machine, and fridge, television, bank account, ATM, LPG connection, working toilet, pumping set, tractor), and 0 indicates that the household owns none of these assets. Intermediate values reflect partial ownership of assets. The reading skill index is an additive index that can take values of 0, 1, or 2 depending on the number of correct responses to two reading comprehension questions in the baseline. Similarly, the science skill index is an additive index that takes integer values from 0 to 8 based on the correct responses to basic science questions. Lastly, prior exposure to mobile internet is measured as an indicator variable that takes a value of 1 if the response to the baseline survey question, "Do you use the internet on your mobile phone?" is "yes," and 0 if "no."

For village characteristics, we include a measure of party preference, BJP+ vote share, which measures the vote share of the BJP and its coalition partners in the 2020 Bihar state elections for each assembly constituency. Since vote share data is at the assembly constituency level and there are many villages in a constituency, we assign the same vote share to all villages within an assembly constituency. For this variable, we rely on assembly constituency data for the 2020 elections in Bihar, compiled by the Trivedi Center for Political Data (Agarwal et al., 2021). Next, to proxy development, we rely on the time series of annual global Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime lights compiled by the National Oceanic and Atmospheric Administration. We accessed these datasets via the Socioeconomic High-resolution Rural-Urban Geographic Platform for data on India (Asher et al., 2021). This data was merged with the villages in our sample. Village nighttime lights provide a measure of the average luminosity in 2021 for each village in our sample that could be mapped to a census village. The results show that the main findings are robust to an alternate specification that includes controls for individual, household, and village-specific characteristics.

Table J27: Main results (adjusted model)

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.320***	0.212***	0.206***	0.212***	0.101***	0.001	-0.022
	(0.0237)	(0.0225)	(0.0214)	(0.0205)	(0.0238)	(0.0196)	(0.0249)
Age	-0.023*	-0.030**	-0.021*	-0.008	-0.012	0.005	-0.013
	(0.0107)	(0.0097)	(0.0100)	(0.0107)	(0.0098)	(0.0090)	(0.0094)
Girl	-0.155***	-0.136***	0.012	-0.063**	-0.067**	-0.114***	-0.027
	(0.0236)	(0.0221)	(0.0198)	(0.0219)	(0.0212)	(0.0206)	(0.0196)
Grade	0.056***	0.054***	0.080***	0.058***	0.041***	0.016	0.060***
	(0.0126)	(0.0125)	(0.0133)	(0.0130)	(0.0121)	(0.0115)	(0.0119)
OBC	0.036	-0.046	-0.082*	-0.065	0.003	0.001	-0.042
	(0.0431)	(0.0369)	(0.0403)	(0.0437)	(0.0365)	(0.0362)	(0.0411)
SC	-0.055	-0.116**	-0.101*	-0.109*	-0.010	0.038	-0.086
	(0.0488)	(0.0433)	(0.0457)	(0.0503)	(0.0422)	(0.0419)	(0.0469)
ST	-0.128	-0.137	-0.244**	-0.101	-0.092	0.075	-0.194*
	(0.0956)	(0.0776)	(0.0930)	(0.0734)	(0.0894)	(0.0764)	(0.0754)
Hindu	0.002	-0.002	0.068	0.052	-0.013	-0.064	0.007
	(0.0431)	(0.0411)	(0.0408)	(0.0388)	(0.0388)	(0.0377)	(0.0388)
Asset Index	0.085***	0.084***	0.023*	0.025*	0.027**	0.010	0.014
	(0.0113)	(0.0114)	(0.0112)	(0.0116)	(0.0102)	(0.0099)	(0.0105)
Hindi medium	-0.098	-0.117	-0.100	-0.142*	-0.114*	-0.124	-0.076
	(0.0538)	(0.0672)	(0.0648)	(0.0611)	(0.0483)	(0.0634)	(0.0477)
Reading skill index	0.110***	0.133***	0.099***	0.032	0.051**	-0.009	0.010
	(0.0218)	(0.0203)	(0.0216)	(0.0201)	(0.0185)	(0.0181)	(0.0194)
Science skill index	0.062***	0.053***	0.040***	0.056***	0.048***	0.012	0.030***
	(0.0081)	(0.0079)	(0.0083)	(0.0081)	(0.0073)	(0.0071)	(0.0077)
Mobile Internet	0.038	0.023	0.050	0.074**	0.066**	0.109***	0.035
	(0.0298)	(0.0271)	(0.0264)	(0.0284)	(0.0239)	(0.0260)	(0.0263)
BJP+ Vote Share	-0.004	-0.007	0.002	-0.007	-0.005	0.000	0.003
	(0.0050)	(0.0048)	(0.0039)	(0.0040)	(0.0041)	(0.0054)	(0.0047)
Village Nightlights	0.033	0.023	0.102***	0.038	-0.024	-0.016	0.016
	(0.0342)	(0.0354)	(0.0283)	(0.0346)	(0.0302)	(0.0256)	(0.0333)
N	10130	10266	10295	10427	10798	10919	10714
R2	0.187	0.129	0.106	0.108	0.160	0.085	0.157
Library-Spillover FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*p<0.05; **p<0.01; ***p<0.001. SEs clustered at the village level in parentheses.

J.3 Multiple hypotheses corrections

Adjusted p-values. To assuage concerns about multiple comparisons, Tables J28 and J29 each report three sets of p-values for Intent-to-Treat (ITT) and Complier Average Causal Effects (CACE) estimates on the main indices, respectively: (1) standard p-values based on the baseline specifications we pre-registered, (2) p-values adjusted for the false discovery rate (FDR), and (3) p-values adjusted using Bonferroni corrections. Our results are robust to these corrections: where we find effects in the baseline models, these effects remain statistically significant after adjusting for multiple corrections.

Table J28: Main results (ITT): correcting for multiple hypotheses

Outcome	**N**	**Estimate**	**SE**	**p**	**p (FDR)**	**p (Bonf.)**
Awareness	11,781	-0.01	0.024	0.64	0.64	>0.9
Accuracy Discernment	11,121	0.32	0.024	<0.001	<0.001	<0.001
Sharing Discernment	11,263	0.21	0.023	<0.001	<0.001	<0.001
Health Preferences	11,309	0.21	0.021	<0.001	<0.001	<0.001
Source Discernment	11,452	0.21	0.020	<0.001	<0.001	<0.001
Engagement Attitude	11,868	0.10	0.023	<0.001	<0.001	<0.001
Engagement Behavior	12,007	0.02	0.020	0.37	0.44	>0.9

P-values adjusted for False Discovery Rate (FDR) and using Bonferroni correction.

DV: Main Indices. Models include library-spillover FEs.

Table J29: Main results (CACE): correcting for multiple hypotheses

Outcome	**N**	**Estimate**	**SE**	**p**	**p (FDR)**	**p (Bonf.)**
Awareness	11,781	-0.01	0.026	0.64	0.64	>0.9
Accuracy Discernment	11,121	0.34	0.026	<0.001	<0.001	<0.001
Sharing Discernment	11,263	0.23	0.024	<0.001	<0.001	<0.001
Health Preferences	11,309	0.23	0.023	<0.001	<0.001	<0.001
Source Discernment	11,452	0.23	0.022	<0.001	<0.001	<0.001
Engagement Attitude	11,868	0.11	0.025	<0.001	<0.001	<0.001
Engagement Behavior	12,007	0.02	0.022	0.37	0.44	>0.9

P-values adjusted for False Discovery Rate (FDR) and using Bonferroni correction.

DV: Main Indices. Models include library-spillover FEs.

Index of Indices. In addition to the above analyses, we also summarize all outcome indices in a single 'index of indices.' We do so by computing an inverse-covariance weighted index of the respective indices of the main outcome families. We then standardize this summary index with respect to the control group and estimate the intent-to-treat effects of assignment to the treatment on this index. The results are presented in Tables J30 and J31 for the first endline and the follow-up surveys, respectively. Results are robust to this alternative specification, with ITT estimates of 0.27/0.34 in the first endline/follow-up survey.

Table J30: ITT effects for summary index of indices

Outcome	**N**	**Estimate**	**SE**	**p-value**
Index of Indices	9,883	0.27	0.026	<0.001

ITT effect estimates for index of indices.

DV: standardized, ICW-weighted index of main indices.

Table J31: Second endline ITT effects for summary index of indices

Outcome	**N**	**Estimate**	**SE**	**p-value**
Index of Indices	1,760	0.34	0.057	<0.001

ITT effect estimates for index of indices.

DV: standardized, ICW-weighted index of main indices.

J.4 Alternative discernment index

Table J32: Discernment (excluding 2 items)

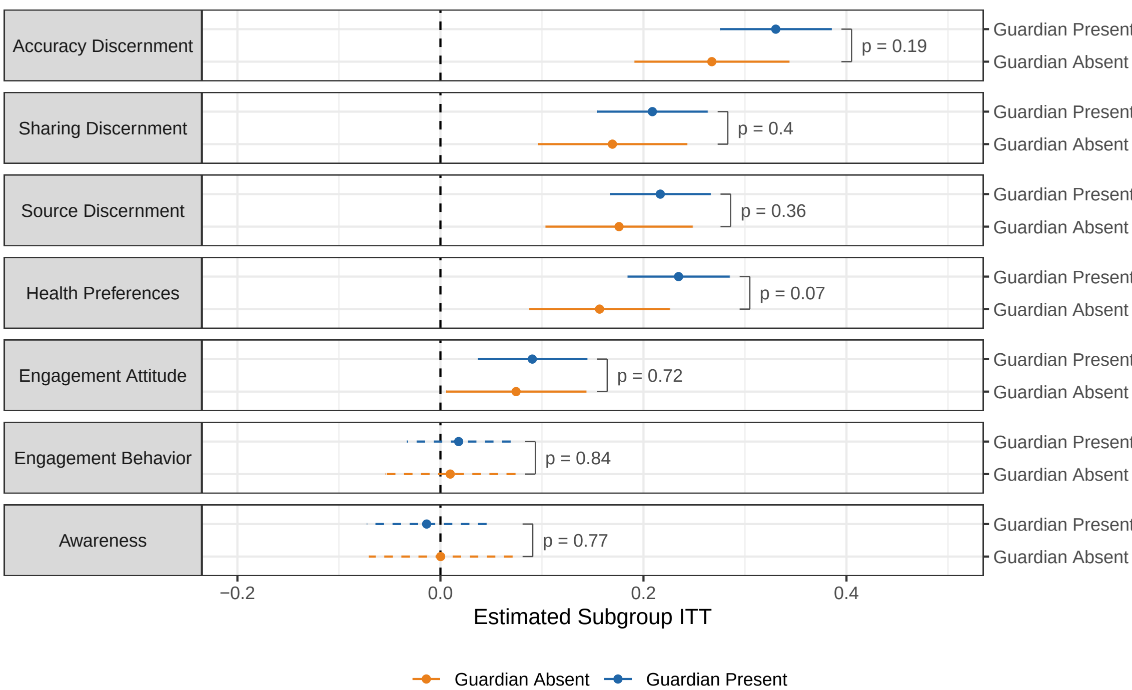
Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**
Accuracy Discernment	Index	11,149	0.27***	0.024	<0.001
Sharing Discernment	Index	11,312	0.18***	0.022	<0.001

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

J.5 Parental presence

To address concerns around parental presence prompting respondent answers during the first endline, we conducted subgroup ITT analyses based on the number of individuals present during the interview. Specifically, we constructed a binary variable coded as 0 if no one else was present during the interview and 1 if one or more individuals were present (N = 7,954 with others present; N = 4,054 with no one present). The results (Figure J19) show that findings hold regardless of parent/guardian presence. While treatment effects are slightly smaller when no others are present, they are not significantly different at the $p < 0.05$ level. Interestingly, we find that guardians are marginally less likely to be present during interviews with students assigned to the treatment group ($beta = -0.03, p = 0.04$). If guardian presence were encouraging socially desirable or "correct" answers, this pattern would bias our results against finding significant effects – further increasing our confidence in the findings.

Figure J19: Subgroup ITT by parent/guardian presence



K Attrition and Bounds

K.1 Endline 1

Table K33 highlights that girls, Hindu students, and students with higher levels of parental education attrited at lower rates (i.e., they were more likely to complete the first endline survey). With Hindu respondents, we underscore that the effect is small (about 2 percentage points) and, importantly, there is no differential attrition between treatment and control. On the other hand, older students and those in higher class grades attrited at marginally higher rates. Table K34 highlights that attrition was not different between treatment and control groups. From an inference point of view, since our main specification estimates the ITT and not the ATE, (lack of) differential effects on attrition are more crucial relative to the few differential effects on compliance that we detect.

Table K33: Attrition predictors

Predictor	**N**	**Estimate*100**	**SE**	**p-value**	**p-value (FDR)**
Treatment - Media Literacy	13,591	0.37	0.005	0.42	0.69
Gender - Girl	13,591	-5.73***	0.006	<0.001	<0.001
Grade	13,590	1.72***	0.002	<0.001	<0.001
Age	13,591	1.74***	0.002	<0.001	<0.001
Religion - Hindu	13,591	-3.36**	0.012	0.005	0.02
Language - Hindi	13,591	0.14	0.007	0.85	0.9
Asset Index	13,591	0.56	0.003	0.06	0.16
Father's Education	12,891	-0.19**	0.001	0.006	0.02
Mother's Education	12,950	-0.15*	0.001	0.04	0.11
Government School	13,591	-1.52	0.016	0.34	0.62
Science Knowledge	13,591	-0.31	0.002	0.17	0.37
Mobile Internet	13,591	-0.33	0.007	0.66	0.9
Newspapers	13,591	-0.19	0.010	0.84	0.9
Social Media	13,591	0.06	0.006	>0.9	>0.9
TV	13,591	0.27	0.008	0.72	0.9
Friends and Family	13,591	-0.44	0.017	0.8	0.9
Vaccinated	13,591	0.46	0.007	0.52	0.78
Ayurveda	13,591	-0.82	0.009	0.34	0.62

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

Table K34: Attrition predictors (*treatment)

Predictor	**N**	**Estimate*100**	**SE**	**p-value**	**p-value (FDR)**
Gender - Girl * T	13,591	-1.90	0.012	0.13	0.73
Grade * T	13,590	-0.09	0.005	0.86	>0.9
Age * T	13,591	0.32	0.004	0.43	0.79
Religion - Hindu * T	13,591	1.18	0.023	0.61	0.86
Language - Hindi * T	13,591	0.02	0.011	>0.9	>0.9
Asset Index * T	13,591	0.64	0.006	0.27	0.79
Father's Education * T	12,891	0.11	0.001	0.42	0.79
Mother's Education * T	12,950	0.24	0.001	0.07	0.73
Government School * T	13,591	-1.83	0.032	0.56	0.86
Science Knowledge * T	13,591	-0.70	0.004	0.1	0.73
Mobile Internet * T	13,591	1.16	0.014	0.42	0.79
Newspapers * T	13,591	-0.64	0.019	0.73	>0.9
Social Media * T	13,591	-1.07	0.012	0.36	0.79
TV * T	13,591	-0.33	0.015	0.82	>0.9
Friends and Family * T	13,591	-2.45	0.034	0.47	0.79
Vaccinated * T	13,591	-1.23	0.014	0.38	0.79
Ayurveda * T	13,591	0.47	0.017	0.78	>0.9

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

K.2 Endline 2

The final follow-up sample was N=2059 and we attempted to call N=2874 more households we were unable to reach. We see very limited differences between the follow-up sample that we end up with and the households that were called but that we did not reach (see Table K35). There is some evidence that students with better effects on sharing discernment and engagement attitudes were more likely to be reached (see Table K37). While these differences are statistically significant at just below p=0.05, they are substantively very minor.

There is also some evidence suggesting that girls were less likely to be reached on being called (Table K36), which contrasts with findings from the first endline survey, where compliance was higher among girls and attrition rates were lower. Several factors might explain this discrepancy. First, the mode of data collection differed: the first endline was conducted in person, while the second relied on phone calls. Girls may have been less likely to answer a shared "household phone", or parents may have been less willing to allow them to speak with strangers over the phone. Additionally, seasonal variation could play a role; the first endline occurred during the school year, while the second took place during summer holidays, which may have influenced these differences.

Table K35: Second endline sample characteristics

Characteristic	**Overall** *N = 13591*	**Called, part.** *N = 2060*	**Called, not part.** *N = 2874*	**Not called** *N = 8657*
Treatment				
Spoken English	6,774 (50%)	1,028 (50%)	1,422 (49%)	4,324 (50%)
Media Literacy	6,817 (50%)	1,032 (50%)	1,452 (51%)	4,333 (50%)
Gender				
Boy	5,674 (42%)	802 (39%)	1,270 (44%)	3,602 (42%)
Girl	7,917 (58%)	1,258 (61%)	1,604 (56%)	5,055 (58%)
Grade	9.66 (1.29)	9.73 (1.31)	9.62 (1.25)	9.66 (1.29)
Age	14.90 (1.58)	14.95 (1.60)	14.90 (1.56)	14.90 (1.58)
Religion - Hindu	91%	92%	89%	91%
Language - Hindi	43%	44%	37%	44%
Asset Index	0.00 (1.00)	0.03 (0.99)	-0.01 (0.99)	0.00 (1.01)
Father's Education	6.9 (4.6)	7.0 (4.7)	6.8 (4.6)	7.0 (4.6)
Mother's Education	4.1 (4.7)	4.4 (4.7)	3.9 (4.6)	4.1 (4.7)
Government School	96%	96%	96%	96%
Has Mobile Internet	19%	18%	19%	19%
Trust Newspapers	90%	91%	92%	90%
Trust Social Media	61%	62%	62%	60%
Trust TV	84%	83%	86%	84%
Trust Friends	97%	97%	97%	97%
Vaccinated	77%	77%	78%	77%
Ayurveda Effective	87%	86%	88%	87%

[1] n (%); Mean (SD); %

Table K36: Second endline attrition predictors

Predictor	**N**	**Estimate**	**SE**	**p-value**
Treatment	4,934	-0.01	0.016	0.66
Gender - Girl	4,934	0.05***	0.015	<0.001
Grade	4,933	0.01**	0.005	0.009
Age	4,934	0.00	0.005	0.32
Religion - Hindu	4,934	0.08**	0.028	0.004
Language - Hindi	4,934	0.03	0.017	0.06
Asset Index	4,934	0.01	0.008	0.08
Father's Education	4,689	0.00	0.002	0.06
Mother's Education	4,704	0.00**	0.002	0.003
Government School	4,934	0.02	0.040	0.6
Science Knowledge	4,934	0.01*	0.005	0.04
Mobile Internet	4,934	-0.02	0.018	0.18
Newspapers	4,934	-0.03	0.025	0.2
Social Media	4,934	0.00	0.014	0.74
TV	4,934	-0.04*	0.019	0.02
Friends and Family	4,934	0.00	0.043	>0.9
Vaccinated	4,934	-0.02	0.018	0.28
Ayurveda	4,934	-0.02	0.022	0.35
Awareness	4,275	0.00	0.008	0.64
Accuracy Discernment	4,032	0.01	0.007	0.39
Sharing Discernment	4,081	-0.01	0.008	0.31
Health Preferences	4,088	0.01	0.008	0.4
Source Discernment	4,147	0.01	0.008	0.28
Engagement Attitude	4,310	0.00	0.008	>0.9
Engagement Behavior	4,358	0.00	0.008	>0.9

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

Table K37: Second endline attrition predictors (*treatment)

Predictor	**N**	**Estimate**	**SE**	**p-value**
Gender - Girl * T	4,934	0.03	0.029	0.33
Grade * T	4,933	-0.01	0.011	0.21
Age * T	4,934	0.00	0.009	0.72
Religion - Hindu * T	4,934	0.03	0.058	0.62
Language - Hindi * T	4,934	0.01	0.030	0.71
Asset Index * T	4,934	0.00	0.015	0.84
Father's Education * T	4,689	0.00	0.003	0.13
Mother's Education * T	4,704	0.00	0.003	0.77
Government School * T	4,934	0.08	0.080	0.33
Science Knowledge * T	4,934	0.01	0.010	0.35
Mobile Internet * T	4,934	0.01	0.035	0.83
Newspapers * T	4,934	-0.01	0.050	0.84
Social Media * T	4,934	-0.05	0.027	0.08
TV * T	4,934	0.04	0.038	0.32
Friends and Family * T	4,934	0.08	0.085	0.37
Vaccinated * T	4,934	0.04	0.034	0.29
Ayurveda * T	4,934	-0.04	0.043	0.33
Awareness * T	4,275	-0.01	0.016	0.62
Accuracy Discernment * T	4,032	0.02	0.015	0.19
Sharing Discernment * T	4,081	0.04*	0.015	0.01
Health Preferences * T	4,088	0.02	0.015	0.25
Source Discernment * T	4,147	-0.02	0.015	0.29
Engagement Attitude * T	4,310	0.03*	0.015	0.03
Engagement Behavior * T	4,358	0.00	0.015	0.84

*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

K.3 Sensitivity analysis

In this section, we evaluate the robustness of our results under the assumption that our outcome data is not missing at random. While Table K33 suggests that outcome data is *not* missing completely at random (MCAR), Table K34 suggests that baseline characteristics do not affect missingness differentially in treatment and control groups. However, we cannot be completely sure that missingness is not systematically correlated to respondents' potential outcomes beyond what is captured by covariate profiles.

Below, we conduct three analyses to suggest that our results are not driven by differential attrition.

K.3.1 Tipping point analysis

First, we conduct an adapted version of tipping point analyses, which are commonly used in clinical trials (Yan, Lee, and Li, 2009). Specifically, we use a computational approach to investigate how much worse the (unobserved) treated potential outcomes of treatment-group respondents for which data is missing would have to be compared to the (observed) treated potential outcomes of treated respondents for which data is not missing, for our estimates to be an artifact of differential attrition.

Concretely, we start by imputing missing outcome data for all respondents based on treatment assignment, class attendance, and a list of individual-level covariates[3] using the `mice` package in `R`. We then re-estimate our main unadjusted model with library-spillover fixed effects and standard errors clustered at the classroom level using the imputed data.

To test the sensitivity of our results to increasingly unfavorable assumptions about the missing data, we then progressively adjust the imputed outcomes for treatment group subjects, incrementally subtracting 0.01 control-group standard deviations (SDs). This iterative process simulates increasingly severe violations of the assumption that missingness is unrelated to potential outcomes once accounting for treatment assignment, attendance, and covariate profiles of respondents.

For each outcome, we identify and record three tipping points: (1) the point at which the treatment effect ceases to be positive and statistically significant at $p < 0.05$, (2) the point at which the point estimate becomes negative, and (3) the point at which the treatment effect becomes negative and significant at $p < 0.05$.

Table K38 shows that the treated potential outcomes of individuals assigned to treatment for which we do not have outcome data would have to be significantly worse than the treated potential outcomes of otherwise similar treatment-group subjects. For the accuracy discernment measure, for instance, the tipping point at which the estimate becomes statistically indistinguishable from 0 at $p < 0.05$ is 1.38 control-group SDs, implying that the treated potential outcome for missing individuals in the treatment group would have to be 1.38 SDs

[3] List of covariates used for imputation: gender, grade, age, religion, language, HH asset index, father's education, mother's education, type of school, science knowledge, access to mobile internet, trust in: newspapers, social media, TV, friends and family, Covid-19 vaccination status, and trust in ayurveda.

worse than their imputed outcomes. This difference would have to increase to 1.61 SDs/1.85 SDs for the point estimate to become negative/negative and significant, respectively.

Table K38: Tipping point analysis results

	Tipping points (in SD units)		
Outcome	Positive, Not Significant	Estimate below 0	Negative, Significant
Accuracy Discernment	1.38	1.61	1.85
Sharing Discernment	0.99	1.22	1.46
Source Discernment	1.07	1.31	1.55
Health Preferences	1.12	1.37	1.62
Engagement Attitude	0.54	0.86	1.2
Engagement Behavior	–	0.17	0.48
Awareness	–	–	0.16

Note: This table summarizes the tipping points at which the treatment effect loses statistical significance, reaches zero, and becomes significantly negative. The values represent how many (control-group) SD units below the imputed values the treatment group outcomes would need to be in each case. Models include library-spillover FEs.

Overall, this analysis shows that where we do find statistically significant effects on our outcome indices without accounting for missing data, missingness would have to highly correlated to potential outcomes for our findings to no longer be robust. Specifically, the potential outcomes of missing individuals in the treatment group would need to diverge drastically from those of observed treatment-group individuals, far exceeding what is plausible given our data and the covariates used for imputation. This highlights the robustness of our results to even substantial deviations from the assumption of random missingness, reinforcing confidence in the validity of our estimated treatment effects.

K.3.2 Inverse probability weighting

Second, we implement an inverse probability weighting (IPW) approach. Specifically, we estimate the probability that each respondent completed the endline survey using a logistic regression model that includes treatment assignment and its interaction with a rich set of baseline covariates. These covariates capture key demographic, socioeconomic, educational, attitudinal, and behavioral characteristics that might jointly influence both treatment assignment and survey completion.

Missing values in the covariates are imputed using median imputation for numeric variables and the modal category for categorical variables. Predicted probabilities from the attrition model are then used to compute inverse probability weights, where each respondent is weighted by the inverse of their estimated probability of being observed. This gives greater influence to individuals who are underrepresented due to differential attrition and reweights the sample to approximate the full randomized sample.

We use these weights in survey-weighted regressions for each of our main outcome variables in the main endline survey (Table K39) and the follow-up survey (Table K40). The results are robust to this specification. While this approach relies on the assumption that attrition is ignorable conditional on the included covariates, it substantially mitigates concerns that observed treatment effects are driven by differential dropout across groups.

Table K39: ITT results with IPW

Outcome	**N**	**Estimate**	**SE**	**p-value**
Awareness	11,781	-0.01	0.024	0.64
Misinfo Threat Perception	11,947	0.10	0.022	<0.001
News Manipulation Awareness	11,909	-0.14	0.022	<0.001
Confirmation Bias Awareness	11,905	-0.02	0.022	0.33
Accuracy Discernment	11,121	0.32	0.024	<0.001
Accuracy of True Statements	11,932	-0.07	0.021	<0.001
Inaccuracy of False Statements	11,158	0.50	0.026	<0.001
Sharing Discernment	11,263	0.21	0.023	<0.001
Share True Statements	11,894	0.00	0.022	>0.9
Not Share False Statements	11,319	0.24	0.026	<0.001
Health Preferences	11,309	0.21	0.021	<0.001
Interest in Health News	12,007	0.03	0.019	0.12
Perception of Vaccine Safety	11,471	0.07	0.022	0.001
Distrust Non-Scientific Remedies	11,798	0.23	0.024	<0.001
Source Discernment	11,452	0.21	0.020	<0.001
Trust Reliable Sources (generic)	11,922	0.02	0.020	0.3
Distrust Unreliable Sources (generic)	11,893	0.10	0.023	<0.001
Trust Reliable Sources (specific)	11,892	-0.04	0.023	0.07
Distrust Unreliable Sources (specific)	11,899	0.27	0.023	<0.001
Engagement Attitude	11,868	0.11	0.023	<0.001
Correct Friend	11,947	0.01	0.022	0.74
Not Share Friend's Misinformation	11,917	0.11	0.021	<0.001
Fact Checking Important	12,007	-0.02	0.020	0.26
Fact Checked Recently	12,007	0.10	0.024	<0.001
Engagement Behavior	12,007	0.02	0.020	0.34
Prefer Hindustan Newspaper	12,007	0.02	0.020	0.31
Willing to Be a Truth Ambassador	12,007	0.01	0.019	0.68
Cues	11,656	0.09	0.017	<0.001

ITT effects using inverse probability weights to account for attrition.

DV: Main Indices. Models include library-spillover FEs.

Table K40: Second endline ITT results with IPW

Outcome	**N**	**Estimate**	**SE**	**p-value**
Accuracy Discernment Index	1,442	0.26	0.057	<0.001
Political Discernment Index	1,383	0.32	0.055	<0.001
Source Discernment Index	1,493	0.14	0.053	0.008
Accuracy of True Statements	1,508	-0.08	0.046	0.07
Inaccuracy of False Statements	1,455	0.35	0.047	<0.001
Accuracy of True Statements	1,469	0.02	0.048	0.65
Inaccuracy of False Statements	1,396	0.31	0.050	<0.001
Trust Reliable Sources	1,501	-0.08	0.049	0.12
Distrust Unreliable Sources	1,509	0.19	0.044	<0.001

ITT effects using inverse probability weights to account for attrition.

DV: Main Indices. Models include library-spillover FEs.

K.3.3 Lee bounds

Lastly, we examine the sensitivity of our results to potential bias from differential attrition using Lee bounds (Lee, 2009). This approach provides a nonparametric way to estimate worst-case and best-case treatment effects under minimal assumptions about attrition.

The key identifying assumption underlying Lee bounds is monotonicity: that treatment can only weakly increase the probability of being observed (i.e., completing the endline or follow-up). Under this assumption, Lee proposes trimming the outcome distribution of the higher-response group to match the response rate of the lower-response group. The resulting estimates provide sharp bounds on the average treatment effect, reflecting two polar scenarios: the lower bound assumes that all attrited treated individuals would have had the best possible outcomes, while the upper bound assumes they would have had the worst.

We adapt this method using a rank-based trimming approach suited to continuous and index outcomes. For each outcome, we sort treated observations by their outcome value and trim either the upper tail (for the lower bound) or the lower tail (for the upper bound) such that the effective sample reflects a response rate equal to that of the control group. Following McKenzie (2017), we then re-estimate our main specification—a model with library spillover fixed effects and village-clustered standard errors—on these trimmed samples.

We report Lee bounds for both the endline sample in Table K41 and the follow-up sample in Table K42. These results serve two purposes. First, it shows that our estimates remain robust under different subsamples, despite minor differences in attrition patterns across survey rounds. Second, it highlights that even when accounting for attrition through conservative bounding methods, the qualitative conclusions remain unchanged.

The results show that the Lee bounds are relatively tight around the observed treatment effects, suggesting that moderate selection on unobservables would not reverse our conclusions. For instance, in the endline sample, the treatment effect on accuracy discernment is 0.316, with bounds ranging from 0.312 to 0.322. A similar pattern holds in the follow-up sample, where the bounds are again centered near the point estimate.

Taken together, the Lee bounds results reinforce the conclusion from the tipping point analysis: our main findings are not highly sensitive to potential biases introduced by differential attrition. While Lee bounds make fewer assumptions than parametric models, their relatively narrow range in this context supports the credibility of our treatment effects.

Table K41: Lee bounds analysis results

Outcome	Diff.	Lower bound	Upper bound
Accuracy Discernment	0.315	0.312	0.322
Awareness	-0.011	-0.013	-0.006
Engagement Attitude	0.105	0.102	0.111
Engagement Behavior	0.018	0.017	0.025
Health Preferences	0.207	0.203	0.214
Sharing Discernment	0.207	0.201	0.214
Source Discernment	0.21	0.204	0.216

Note:
This table summarizes the difference in means for observed outcomes and Lee bounds estimated using a rank-based trimming procedure. The lower bound assumes that attrited treated units would have had the best outcomes and trims the top of the treated outcome distribution. The upper bound assumes they would have had the worst outcomes and trims the bottom. The trimming proportion equals the difference in response rates between the treatment and control groups.

Table K42: Lee bounds analysis results (second endline)

Outcome	Diff.	Lower bound	Upper bound
Accuracy Discernment	0.259	0.256	0.259
Accuracy of True Statements	-0.071	-0.071	-0.066
Inaccuracy of False Statements	0.335	0.333	0.339
Political Discernment	0.269	0.266	0.273
Accuracy of True Statements	-0.008	-0.008	-0.006
Inaccuracy of False Statements	0.281	0.28	0.284
Source Discernment	0.097	0.095	0.102
Trust Reliable Sources	-0.06	-0.06	-0.056
Distrust Unreliable Sources	0.135	0.133	0.138

Note: This table summarizes the difference in means for observed outcomes and Lee bounds estimated using a rank-based trimming procedure. The lower bound assumes that attrited treated units would have had the best outcomes and trims the top of the treated outcome distribution. The upper bound assumes they would have had the worst outcomes and trims the bottom. The trimming proportion equals the difference in response rates between the treatment and control groups.

L Deviations from PAP

We report three minor deviations from the pre-analysis plan (PAP). For reference our PAP was posted to OSF in February 2024 and is available here.

First, as described in the randomization procedure in Appendix A, we randomized to treatment and control within library-spillover strata. Consequently, our main baseline models in the paper include fixed effects (FEs) for library-spillover strata. However, in our PAP, we mistakenly stated that the main models would include district-level FEs alone. In practice, districts were not part of the randomization procedure. Instead, we used the library-spillover strata to ensure balance on key characteristics defining these strata (e.g., spillover potential, library attributes, proximity to other villages, and development indicators). Therefore, the appropriate specification includes FEs by library-spillover strata.

In the PAP, we had also indicated that we would estimate models separately with stratum FEs and library FEs. Instead, our main specification combines stratum and library FEs into a single set of library-spillover strata FEs. Nevertheless, we report a number of alternate specifications in Appendix J, including the original PAP model with district FE.

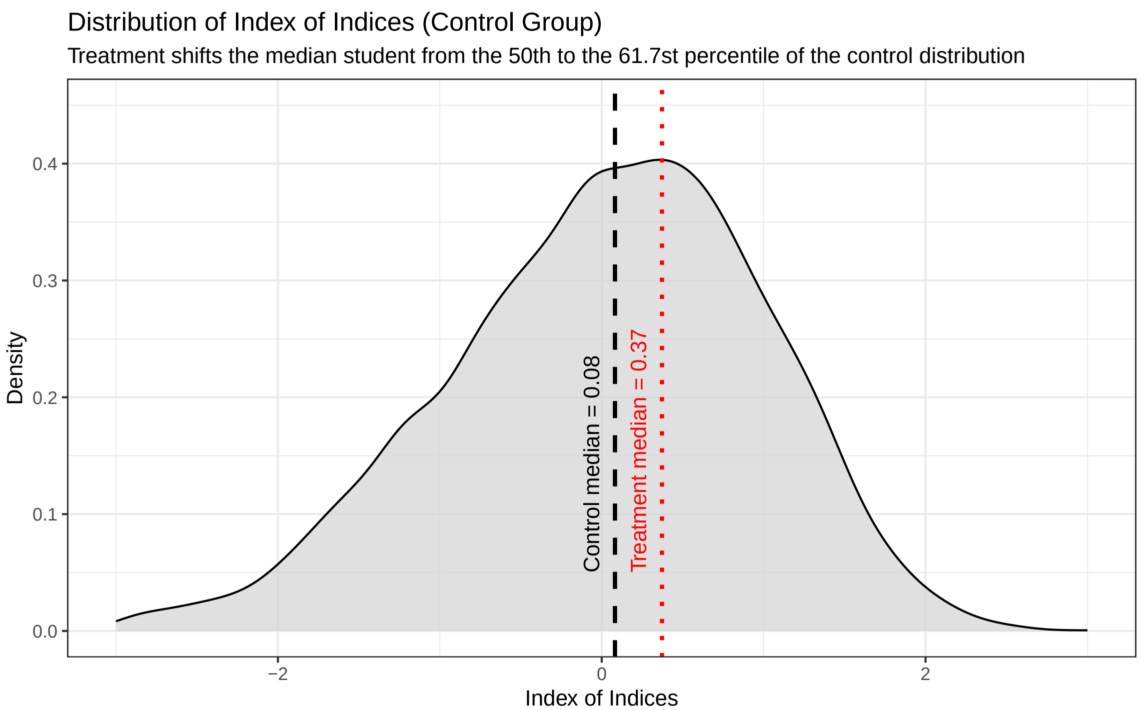
Second, for heterogeneity analyses, we pre-registered two treatment effect moderators: access and use of media, and attachment to non-scientific belief systems. We present results for these as specified. Additionally, we include a few more variables in our heterogeneity treatment effect (HTE) tables, which were framed as research questions (RQs) in the pre-analysis plan rather than formal hypotheses. The additional variables include demographics such as age, grade, gender, and household income. These were listed as research questions rather than formal hypotheses because we lacked strong priors about their potential effects, given the absence of similar studies in our context. However, we aimed to explore heterogeneity for these variables to inform policy. For instance, understanding whether older or younger students, or male or female students, show different learning outcomes is valuable for assessing generalizability, even without specific expectations. We also included caste and religion at the suggestion of a reviewer. Finally, we analyze partisan identity, which was also framed as an RQ in the PAP due to initial uncertainty about our ability to estimate it robustly; however, our data allowed for reliable estimation.

M Cost effectiveness

To illustrate the cost effectiveness of the program, we conducted the following analysis.

First, in line with a reviewer request, we created an 'index of indices' of the individual outcome indices laid out in section D.1. We then computed the control-group median and distribution to express the median treatment-group outcome relative to the control-group distribution. As Figure M20 shows, the treatment shifts the median outcome of the index from the 50th to the 61st percentile of the control-group distribution.

Figure M20: Shift in median outcome relative to control-group distribution



At what cost did we achieve such effects? Or put otherwise, at what cost might governments achieve such effects in the future? To answer these questions, we present in Figure M21 below estimates of the cost of the intervention under different assumptions.

We begin by detailing the various costs incurred during the implementation of the intervention. These included teacher salaries, printing and dissemination of teaching materials, student kits and snacks, curriculum design, teacher training, and post-training support. Teacher salaries covered the period during which a specially hired cohort delivered the intervention. Printing and material costs encompassed over 15,000 high-quality, color copies of student materials, including flyers, homework assignments, lesson plans, classroom posters, and flipcharts. Each student also received a kit containing writing supplies and a blank notebook, along with snacks provided during each of the four sessions. Though the research team provide substantive, theoretical input, the practical aspects of curriculum design were outsourced to an external consulting firm, DataLeads, which developed the BIMLI logo, formatted the lesson plans into textbook-style materials, designed homework sheets and class content, and managed translation. For teacher training, a large pool of selected teachers was brought to Patna, the capital of Bihar, for a two-day training attended by the coauthors. These sessions included walk-throughs of each module as well as guidance on logistics, ethics, and child safety. The associated costs included venue rental, teacher accommodation for three days, and meals. Post-training support involved a team of research assistants and field managers who provided ongoing logistical help and troubleshooting throughout the program.

Table M21 breaks down the total program cost under four implementation scenarios. The first two columns reflect the full costs, including salaries for hired teachers. As noted in the main text, we opted to hire a dedicated pool of teachers through an open call, given that government schoolteachers were already overburdened and scheduling additional class sessions with them was infeasible. Columns 1 and 3 reflect total expenditures, while columns 2 and 4 exclude one-time fixed costs – specifically, curriculum development and teacher recruitment – which would not need to be repeated in future implementations. The per-student cost in the full-cost model, with hired teachers and fixed costs included, is approximately \$4.84. In contrast, columns 3 and 4 simulate a more scalable and realistic model in which the intervention is delivered by existing public-school teachers. Under this model, the per-student cost drops significantly, to below \$4 and even as low as \$1, depending on these assumptions. We believe this latter scenario better reflects how such an intervention would likely be implemented at scale, and therefore provides a useful benchmark for cost-effectiveness.

These estimates in our view show that relatively modest investments can shift the median student's outcomes meaningfully, suggesting that the program is highly cost-effective relative to its impact.

Figure M21: Cost Estimates of BIMLI Under Different Assumptions

	Modality 1	Modality 2	Modality 3	Modality 4
	With specially hired teachers		***Without specially hired teachers***	
	Short term cost (total for 300 treatment classrooms, in INR)	Long term cost (total for 300 treatment classrooms, in INR)	Short term cost (total for 300 treatment classrooms, in INR)	Long term cost (total for 300 treatment classrooms, in INR)
Fees to hired teachers	600000	600000	N/A	N/A
Printing and dissemination of teaching materials	80000	80000	80000	80000
Student kits & snacks	30000	30000	30000	30000
Curriculum Design	1170000	N/A	1170000	N/A
Training of teachers	600000	N/A	600000	N/A
Post-training support	500000	500000	500000	500000
TOTAL COST (in INR)	2980000	1210000	2380000	610000
TOTAL COST *per student* (in INR)	413.89	168.06	330.56	84.72
TOTAL COST *per student* (in USD)	**4.84**	**1.97**	**3.87**	**0.99**

References

Agarwal, Ananay, Neelesh Agrawal, Saloni Bhogale, Sudheendra Hangal, Francesca Refsum Jensenius, Mohit Kumar, Chinmay Narayan, Priyamvada Trivedi, and Gilles Verniers. 2021. "TCPD Indian Elections Data v2.0." *Trivedi Centre for Political Data, Ashoka University.*

Ali, Ayesha, and Ihsan Ayyub Qazi. 2023. "Countering misinformation on social media through educational interventions: Evidence from a randomized experiment in Pakistan." *Journal of Development Economics* 163: 103108.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 2021. "Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform." *The World Bank Economic Review* 35 (4): 845–871.

Badrinathan, Sumitra. 2021. "Educative Interventions to Combat Misinformation: Evidence From a Field Experiment in India." *American Political Science Review* 115 (4): 1325–1341.

Biswas, Firoj. 2023. "Electoral patterns and voting behavior of Bihar in Assembly elections from 2010 to 2020: a spatial analysis." *GeoJournal* 88 (1): 655–689.

Blair, Harry W. 1972. "Ethnicity and democratic politics in India: Caste as a differential mobilizer in Bihar." *Comparative Politics* 5 (1): 107–127.

Capewell, Georgia, Rakoen Maertens, Miriam Remshard, Sander Van Der Linden, Josh Compton, Stephan Lewandowsky, and Jon Roozenbeek. 2024. "Misinformation interventions decay rapidly without an immediate posttest." *Journal of Applied Social Psychology* 54 (8): 441–454.

Chauchard, Simon, and Sumitra Badrinathan. 2025. "The Religious Roots of Belief in Misinformation: Experimental Evidence from India." *Forthcoming, British Journal of Political Science.*

Jaffrelot, Christophe. 2013. "Caste and political parties in India: do Indians vote their caste—while casting their vote?" In *Routledge Handbook of Indian politics.* Routledge.

Jaffrelot, Christophe. 2023. "Do Indians Vote Their Caste-or Their Jati, or Their Class, or...?" *The Oxford Handbook of Caste*: 331.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3).

McKenzie, David. 2017. "Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition." *American Economic Review* 107 (8): 2278–2307.

Yan, Xu, Shiowjen Lee, and Ning Li. 2009. "Missing data handling methods in medical device clinical trials." *Journal of Biopharmaceutical Statistics* 19 (6): 1085–1098.