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

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

Word Count: 12,915

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 spillover effects within households, 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

---

*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 UC3M's 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, Erik Kramon, Horacio Larreguy, Gareth Nellis, Laura Paler, Casey Petroff, Cyrus Samii, Laura Schechter, and seminar participants at George Washington, Penn, American, Rochester, Brown, Michigan, Sciences Po, UMontreal, Humboldt, University of Oslo, IC3JM, CUNEF, Namur, and APSA and ICA conferences.

[†]PhD Candidate, University of Wisconsin–Madison

[‡]Assistant Professor, American University

[§]Associate Professor, University Carlos III Madrid & Instituto Carlos 3 - Juan March

[¶]PhD Candidate, Princeton University

1 Introduction

Around the world, educational programs have long been seen as potential catalysts for societal transformation. Political leaders 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 China, Cantoni et al. (2017) found that school curriculum reforms fostered positive attitudes towards the nation. 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) found that an additional year of schooling reduced anti-immigration sentiments later in life. And in Mali, Gottlieb (2016)'s work on civic education demonstrated that such learning 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?

We address this question in the context of India, where misinformation has had severe consequences. While misinformation is a global issue (Persily, Tucker, and Tucker, 2020), it presents unique challenges in developing economies, settings often characterized by lower digital literacy, expanding internet access, informal information networks, and identity-driven misinformation (Badrinathan, 2021; Chauchard and Garimella, 2022; Gottlieb, Adida, and Moussa, 2022). Despite recent progress in misinformation research, most studies focus on Western contexts, leaving a gap in understanding effective solutions for the Global South (Blair et al., 2023). India exemplifies the multifaceted nature of this problem, where misinformation can affect political violence and health behaviors (Bridgman et al., 2020). Low state capacity, particularly in healthcare, and shrinking independent media further limit access to credible information (Mohan, 2021;

Sen, 2023; Sharma, 2015). Political and religious elites spread misinformation to advance partisan agendas, making it salient in a highly polarized environment (Siddiqui, 2020). This combination of elite-driven disinformation, restricted media access, and weak institutions enables misinformation to flourish with impunity.

A substantial body of academic research has tested the efficacy of misinformation countermeasures, including fact-checking and corrections (Porter and Wood, 2019; Bowles et al., 2024; Clayton et al., 2019), accuracy reminders (Pennycook and Rand, 2019), preemptive debunking (Pereira et al., 2024; Roozenbeek et al., 2022), and providing tips to respondents (Guess et al., 2020). While many of these methods are effective, they are largely one-off, online strategies targeting digitally literate, urban populations and are seldom adapted for offline or vulnerable communities (Blair et al., 2023; Brashier, 2024). Meanwhile, governments and NGOs have increasingly promoted classroom-based media and information literacy programs targeting youth, with a global surge in such initiatives after 2016. For instance, New Jersey is advancing mandatory K-12 media literacy education (Sitrin, 2020), mirroring efforts in California, Estonia, and Finland. Yet, remarkably, there is a lack of causal evidence evaluating these interventions despite their resource-intensive nature and cost: to date, no study has estimated the causal effects of sustained classroom-based education in combating misinformation.[1]

To fill this gap, we conducted a field experiment in Bihar, one of India's least developed states, involving over 13,500 adolescents aged 13–18 across 583 villages. We focus on adolescents because this developmental stage is crucial for moral and identity formation; young people are still open to shaping their attitudes, while adults with entrenched partisan or social identities often resist change when confronted with misinformation corrections (Niemi and Jennings, 1991; Markus and Nurius, 1986; Flynn, Nyhan, and Reifler, 2017). 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 study, focused on health misinformation and aimed to (1) enhance scientific knowledge

[1] A partial exception is Apuke, Omar, and Asude Tunca (2023), where a six-week media literacy course in Nigeria found positive effects, though its limited sample size and potential validity issues (no discussion of spillover or compliance) remain significant limitations.

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.

Our intervention was, by design, a bundled treatment combining several elements into a comprehensive learning course. We recruited and trained educators to deliver lessons in government-owned libraries across the state. The Bihar Rural Livelihoods Promotion Society (BRLPS, or as it is commonly known, Jeevika), an autonomous body under the Bihar State Government, sanctioned the program to be administered as an official government certificate course, thereby extending the reach and legitimacy of the intervention and boosting enrollment.[2] We randomized respondents into treatment at the village level, with the study targeting one classroom (20 to 24 respondents) per village. Control group villages received placebo classes on basic conversational English, designed to hold constant attendance and compliance with a long-term program and only vary the content of instruction.

We evaluate the effect of this intervention on a variety of misinformation-related outcomes, including attitudes, preferences, and behaviors. Intent-to-treat estimates measured soon after the intervention indicate that the intervention 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 participants demonstrated heightened discernment in evaluating information and making decisions regarding the sharing of news items (0.32 SD), with effects sizes substantially larger than those previously identified. Notably, the intervention also brought about changes in their health preferences (0.21 SD), diminishing reliance on alternative medical approaches to cure serious illnesses. This is crucial as previous studies show that misinformation related to health, and especially misinformation tied to traditional beliefs systems, is more difficult to change (Chauchard and Badrinathan, 2024). Further, the treatment enhanced students' ability to assess the credibility of different types of sources, including mediums, platforms, and transmitters of news. Finally, while intent-to-treat estimates show no overall effect of the treatment on behaviors regarding misinformation coun-

[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.

termeasures, 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 women.

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 a large effect 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 are better able to discern true from false political news 4 months after an intervention that focused entirely on health misinformation (0.31 SD), demonstrating that they were able to learn from the treatment, retain its lessons, and apply it to entirely 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 network-level spillover effects, and trickle-up socialization from children to parents (Carlos, 2021; Dahlgaard, 2018). Several of the outcomes we measure assess and require the acquisition 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 application rather than simple recall-based responses.

Our study contributes to the expanding body of literature addressing strategies to mitigate misinformation on a global scale. But what distinguishes our research from existing work is its departure from the assumption that misinformation predominantly manifests a problem to be addressed online. Prior research has largely operated under this presumption, leading to the development of numerous interventions tailored to online demographics. These interventions include, for instance, online gaming as a strategy to combat misinformation (Roozenbeek and van der Linden, 2019) and the provision of corrective nudges through direct messaging on Twitter (Pennycook et al., 2021). However, such interventions are inherently ill-suited for individuals who do not have regular or sustained access to the internet: in our sample, fewer than 1 in 5 households report access to a mobile phone with internet. Moreover, in contexts like India, those who do have access often use encrypted messaging platforms like WhatsApp, where interventions reliant on platform-based changes are not feasible (Badrinathan, 2021). Indeed, prior research in developing contexts has underscored the significance of offline sources and communal spaces for news dissemination and consumption, such as interpersonal communication, the role of the family, and face-to-face interactions in community settings like marketplaces (Gadjanova, Lynch, and Saibu, 2022). Operating under this premise, one of our primary contributions is studying adolescents. In doing so, we target a demographic with potentially less entrenched attitudes. Insofar as studies on misinformation consistently indicate that a key impediment to the effectiveness of interventions is motivated reasoning – the inclination to accept or reject information based on pre-existing beliefs and identities (Taber and Lodge, 2006) – we thus potentially engage with a population that has not yet developed resistance to altering entrenched viewpoints (Niemi and Jennings, 1991). Finally, partnering directly with a government agency to implement this experiment as an official program boosted both the legitimacy and scalability of our intervention, reducing non-compliance, and simulating a real-world rollout of a school-based program.

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 (Ghosh et al., 2024; Cheema et al., 2023; Banerjee et al., 2014).

2 Sustained Education Against Misinformation

The global rise of misinformation has attracted substantial academic and policy attention (Persily, Tucker, and Tucker, 2020), sparking a surge of experimental studies identifying solutions to the problem. However, many misinformation interventions have limitations. Their effects often decay rapidly (Capewell et al., 2024), and they primarily target digital or online populations, assuming misinformation is an online phenomenon. This overlooks offline communities, especially in rural or underconnected areas, where in-person interactions dominate (Gadjanova, Lynch, and Saibu, 2022). This focus also overlooks online populations where exposure to information happens via WhatsApp or other Mobile Instant Messengers (MIMs) – as is the case in many developing countries – as encrypted services do not lend themselves to corrections, fact-checking, or algorithmic interventions (Badrinathan, 2021; Rossini et al., 2020; Valeriani and Vaccari, 2018).[3] Moreover, many popular interventions rely on short-term nudges or priming, which fail to address entrenched norms or build lasting skills (Pennycook and Rand, 2019). Relatively more intense approaches, like inoculation strategies (Roozenbeek et al., 2022), have shown some promise in the Global North but have yielded mixed results in the Global South and other non-Western contexts (Blair et al., 2023; Badrinathan and Chauchard, 2023a).

These limitations imply that for contexts like ours, interventions may need to be conducted offline rather than online, targeting populations with limited internet access, and avoiding assumptions that citizens have skills and motivation to counter misinformation. Taking these reflections into account, we developed a classroom-based field experiment relying on an educational program entitled the *Bihar Information and Media Literacy Initiative* (BIMLI).

This intervention was aimed at achieving two primary objectives: (a) enhancing knowledge through skills and factual learning, and (b) shifting norms surrounding misinformation. By *knowledge*, we refer to two components: (1) *recall* – the ability to remember specific information learned in the classroom, and (2) *application* – the ability to use general skills and tools acquired in class to critically assess new information beyond the classroom. In the first case, the intervention may enhance knowledge by exposing students to specific, accurate facts they can later recall. For example, since the curriculum focuses on health-related misinformation, students might remember that there is no evidence supporting the claim that papaya leaves cure dengue fever. In the second case, students may develop more general skills, such as effective strategies for detecting

[3]While scholars have looked at social corrections or peer corrections to misinformation on encrypted apps (Badrinathan and Chauchard, 2023b; Bode and Vraga, 2015), such techniques assume that people already have the knowledge, desire and skills to fact check information, such that when they receive primes about accuracy or corrections, they are able to change beliefs and behavior. These assumptions may not hold for populations new to the internet, with limited digital literacy, or in rural, low-income settings.

misinformation in the future or consuming information critically by using cues like emotionality or source, or simply learning to pause before sharing information impulsively. These broader skills could significantly influence how they evaluate new information encountered even after the program, whether or not it is health-related.

With respect to *norms*, the intervention aimed to influence perceptions of appropriate behaviors and attitudes toward misinformation. Several elements in the curriculum explicitly or implicitly targeted normative change, including modules designed to (1) elevate the perceived importance of misinformation as a societal issue, (2) increase awareness of the risks associated with misinformation, (3) shift what individuals consider acceptable to say, believe, and share within their social networks, and (4) suggest ways and strategies to intervene when faced with acquaintances spreading misinformation. Given that educational institutions often serve as strong sources of normative influence (Tankard and Paluck, 2016), the program's affiliation with a government agency and the involvement of local authority figures likely amplified these signals. Additionally, teachers – often local role models – may themselves have been seen as influential social referents for students (Paluck and Shepherd, 2012).

Overall, the BIMLI program was thus designed to achieve multiple objectives. It sought not only to enhance knowledge – both specific and general – about misinformation but also to shift social norms around how misinformation is perceived and responded to. By targeting both cognitive and normative dimensions, our goal was to foster long-lasting changes in behaviors and attitudes, addressing the deeper, more ingrained habits that contribute to the persistence of misinformed beliefs. Our curriculum (described in detail below) emphasized interactive teaching, analytical skills, and critical thinking – features that are notably absent from the typical schooling approach in India, which often relies on one-way communication from teacher to student, rote learning, and memory-based exams (Kumar, 1986; Bhattacharya, 2022).

BIMLI was, by design, a significantly more intensive and substantively different intervention against misinformation than those typically tested. We introduced several key design and theoretical innovations to aid the successful implementation of the program. First, in terms of mode of delivery, we administered the program face-to-face in classroom settings, fostering a peer-based, interactive environment with iterative learning, where respondents encountered key lessons repeatedly over multiple sessions. 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). Additionally, partnering with an agency of the Bihar state government to roll out the program as an official government-endorsed certificate course increased its legitimacy. This official recognition was aimed at incentivizing participation and higher compliance.

Our second innovation was with regards to the intensity of the treatment. The total dosage of our program amounted to approximately 10 contact hours between teachers and students. This is significantly more labor-intensive than existing studies: for example, Hameleers (2020) and Guess et al. (2020) provided respondents with tips to spot misinformation via flyers; Ali and Qazi (2023) used a brief 3-4 minute media literacy video. Closer in design to our study, Badrinathan (2021) engaged respondents in a one-hour discussion on media literacy, and Moore and Hancock (2022) offered a self-directed hour-long digital literacy module. These are typically short-term, one-off treatments without continued engagement, and many other interventions rely on primes or reminders about critical thinking rather than providing actual tools or techniques for long-term learning. Consequently, we opted for this sustained, iterative approach involving multiple sessions of learning as well as homework assignments between sessions, thereby departing from almost all other educative interventions in the misinformation literature (Blair et al., 2023).

Third, our target population was adolescents. Existing research shows that adults, especially in polarized contexts, are significantly harder to influence due to cognitive biases such as motivated reasoning (Badrinathan, 2021). Adolescents, by contrast, are still forming their worldviews and may be more open to new information and perspectives. This demographic, therefore, offers a unique opportunity to intervene before attitudes and behaviors related to misinformation become deeply entrenched (Niemi and Jennings, 1991).

3 The Politics of Misinformation in India

Health-related misinformation is widespread in India. 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, 2024), 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, such as hospital treatment, leading to potentially fatal outcomes (Bridgman et al., 2020). Such health misinformation may also lead to belief in falsehoods and conspiracies in other domains, highlighting its contagion potential (van Prooijen and Böhm, 2024). 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 tied to social identities, and are often manipulated by political elites to gain electoral support. Traditional health remedies, many of which are linked to ancient Hindu culture, are used as a tool to garner support from the Hindu voter base. This is especially evident with the rise to power of the Hindu nationalist Bharatiya Janata Party (BJP), the party leading the current federal government in the country, which has positioned itself as a champion of Hindu values (Jaffrelot, 2021). A recent example of politically-driven health disinformation involves a BJP politician hosting a public event where participants drank cow urine, believing it to be a cure for covid – an event that ended with several attendees being hospitalized (Siddiqui, 2020). Politicians from the BJP have even boasted about their ability to make anything go viral, true or false (Kumar, 2019). Previous research demonstrates that misinformation that resonates with long-standing identities can be particularly hard to dislodge (Nyhan, 2021; Chauchard and Badrinathan, 2024). Indian politics has long been dominated by a fundamental cleavage between Hindus and Muslims, and the prominence of religion as a social identity has been central (Brass, 2011). Misinformed beliefs, particularly when propagated by elites seeking to further divide already polarized populations, may thus be very difficult to change. Some research in this context has shown that Indian citizens do hold deeply entrenched misinformed beliefs, with partisan motivated reasoning (Taber and Lodge, 2006) posing a significant barrier to efforts by civil society and other actors to correct falsehoods (Badrinathan, 2021). While Bihar, the context of our study, is not directly under single-party BJP rule, it is governed by a BJP-led coalition. Additionally, its geographical location in northern India places it within the broader media landscape of neighboring states such as Uttar Pradesh — regions that have witnessed multiple instances of elite-driven, high-profile disinformation, at times resulting in violence and fatalities (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, home to 127 million people, and with over one-third of the population living 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). Children in Bihar, especially girls, are significantly less likely to attend school compared to those in other states (Muralidharan and Prakash, 2017). Access to the internet is also limited: as per our baseline data, only 11.5% of respondents reported owning a personal cellphone; while many more had access to an internet-enabled smartphone, only 19% of them reported using the internet on that device. With most interactions and sources of information being offline, children primarily rely on their families for information. However, adults may themselves be misinformed, and strong cultural norms around deference to elders and family make such misinformation harder for children to challenge (Malhotra and Pearce, 2022). Even in households with internet access in our sample, it was often through a shared mobile phone, highlighting the stark contrast with Western settings, where internet access is more individualized (Steenson and Donner, 2017). This limited connectivity is compounded by a deteriorating informational environment across India. Independent media and spaces for dissent are shrinking, as state capture of institutions, including news sources, intensifies (Mohan, 2021; Sen, 2023). This aligns with broader observations of democratic decline in the country in recent years, where the space for credible information has narrowed significantly alongside decreasing state capacity (Tudor, 2023).

While vulnerability to misinformation can be thought of as a country-wide problem, Bihar thus does faces distinct structural challenges related to state capacity that set it apart from the rest of the country. This underscores the need for misinformation interventions that are tailored to the context. In Bihar, low state capacity, the instrumental use of misinformation by some political elites, entrenched cultural myths, and generational beliefs all contribute to the propagation of falsehoods, heightening the complexity of combating misinformation.

Despite these challenges, public trust in state institutions appears to be notably high in Bihar. During the study's design phase, for example, parents indicated their trust in Jeevika, our government partner organization, as a key reason for their support of the program. Our data also show that a substantial proportion of parents cited trust in Jeevika as their primary motivation for participating. This aligns with previous research highlighting Jeevika's successful initiatives in Bihar, including lending programs (Hoffmann et al., 2021), promoting women's financial autonomy (Datta, 2015), and supporting agricultural livelihoods (Baul et al., 2024). These findings suggest that citizen interaction with Jeevika and the government is both commonplace and effective in Bihar, potentially fostering trust. This further mirrors research indicating lower socio-economic households and residents of less developed states, like Bihar, may trust state governments more, likely due to reliance on government programs (Kumar, Pratap, and Aggarwal, 2021). Thus, while Bihar's structural challenges may amplify the impact of misinformation, citizen trust in its institutions provides an opportunity to leverage this reliance on both political and educational institutions to counter misinformation effectively. The next sections describe our field experiment that aimed to do so.

4 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 control respondents received conversational English classroom lessons (additional details below).

4.1 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 BIMLI 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. Media and information literacy is broadly defined as the skills and competencies that promote critical and responsible engagement with information and media sources (Jones-Jang, Mortensen, and Liu, 2021). In practice, however, this concept is operationalized in a myriad of ways, with scholars prioritizing different elements of the concept such as critical thinking, knowledge about media industries, or knowledge about psychological biases in thinking. In Table 1 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 treatment to the local context.

Substantively, our curriculum focused on fostering critical thinking rather than offering prescriptive tips to spot misinformation. Such tips appear to be the way in which many misinformation interventions operationalize "digital literacy" (Guess et al., 2020) or "media literacy" (Hameleers, 2020). Instead, we provided a theoretical framework for evaluating information broadly, regardless of its source. 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 newspaper over another) ineffective. Given the decline in mass media credibility amidst 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 for specific topics – for instance, relying on community health workers employed by the government (called ASHA workers) for health-related information.

We collaborated with DataLeads, a Delhi-based organization specializing in media literacy, alongside local Bihar educators and Indian media literacy experts, to design this curriculum. The modules consistently incorporated discussion-based activities to ensure engagement, and time-use lesson plans for teachers ensured that their delivery was highly standardized across classrooms. A central component of our teacher training emphasized creating interactive classroom sessions that encouraged engagement between teachers and students, as well as among students themselves. This approach aimed to cultivate analytical thinking, critical reasoning, and deep learning rather than relying solely on passive reception of information from an authority figure. This method represents 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

Table 1: 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; 2. Develop awareness of human biases in consumption information; 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 (2023).	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: *"Talking About Misinformation with Family and Vaccine Importance"*	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.

and Pearce, 2022). We developed extensive supporting materials, including detailed slide decks, teacher guidelines for working with children, time-structured lesson plans, and photo banks with contextualized examples. These resources, while designed for the local context and delivered in Hindi, could serve as a foundation for sustained educational efforts to combat misinformation in comparable contexts. Appendix B provides an overview of the materials used in the treatment.

Apart from the four in-person sessions, to complement in-class materials and encourage further reflection we also designed homework assignments that students completed between sessions. These included writing stories and recording observations, along with talking to family members about what was learned in the classroom. Finally, after each session, we provided succinctly designed take-home summary sheets for students that recap the goals and objectives of each lesson, designed to act as a guide or reference booklet.

4.2 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. Our engagement with Jeevika commenced in mid-2022, driven by a shared interest in combating health misinformation, particularly among youth.

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.[4] We ran our classes in these libraries from November 2023 to March 2024. We opted to deliver these classes after school hours and to not use existing public school teachers or school settings for this program. Our meetings with government officials revealed that public school teachers were already overburdened and that absenteeism (in students and teachers) was common in Bihar, which would have made compliance a big challenge. Besides, adding a module or topic in public school curriculum requires extensive bureaucratic authorizations and oversight. Recognizing that the success of such a program is con-

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

tingent not only on student attendance and compliance, but also crucially on teacher attendance and quality, we recruited a separate pool of teachers for this intervention, in partnership with DataLeads. Each teacher visited each classroom roughly once every two to three weeks.[5] Given our context, the curriculum was designed to be taught entirely offline, relying on face-to-face discussions, printouts of photos and examples when needed, and minimal digital technology. In many ways, therefore, the logistical aspects of our intervention were designed to mimic the prototypical learning environment of the average school-going child in rural India.[6]

To ensure its broad acceptance, Jeevika promoted the program as an officially recognized government-certified course, enhancing its credibility and external validity. This allowed us to reach remote rural populations often underrepresented in misinformation research. The libraries Jeevika provided access to were equipped with essential infrastructure – seating, blackboards, and other class equipment – which offered a level of standardization we would not have easily achieved across public schools. These libraries were relatively new constructions which allowed for conducive classroom settings that may have encouraged attendance, otherwise a major problem across the state's public schools.[7]

These efforts collectively aimed to incentivize enrollment and continued participation in the program. Notably, past studies using educational methods to combat misinformation, particularly in the Global South, have largely reported null results (Badrinathan, 2021; Ali and Qazi, 2023; Blair et al., 2023). This aligns with a broader literature in comparative politics suggesting that information provision alone often fails to change beliefs and behaviors. Citizens face constraints that hinder translating new information into action, requiring that treatments provide not only exposure to information but also incentives to prioritize the issue and promote confidence in the efficacy of citizen action (Kosec and Wantchekon, 2020). Group dynamics also play

[5]DataLeads received 400 applications and selected 50 teachers through an interview process, followed by a 2-day training to assess their skills. The selection process prioritized dynamic teaching styles to encourage classroom engagement. The final group included school teachers, journalists, professors, and fact-checkers, among other occupations. Each teacher was assigned 6-9 classrooms across 2-3 libraries, with each classroom having the same teacher throughout the intervention.

[6]See Appendix B1 for an example of what a classroom session in progress looks like.

[7]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; that number is the lowest in the 28 states in India that the survey spans.

a role; knowing others are informed can create social pressure to facilitate change (Lieberman, Posner, and Tsai, 2014). In our context, the educational classroom setting and iterative intervention, featuring authoritative figures and peer interactions, all aimed to enhance receptivity. Null effects in interventions may also stem from inadequate support from governing authorities and implementation disparities, especially in developing countries where elite involvement is critical for mobilizing rural citizens (Rao, Ananthpur, and Malik, 2017). Partnering with Jeevika helped address these challenges, along with using external educators and conducting the intervention in well-equipped library settings.

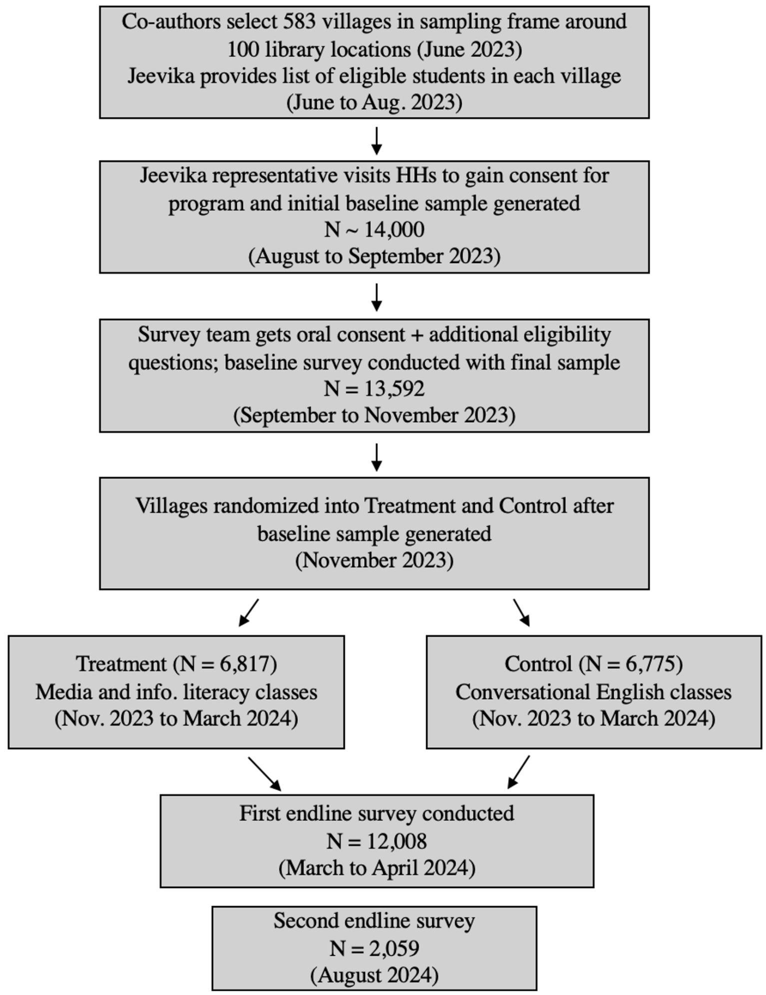
4.3 Sampling, enrolment and baseline data

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 (typically, 3) around each library to receive our treatment; the remaining served as control villages.[8] Our sampling procedures ultimately allowed us to select 583 villages which served as the sampling frame for this study. We determined pre-randomization whether each village had high- or low-spillover potential, and then assigned to treatment and control within each library and spillover strata (see Appendix A for a detailed note on sampling).

In each of the 583 selected villages, Jeevika provided a list of households based on data from enrollment in state programs. From this list, eligibility criteria for our study included households with children enrolled in school in grades 8–12. Jeevika staff visited these households to confirm the presence of an eligible child and explained the program. After these visits, a final list of 20 to 24 interested students was created for each village. Next, our survey team conducted a pre-randomization and in-person baseline survey from September to November 2023. Enumerators visited homes from the Jeevika list, applying three additional eligibility criteria: obtaining oral consent from the student and a parent/guardian, assessing basic (third-grade) Hindi reading comprehension, and ensuring the student could attend all four library sessions. Households failing any criterion were replaced. Figure 2 shows location of treatment and control villages across

[8] We selected fewer than 6 villages across 7 libraries, because there were not enough villages within the defined radius. Hence the final village sample is N = 583, relative to our original target N of 600.

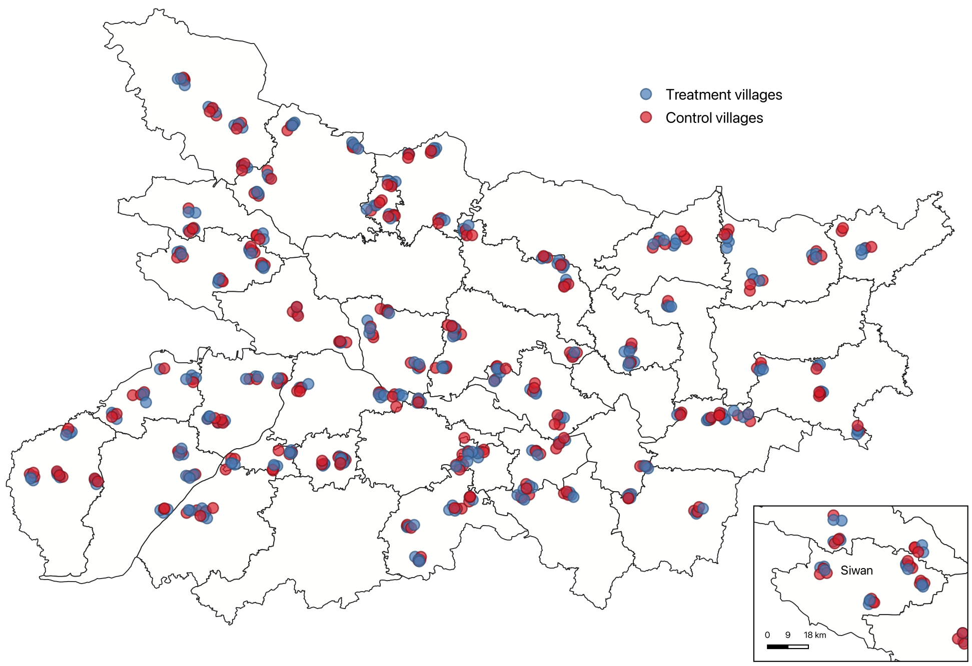
Figure 1: Study Flow and Timeline



Bihar. Crucially, we note that randomization occurred *after* students opted in, avoiding issues with differential opt-in rates between treatment and control.[9] 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. Appendix A details sampling procedures.

[9]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.

Figure 2: Map of treatment and control villages



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 in key demographic variables based on census parameters.

[10]The sample was 58% female, 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.

4.4 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. Encouraged by Jeevika, we additionally wanted control group respondents to benefit from the program and hence chose a topic that fostered engagement without confounding misinformation outcomes. Subjects like math, science, and history were excluded due to overlap with standard curricula or national identity narratives, and several other topics like cooking were discarded due to potential gender biases. We ultimately implemented a 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. 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. Appendix E summarizes teacher demographics by group.

4.5 Endline data and compliance

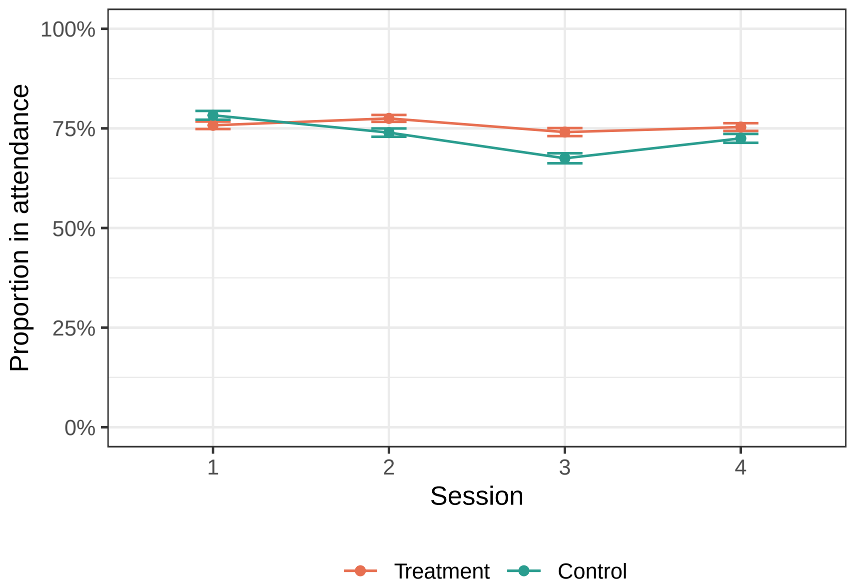
Our first endline survey was conducted in-person in the weeks following the end of the fourth and last session. Because of the large sample size, the endline took 5 weeks to complete. Although we did not randomize the order in which villages were surveyed at endline, we expect that endline survey timing, while dependent on a number of village-level covariates, is independent of treatment status. At the first endline, we were able to re-contact 12,008 of the total 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 female students and higher among older respondents (see Appendix C). 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 or parent 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.[11]

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. They also received notebooks, pens, flyers, homework, and snacks during each session. External monitors also made random visits to verify teacher presence and adherence to class schedule. Co-authors also visited during initial and final sessions. 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. This attendance allows us to measure whether students were fully non-compliant (attended 0 sessions), fully compliant (attended all 4 sessions) or partially compliant (1-3 sessions). Crucially, our definition of attrition does not hinge on compliance with the treatment. Even participants who are (partially) non-compliant with the treatment but continue to engage with the study by completing the endline survey are not classified as having attrited. Compliance data demonstrates that on average, 76% of the sample attended the first class. We detect no significant difference in attendance 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 3). Further, we find that girls were more likely to attend classes (both treatment

[11] 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.

Figure 3: Compliance data across treatment and control



and control) compared to boys, whereas older students were significantly less likely to attend compared to younger ones.[12] Importantly, there is no differential attendance for treatment and control as a function of these variables, nor do we detect any differential attrition (Appendix C).

4.6 Outcome measures

We pre-specified and included seven 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.[13] Since each outcome family includes multiple items, we compute inverse-covariance weighted (ICW) indices that combine and weight individual items, that are standardized relative to the control group mean and SD. Ultimately our main analysis focuses on seven outcome indices. Appendix D details the rationale for ICW indices and their pre-specified construction. The full endline instrument is posted to our pre-analysis plan. In

[12]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.

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

the second endline survey, we measured accuracy discernment for both respondents and one parent/guardian, alongside self-reported items seeking to elucidate mechanisms.

5 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:

$$Y_{ijk} = \beta_0 + \beta_1 T_{ijk} + \sum_{k=1}^{m-1} \gamma_k + \varepsilon_{ijk} \tag{5.1}$$

$$Y_{ijk} = \beta_0 + \beta_1 T_{ijk} + \sum_{c} \alpha_c X_{ci} + \sum_{k=1}^{m-1} \gamma_k + \varepsilon_{ijk} \tag{5.2}$$

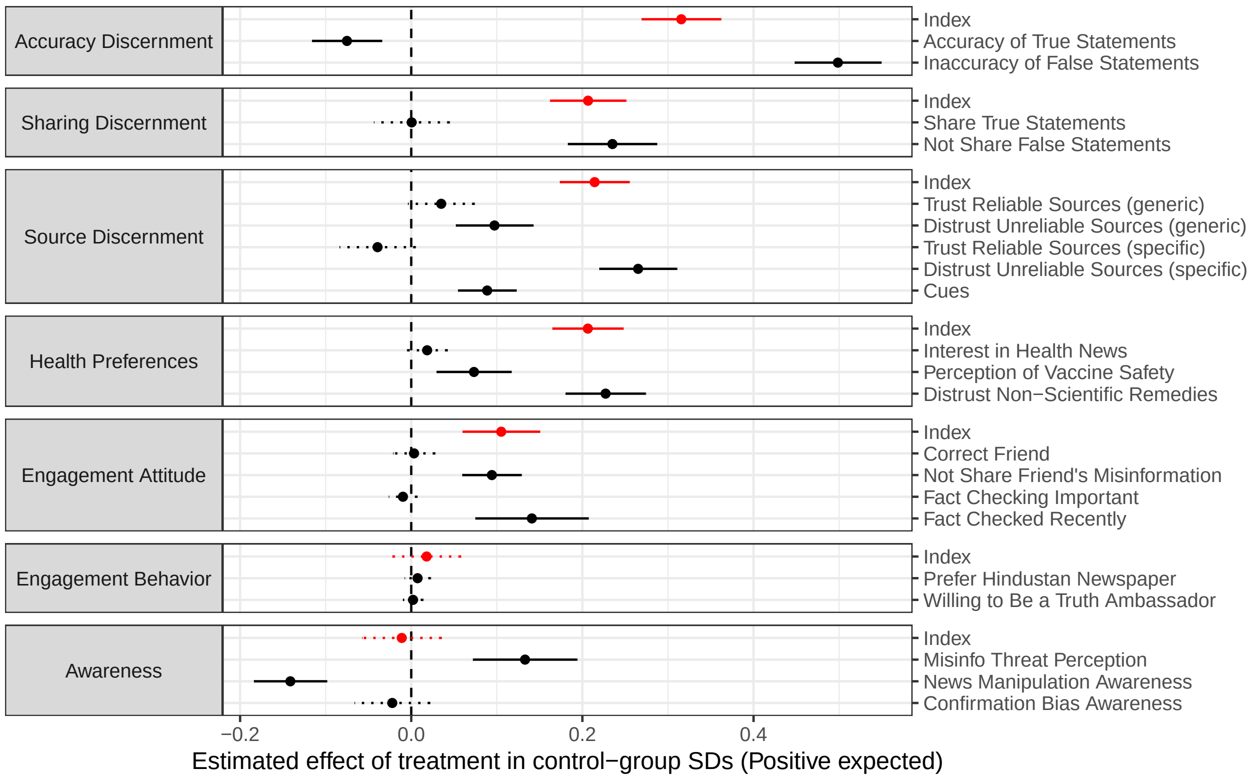
where Y_{ijk} is the primary outcome of interest Y for student i in classroom j and library-spillover strata $k \in \{1, ..., m\}$, β_0 is the intercept, T_{ijk} is a treatment indicator, α_c denotes the coefficient for the control variable X_c, γ_k denotes fixed effects for each library-spillover strata 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) using a 2SLS model and operationalizing compliance continuously ranging from 0 (0/4 sessions attended) to 1 (4/4 sessions attended).[14]

5.1 First Endline

We first examine the overall effect of BIMLI on outcomes from the first endline survey. Our main results are summarized in Figure 4, which shows the estimated effect of assignment to treatment on seven outcomes. The estimates of treatment effect we present in Figure 4 can be seen as conservative because of dilution due to partial non-compliance. We find that the treatment positively affected five of the seven outcomes and detail these results below. In Figure

[14] See Appendix J for CACE specification.

Figure 4: Estimated effect of assignment to BIMLI treatment



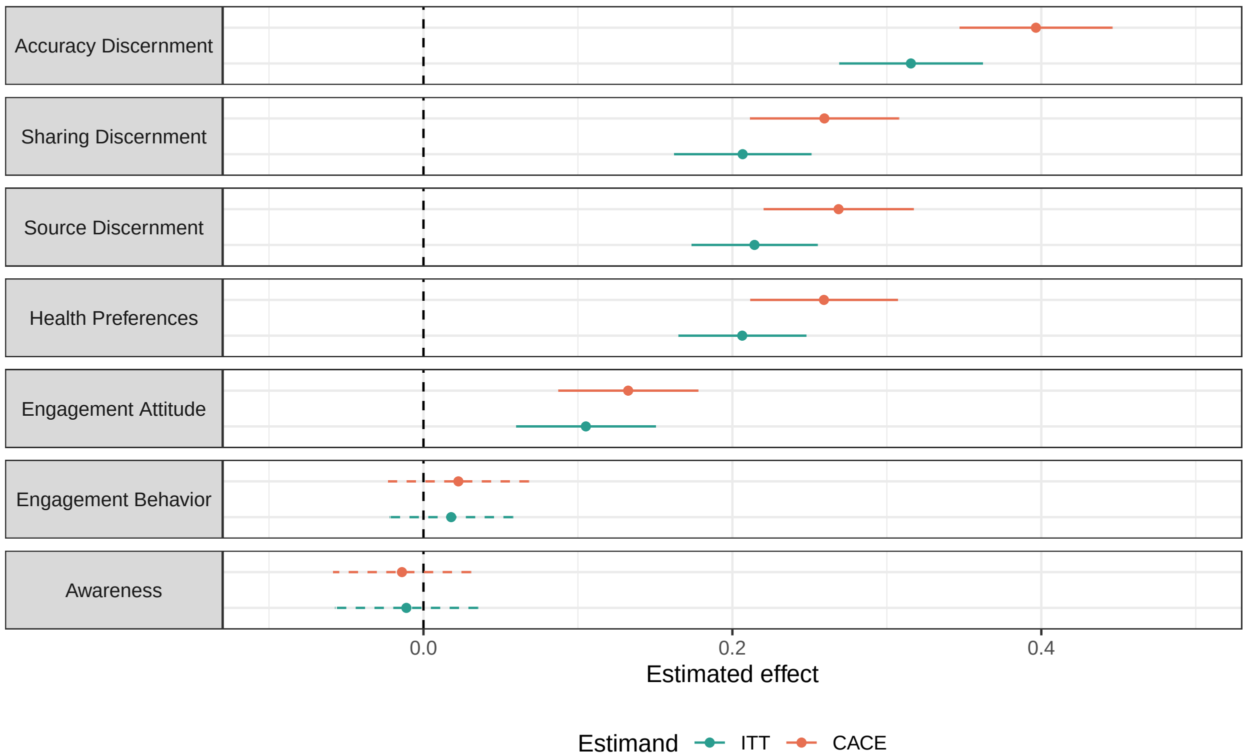
Notes: This figure plots the effect of assignment to BIMLI for 7 outcome families. Each index is an ICW calculation of components within an outcome family. Each component is standardized relative to the control mean and SD. Confidence intervals are at the 95% level and are based on standard errors clustered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

5 we compare ITT estimates to CACE, the causal effect among compliers. Tabular results for the ITT model are in Appendix G and for the CACE model are in Appendix J.

5.1.1 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 standard, 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

Figure 5: Estimated effect of BIMLI treatment on compliers



Notes: This figure plots a comparison of ITT versus CACE effects for 7 outcome families. Each index is an ICW calculation of components within an outcome family. Each component is standardized relative to 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 and model specification for CACE estimates are in Appendix J.

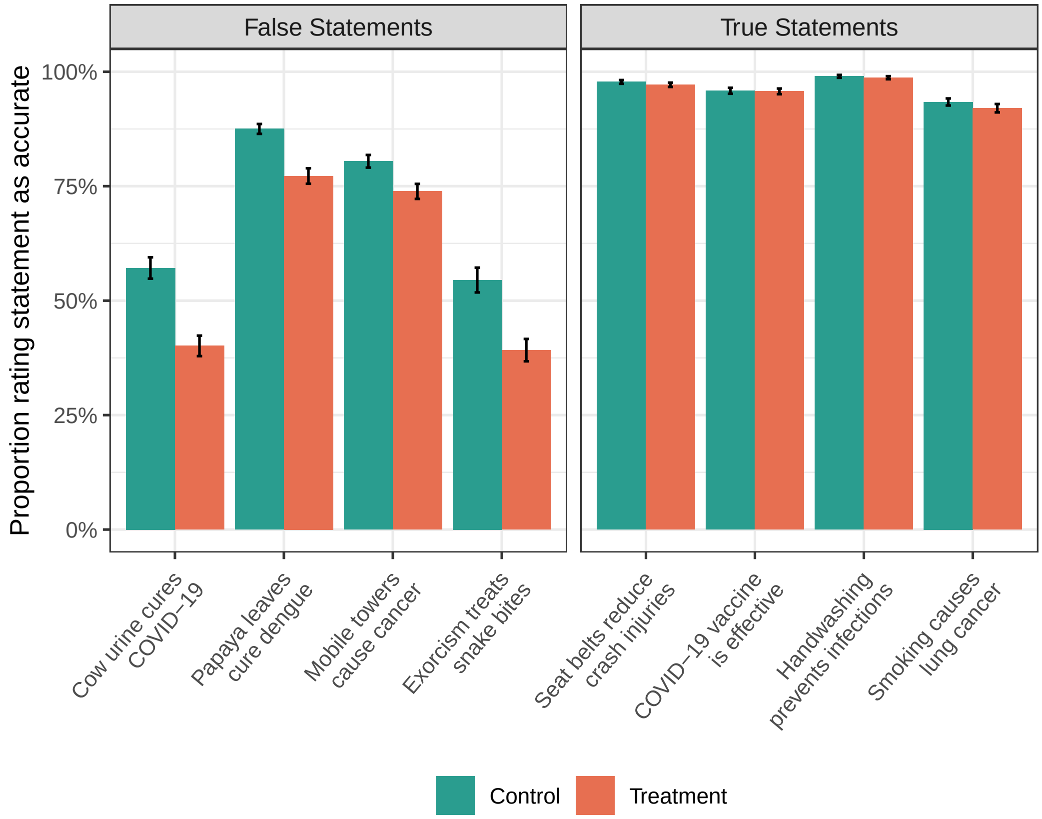
skill application rather than mere recall.[15] We also measured sharing discernment (intention to share each story) using the same items.[16] The selection for these stories was based on familiarity, cultural relevance, and minimizing the introduction of new misinformation. After fieldwork and piloting, we identified 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 4). The magnitude of this effect, a 0.32 SD increase in discernment relative to the control group, is substantively large

[15] We re-estimate effects dropping the 2 items discussed in class and find that results hold (Appendix J.6).

[16] 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 6: Accuracy discernment by headline



compared to discernment effects detected in research on misinformation in comparable contexts. 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. Further, when we compare ITT to CACE estimates, we find that the effect on accuracy discernment is even larger (0.39 SD) among compliers (see Figure 5). We also see variation in the true and false components of the discernment measure: we find that while the treatment made respondents marginally more skeptical of all news, the magnitude of this effect is much larger for the false statements, and hence the overall effect is a net positive. In Figure 6 we break down the true and false components into individual headlines. The graph demonstrates that large proportions of respondents in the control group thought that false statements were accurate 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.

While we observe varying effects on true and false stories, we attribute this to a ceiling effect for true stories, as belief in these was already high, leaving little room for further increase. Substantively, we do not view the effects on true stories as very concerning, insofar as effects are small, especially before standardization, and arguably caused by these ceiling effects. Moreover, in the current informational environment in India, we view a small degree of skepticism towards all information – especially given high baseline levels of trust in mass media in our sample – as normatively acceptable. Over the past decade, the quality of democracy has significantly declined, affecting the credibility of news sources. Many traditional media outlets have been acquired by government allies, leading to a pro-government bias across these outlets, with ideological diversity in available news sources narrowing (Mohan, 2021). While credible sources do exist, they are predominantly in English and paywalled, making them inaccessible to our sample. In such an environment, generalized skepticism may be warranted. Finally, we note that our second endline yields precise null effects on discernment measures for true information (Table 2). 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.

5.1.2 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 – both the mediums (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 two sources we expect to increase trust in (MBBS doctors, healthcare workers) and two we hope to decrease trust in (ayurvedic doctors, 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 remedies, WhatsApp forwards, TV interviews with ayurvedic doctors). This helps distinguish between general and specific trust and separates transmitters from mediums. Finally, we explore which factors foster trust in information, examining signals like likes/shares online, shared ethnicity, as well as message tone and emotionality. Our results show that BIMLI, overall, significantly changed how respondents interact with and trust sources, with a notable shift in the index (SD = 0.21). As seen from Figure 4, these effects are primarily driven by increased distrust in unreliable sources (both generic and specific).

5.1.3 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 chickenpox vaccines. To assess reliance on alternative medicine, respondents were asked if they would visit a traditional healer 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, we show that BIMLI was able to significantly alter respondents' health preferences (index 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 for several reasons. First, 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, 2024), motivated reasoning may impede the effectiveness of misinformation countermeasures (Taber and Lodge, 2006). Despite this, BIMLI had a significant impact on altering health preferences, including preferences regarding which forms of medicine to rely on during crises and beliefs about the efficacy of alternative treatments.

5.1.4 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.

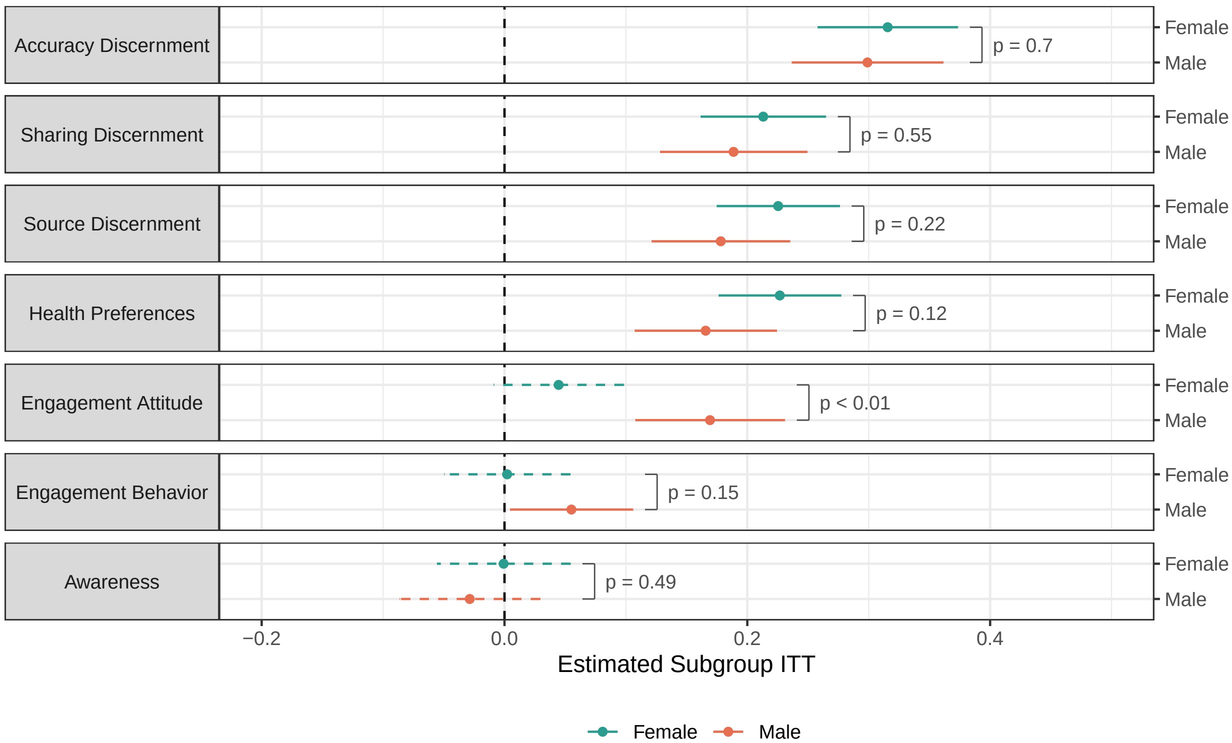
Children in India are accustomed to tests and often excel in educational settings. To ensure our findings were not solely driven by this familiarity, we incorporated two behavioral measures. First, respondents entered a lottery to choose between two subscriptions: a credible Hindi newspaper, *Hindustan*, or a popular magazine, *Manohar Kahaniyan*, featuring horror, mystery, and romance stories. We hypothesized greater demand for news among the treatment group, while noting that this measure is unrelated to misinformation directly. Second, we invited respondents to become “truth ambassadors,” a community role described as supporting local government by dispelling misinformation during crises, 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 reveals notable differences in misinformation engagement measures, even though indices for other outcomes show no such variation (Figure 7). 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 H.2). 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 broadly, public behaviors improved only among boys. Achieving similar changes among girls may require interventions that address gender norms alongside misinformation.

5.1.5 Awareness

This index assessed awareness of misinformation and recall of classroom material through five items. The first measured perceptions of misinformation as a threat. While exposure to BIMLI

Figure 7: 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 male and female coefficients.

significantly increased this perception, 78% of respondents already viewed misinformation as a threat, limiting room for further change. Awareness of media and cognitive biases was also measured using items adapted to the Indian context from Ashley, Maksl, and Craft (2013). These definitional items focused on 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. Treated respondents were also less likely to express agreement with statements suggesting ways in which the reporting of news in the media can be biased – a point that had been covered in session 2 (see Table 1). Overall we find a null effect on the awareness index. 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.

5.1.6 Heterogeneous treatment effects

We pre-registered heterogeneous effects analyses based on a number of variables. Most importantly, to proxy motivated reasoning, we examined 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 for recent elections. Matching this data back to our baseline, we were able to estimate party ID at the household level.[17] 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 looked at socio-economic status (proxied by an asset index), age (student grade) and gender. We also looked at basic science knowledge. Finally, we looked at whether results are different as function of being in a high- or low-spillover village. 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 mobile internet usage, age, or partisanship. 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 (Coppock, 2023), consistent with our argument about adolescents having less fixed political attachments.

5.1.7 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, lan-

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

guage of schooling), household-level variables (asset index as a proxy for wealth 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. Next, to determine whether the effects are concentrated in specific districts, we examine ITT estimates by district subgroup. The results show no clear pattern. Notably, given significant variation in districts' socio-economic development, we confirm that these findings are not correlated with development outcomes. Following this, we apply multiple corrections across indices, as pre-registered. Results on our main dependent variables remain significant. 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. We show that our analyses are robust to very high levels of differential attrition based on unobservables and find no evidence of differential attrition based on observables. All results on robustness checks are reported in Appendix J.

5.2 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 and potential network spillover impacts.[18] 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 government,[19] and (3) to examine spillover learning effects within respondents' networks. Given that students were encouraged to share classroom learnings with families, we interviewed one parent or guardian to evaluate whether treated students' families showed improved discernment abilities.

[18] Appendix I describes sampling for the second endline, including attrition and compliance, and information on those recontacted vs those who eventually answered.

[19] Teachers were explicitly instructed to avoid political content.

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 2. Moreover, treated respondents exhibited a significantly higher capacity to accurately assess the veracity of political stories (0.31 SD). This result is particularly striking given that the intervention exclusively targeted health-related content and did not address political claims. Notably, the political stories included in this survey were entirely novel and had not been evaluated during the initial endline assessment. Despite this, treated respondents demonstrated an enhanced ability to differentiate between true and false information on political topics. These findings underscore the broader implications of misinformation education: even when narrowly focused on a specific domain (such as health), educational interventions can yield transferable benefits across other domains. Further, these effects persist over time, highlighting the potential of such interventions to foster broad, enduring resilience against various forms of misinformation. We note that we observe very limited differences between the random follow-up sample that we recontacted versus those who eventually answered (including on outcomes from the first endline), implying that the persistence we observe likely generalizes to the whole sample.

Next, we find that parents/guardians of treated students were also significantly better at discerning true from false health information (0.27 SD), as demonstrated in Table 3. This result is particularly notable as it highlights the potential for "trickle-up" socialization, where children's learning influences their parents (Dahlgaard, 2018). It also suggests that sustained learning may generate valuable network spillover 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 flyers that family members could potentially view or discuss with them. We view this finding as noteworthy, underscoring that educative interventions, even with vulnerable populations such as children, can have effects that spill over 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).

Table 2: Effect of assignment to BIMLI treatment on 4-month follow-up

Outcome	Type	N	Estimate	SE	p-value
Accuracy Discernment	Index	1,944	0.26***	0.048	<0.001
Accuracy of True Statements	Sub-index	2,052	-0.06	0.040	0.14
Inaccuracy of False Statements	Sub-index	1,961	0.33***	0.041	<0.001
Political Discernment	Index	1,863	0.31***	0.049	<0.001
Accuracy of True Statements	Sub-index	1,991	-0.01	0.041	0.88
Inaccuracy of False Statements	Sub-index	1,887	0.31***	0.043	<0.001
Source Discernment	Index	2,027	0.10*	0.044	0.03
Trust Reliable Sources	Sub-index	2,039	-0.07	0.042	0.08
Distrust Unreliable Sources	Sub-index	2,055	0.14***	0.041	<0.001

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

Table 3: Effect of assignment to BIMLI on treatment group parents

Outcome	Type	N	Estimate	SE	p-value
Accuracy Discernment Index	Index	1,785	0.27***	0.054	<0.001
Accuracy of True Statements	Sub-index	2,019	-0.01	0.047	0.89
Inaccuracy of False Statements	Sub-index	1,803	0.27***	0.049	<0.001

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

5.3 Mechanisms

What explains these substantial and lasting effects? This section explores potential mechanisms underlying our findings. Respondents in the treatment group may have exhibited changes in attitudes and behavior through at least three pathways: reinforcement of classroom learning (*recall* mechanism), application of strategies and skills acquired during the intervention (*application* mechanism), or shifts in their perceptions of misinformation and expressions about it due to evolving social norms (*norms* mechanism).

Several factors suggest that our findings stem not just from recall of classroom content but also from substantive learning and application of new knowledge, and shifts in norms. Many outcome measures, such as willingness to correct friends, involve costly behaviors and intentions. In India, social norms of deference discourage confrontation, even when faced with misinformation (Malhotra and Pearce, 2022; Giles et al., 2003). The observed changes in behavioral intentions thus indicate shifts in norms, making social desirability or demand effects unlikely explanations; culturally, avoiding confrontation is often the socially desirable response (Savani, Morris, and Naidu, 2012). Additionally, the second endline survey included political stories—an entirely new topic not covered during the intervention or first endline. The findings here suggest respondents are not merely recalling information but are learning, retaining, and applying principles across real-world contexts. Finally, our discernment measure included two stories discussed in the classroom and six new ones. To validate the learning mechanism empirically, we recalculated discernment by excluding the two classroom stories, analyzing only the six introduced at the endline (Appendix J.6). The results are still large and significant, underscoring that the observed effects stem from respondents' ability to learn and apply new information rather than mere recall.

Lastly, to further examine underlying mechanisms, we included several descriptive questions in the second endline. First, we asked respondents in the treatment group to identify the pathways through which they believed the intervention had influenced them. Self-reported results indicate that 40 percent perceived the treatment to have taught them new strategies to consume information, 29 percent felt it corrected their existing beliefs, and 32 percent reported having better understood the normative importance of *not* sharing misinformation, suggesting the presence of multiple contributing mechanisms. Additionally, we provided respondents with a vignette scenario in which an acquaintance shares a piece of misinformation, and asked how they would respond. Treatment group respondents were significantly more inclined to report that they would demonstrate a verification strategy for future use and emphasize refraining from sharing misinformation (Appendix I). These findings suggest that mechanisms of learning and skill application, as well as norm shifts, were actively contributing to effects.[20]

6 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 in students' critical thinking and application of classroom lessons to new areas. By the program's

[20]We acknowledge that norms and learning are inherently intertwined, making it difficult to separate the two concepts. When a social referent, such as a teacher in this context, imparts a skill, it simultaneously conveys a norm due to their position of authority.

end, treatment 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 skills learned. Finally, we found that parents/guardians of treated students were significantly better at discernment, indicating that such educational interventions can have spillover 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.

They are significant given the mixed or null outcomes often seen in educative interventions so far, even when they focus on the Global North (Blair et al., 2023). By contrast, our program proved effective in a challenging context. Bihar, where the study took place, has low educational prioritization and high dropout rates, with 42% of students leaving school before 10th grade (Muralidharan and Prakash, 2017). Despite achieving about 70% session compliance, this lower rate reflects the challenges of rural areas with limited state capacity. Bihar consistently underperforms on living standards and public programs often show weak outcomes (Desai, 2019; Rasul and Sharma, 2014; Mathew and Moore, 2011; Jha, 2023). Thus it was not readily foreseeable that a curriculum like BIMLI would have such notable effects. Despite the politicization of health issues, adolescents engaged with the program, influencing their preferences, behaviors, and information consumption. While adults may be more resistant to change due to motivated reasoning (Taber and Lodge, 2006; Badrinathan, 2021), these findings support our argument that interventions should target populations more open to change—like young people seeking education in underserved areas.

Despite these encouraging findings, we recognize several limitations of the study. The intervention was delivered as a bundled treatment with multiple components, making it challenging to look at treatment effect heterogeneity based on the content of classes. Using session-wise attendance to pinpoint effective elements proves difficult, as session topics are confounded with factors like peer effects. For instance, students attending the first class may form distinct social networks, leading to downstream effects that are endogenous to the first class they attend. Additionally, each session's content was recapped in subsequent sessions, further complicating efforts to isolate the impact of specific topics. Our goal was to design a comprehensive, high-dosage intervention, addressing the limited success of many media and information literacy programs. Future extensions of this work could unbundle the curriculum to determine where the effects are most concentrated. Another limitation is the high cost and effort required to set up such an intensive and sustained intervention, raising the question of whether this dosage of treatment is necessary. Indeed, many studies use brief fact-checks or primes and nudges and find effects. To provide suggestive evidence of the required dosage to significantly affect outcomes, we computed subgroup ITTs by session attendance (Appendix Figure C.10), and found that attending one session had no significant effects, consistent with other research in this context showing that one-time treatments are insufficient (Badrinathan, 2021). Two sessions significantly improved accuracy, sharing discernment, and source discernment. However, shifting health preferences and attitudes toward counter-misinformation efforts required at least three sessions.[21]

These findings highlight two key points: First, discernment tasks are more easily influenced but require at least two sessions (that is, at least three contact hours) for significant change. Second, outcomes requiring three or more sessions typically involve deeply ingrained attitudes, such as beliefs in Ayurveda, which are harder to shift than newer misinformation narratives. Research on polarization and social contact supports the difficulty of changing attitudes on polarizing topics, even with intensive interventions (Scacco and Warren, 2018; Mousa, 2020). Our results demonstrate that while challenging, deep-rooted attitudes can change through sustained, education-based efforts. This underscores the need to move beyond light-touch interventions and prioritize sustained engagement for meaningful attitude shifts.

Finally, we discuss the generalizability of our results. As stated earlier, our study sample is characterized by low state capacity, limited access to credible news, and low socioeconomic status. To make the intervention work in this challenging context, we made specific design choices, such as bringing in external teachers to deliver the curriculum and partnering with a

[21]For a detailed explanation of these results and why non-random attendance warrants caution when drawing conclusions from this analysis, see Appendix C.

well-known government agency for rollout. One reason the program may have been successful is that it stood out in this context—whether due to its teaching style, or being a rare high-quality educational opportunity. Data suggest it might be the case: over 95% of parents surveyed, in both treatment and control groups, indicated they would enroll their children in such a program again. Among these, 70% cited a desire for their children to learn, while 27% highlighted trust in Jeevika, the state agency, as the key reasons for their willingness to enroll their children again (Appendix I.2). Further, we acknowledge that our village selection was based on convenience, specifically proximity to libraries, so students could walk to classes. As a result, we cannot ascertain the prior vulnerability to misinformation in this sample, as we lacked baseline measures and did not select villages based on this factor. Census data show that villages in our sample are slightly more developed than those further away (Appendix A), though it is unclear if this correlates with vulnerability to misinformation.[22] While there are some differences between villages within and outside our sample, these are not large, and statistically significant differences should not be interpreted as evidence that we sampled from a privileged population. Jeevika's libraries are in fact by definition located in rural or semi-rural locations, across India's poorest state.

These points nonetheless raise questions about the intervention's generalizability: (1) its effectiveness within existing public systems in India or Bihar, and (2) its applicability to children in these systems, who may differ from our sample. While we acknowledge these limitations, we note that our data shows minimal variation in treatment effects based on income, socioeconomic status, or political variables, suggesting the intervention could have similar impacts across diverse populations (see Appendix H). Many regions in the Global South share challenges similar to Bihar – low state capacity, limited access to credible news, and socioeconomic inequalities – where misinformation spreads through informal networks rather than online platforms (Gadjanova, Lynch, and Saibu, 2022). Given these shared issues, our findings suggest that education-based interventions like ours could be effective in other regions with similar conditions.

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

[22] Previous research suggests that vulnerability to misinformation is not determined by income or socio-economic status (Maffioli and Gonzalez, 2022).

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, even if intensive and costly. 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. To provide more insight into the applicability of such an intervention as policy, and recognizing that its success is a function of both teaching quality as well as curriculum, we investigated whether certain teacher demographics influenced outcomes. We found that female teachers delivered significantly better results across many measures (see Appendix E). Interestingly, despite significant religious polarization in north India and evidence of discrimination against religious and caste minorities, teacher religion and caste identity did not significantly affect outcomes.

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. Therefore, our study not only contributes to 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.

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.

Baul, Tushi, Dean Karlan, Kentaro Toyama, and Kathryn Vasilaky. 2024. "Improving smallholder agriculture via video-based group extension." *Journal of Development Economics*: 103267.

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. 2015. "In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media." *Journal of Communication* 65 (4): 619–638.

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. 2024. "Sustained exposure to fact-checks can innoculate citizens against misinformation in the Global South." *Forthcoming, American Political Science Review,*.

Brashier, Nadia M. 2024. "Fighting Misinformation Among the Most Vulnerable Users." *Current Opinion in Psychology*: 101813.

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*. 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 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.

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.

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 Kiran Garimella. 2022. "What circulates on partisan WhatsApp in India? Insights from an unusual dataset." *Journal of Quantitative Description: Digital Media* 2.

Chauchard, Simon, and Sumitra Badrinathan. 2024. "The Religious Roots of Belief in Misinformation: Experimental Evidence from India." *Working paper*.

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*. 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.

Datta, Upamanyu. 2015. "Socio-economic impacts of JEEViKA: A large-scale self-help group project in Bihar, India." *World Development* 68: 1–18.

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.

Desai, Sonalde et al. 2019. *Human Development in India: Challenges for a Society in Transition*. 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. 2024. "Creating cohesive communities: A youth camp experiment in india." *Working Paper*.

Giles, Howard, Kimberly A Noels, Angie Williams, Hiroshi Ota, Tae-Seop Lim, Sik Hung Ng, Ellen B Ryan, and Lilnabeth Somera. 2003. "Intergenerational communication across cultures: Young people's perceptions of conversations with family elders, non-family elders and same-age peers." *Journal of cross-cultural gerontology* 18: 1–32.

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. 2023. Less Partisan but No More Competent: Expressive Responding and Fact-Opinion Discernment. Technical report.

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.

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

Hoffmann, Vivian, Vijayendra Rao, Vaishnavi Surendra, and Upamanyu Datta. 2021. "Relief from usury: Impact of a self-help group lending program in rural India." *Journal of Development Economics* 148: 102567.

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 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.

Jones-Jang, S Mo, Tara Mortensen, and Jingjing Liu. 2021. "Does media literacy help identification of fake news? Information literacy helps, but other literacies don't." *American Behavioral Scientist* 65 (2): 371–388.

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

Kumar, Deepak, Bhanu Pratap, and Archana Aggarwal. 2021. "Public trust in state governments in India: Who are more confident and what makes them confident about the government?" *Asian Journal of Comparative Politics* 6 (2): 154–174.

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

Kumar, Sanjay. 2019. "In India, Hindu pride boosts pseudoscience." *Science Magazine* 363: 679–680.

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.

Maffioli, Elisa M, and Robert Gonzalez. 2022. "Are socio-demographic and economic characteristics good predictors of misinformation during an epidemic?" *PLoS Global Public Health* 2 (3): e0000279.

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

Markus, Hazel, and Paula Nurius. 1986. "Possible selves." *American Psychologist* 41 (9): 954.

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/.

Moore, Ryan C, and Jeffrey T Hancock. 2022. "A digital media literacy intervention for older adults improves resilience to fake news." *Scientific Reports* 12 (1): 6008.

Mousa, Salma. 2020. "Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq." *Science* 369 (6505): 866–870.

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.

Niemi, Richard G, and M Kent Jennings. 1991. "Issues and inheritance in the formation of party identification." *American Journal of Political Science*: 970–988.

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 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, and Sander van der Linden. 2019. “Fake news game confers psychological resistance against online misinformation.” *Palgrave Communications* 5 (1): 1–10.

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.

Rossini, Patrícia, Jennifer Stromer-Galley, Erica Anita Baptista, and Vanessa Veiga de Oliveira. 2020. “Dysfunctional information sharing on WhatsApp and Facebook: The role of political talk, cross-cutting exposure and social corrections.” *New Media & Society*: 14-29.

Savani, Krishna, Michael W Morris, and NVR Naidu. 2012. “Deference in Indians’ decision making: Introjected goals or injunctive norms?” *Journal of Personality and Social Psychology* 102 (4): 685.

Scacco, Alexandra, and Shana S Warren. 2018. “Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria.” *American Political Science Review* 112 (3): 654–677.

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.

Valeriani, Augusto, and Cristian Vaccari. 2018. “Political talk on mobile instant messaging services: A comparative analysis of Germany, Italy, and the UK.” *Information, Communication & Society* 21 (11): 1715–1731.

van Prooijen, Jan-Willem, and Nienke Böhm. 2024. “Do conspiracy theories shape or rationalize vaccination hesitancy over time?” *Social Psychological and Personality Science* 15 (4): 421–429.

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

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 is 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 in the second round 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 within 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 H17).

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.

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.

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, 2024). 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 = ≈ 300), we calculated that we would need to sample 14,000 students in about 600 clusters in order to achieve 80% power (see PAP for more details).

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 (means)

	N	**Treatment**	**Control**	**Diff.**	**SE**	**p-value**
No. of households	546	1630.541	1830.761	-200.22	483.34	0.679
Total population	546	8767.474	9722.783	-955.31	2486.25	0.701
Share SC population	546	.167	.17	-0.00	0.01	0.780
Share ST population	546	.009	.008	0.00	0.00	0.603
Share literate	546	.505	.516	-0.01	0.01	0.165
Total area (km)	546	477.478	505.17	-27.69	55.19	0.616
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	0.416
No. of secondary schools	546	.474	.37	0.10	0.06	0.104
Pucca road	546	.837	.801	0.04	0.03	0.271
Power supply (domestic)	546	.804	.772	0.03	0.04	0.362
Power supply (agricultural)	546	.444	.435	0.01	0.04	0.821

*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 - Female	13,590	0.590	0.580	0.01	0.011	0.62	0.72
Grade	13,589	9.700	9.620	0.08	0.034	0.01*	0.12
Age	13,590	14.960	14.850	0.09	0.040	0.02*	0.12
Religion - Hindu	13,590	0.900	0.910	-0.01	0.013	0.51	0.72
Language - Hindi	13,590	0.420	0.440	-0.02	0.014	0.13	0.31
Caste - GEN	13,392	0.080	0.090	-0.01	0.010	0.26	0.54
Caste - OBC/EBC	13,392	0.690	0.700	-0.01	0.019	0.48	0.72
Caste - SC	13,392	0.210	0.190	0.03	0.017	0.13	0.31
Caste - ST	13,392	0.020	0.020	0.00	0.004	>0.9	>0.9
Asset Index	13,590	-0.010	0.010	-0.05	0.028	0.07	0.3
Father's Education	12,890	6.920	6.920	-0.03	0.117	0.76	0.84
Mother's Education	12,950	4.070	4.160	-0.13	0.122	0.3	0.58
Government School	13,590	0.960	0.960	0.01	0.004	0.1	0.3
Science Knowledge	13,590	4.290	4.290	-0.02	0.038	0.58	0.72
Mobile Internet	13,590	0.180	0.190	0.00	0.008	>0.9	>0.9
Trust Newspapers	13,590	0.910	0.900	0.01	0.006	0.39	0.64
Trust Social Media	13,590	0.620	0.590	0.03	0.011	0.02*	0.12
Trust TV	13,590	0.850	0.830	0.01	0.008	0.08	0.3
Trust Friends and Family	13,590	0.970	0.970	0.00	0.004	0.37	0.64
Trust Vaccinated	13,590	0.780	0.760	0.02	0.011	0.08	0.3
Trust Ayurveda	13,590	0.870	0.870	0.00	0.007	0.61	0.72

*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 (means)

	N	**Sample**	**Non-Sample**
No. of households	39192	1731.751	439.499
Total population	39192	9250.377	2404.27
Share SC population	39192	.169	.185
Share ST population	39192	.008	.026
Share literate	39192	.511	.493
Total area (km)	39192	491.476	220.99
No. of primary schools	39192	2.965	1.389
No. of middle schools	39192	1.538	.677
No. of secondary schools	39192	.421	.158
Pucca road	39192	.819	.637
Power supply (domestic)	39192	.788	.614
Power supply (agricultural)	39192	.44	.325

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 B1 the slide discusses some common reasons why people are vulnerable to misinformation; this was included in the first module. In Figure B2 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 B3 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 B1: Slide deck example outlining reasons for vulnerability to misinformation

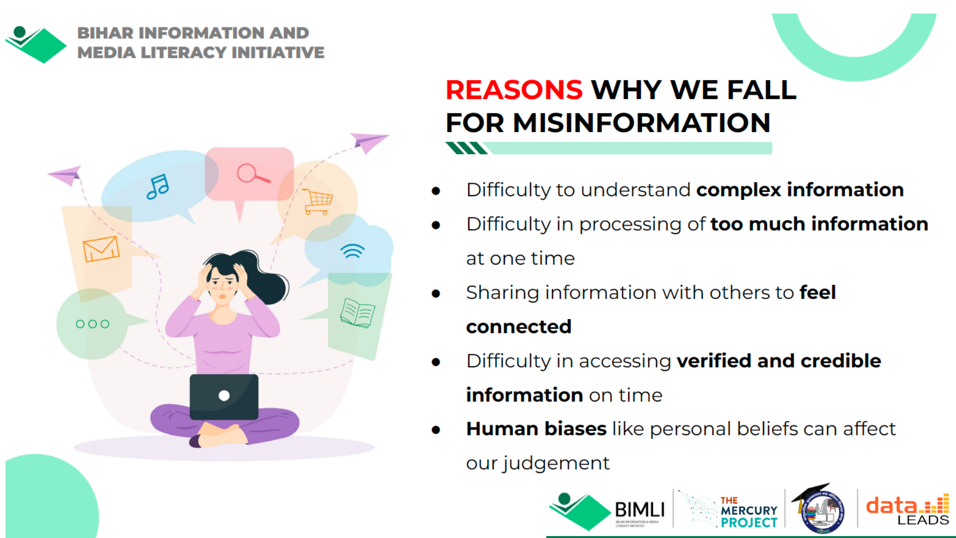
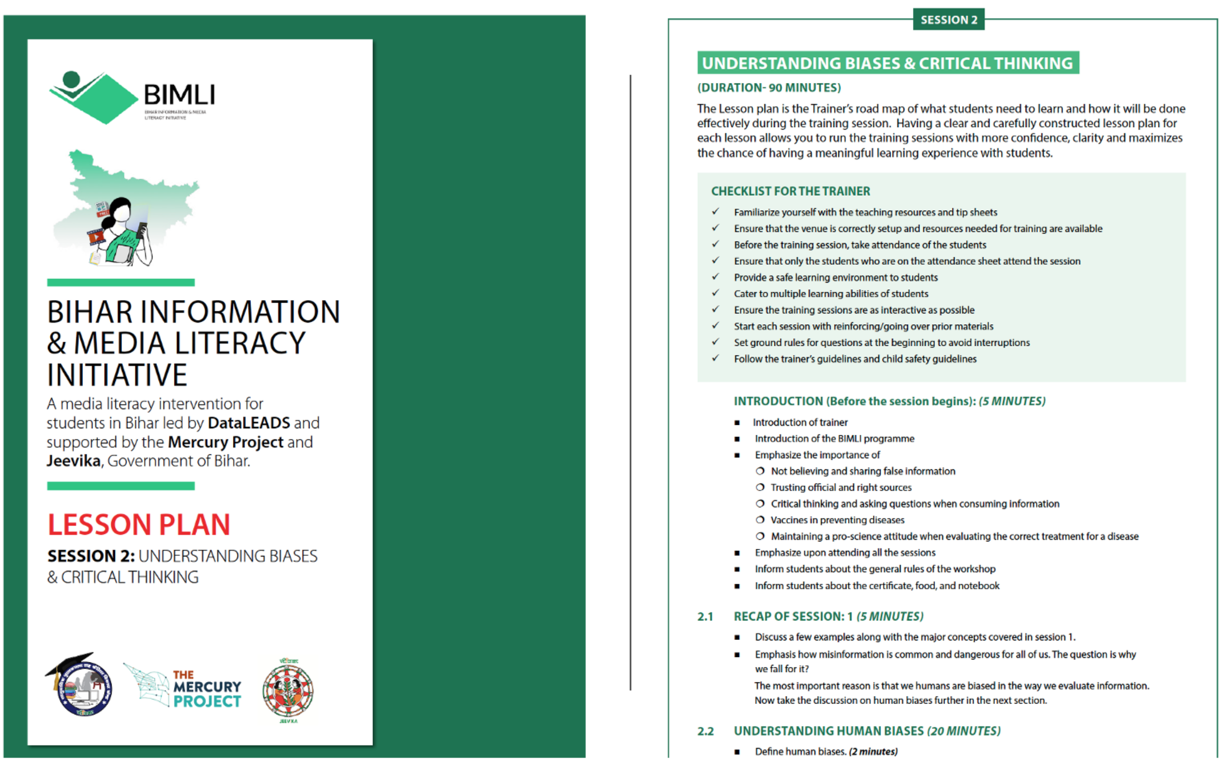


Figure B2: Slide deck example of a real-life consequence of misinformation in India

Figure B3: 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 B4 that was used for the second session.

Figure B4: Session 2 lesson plan example


BIMLI

BIHAR INFORMATION & MEDIA LITERACY INITIATIVE

A media literacy intervention for students in Bihar led by **DataLEADS** and supported by the **Mercury Project** and **Jeevika**, Government of Bihar.

LESSON PLAN

SESSION 2: UNDERSTANDING BIASES & CRITICAL THINKING

THE MERCURY PROJECT

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 B5: 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 B6 provides an example.

Figure B6: Lesson summary tip-sheets distributed to treatment students

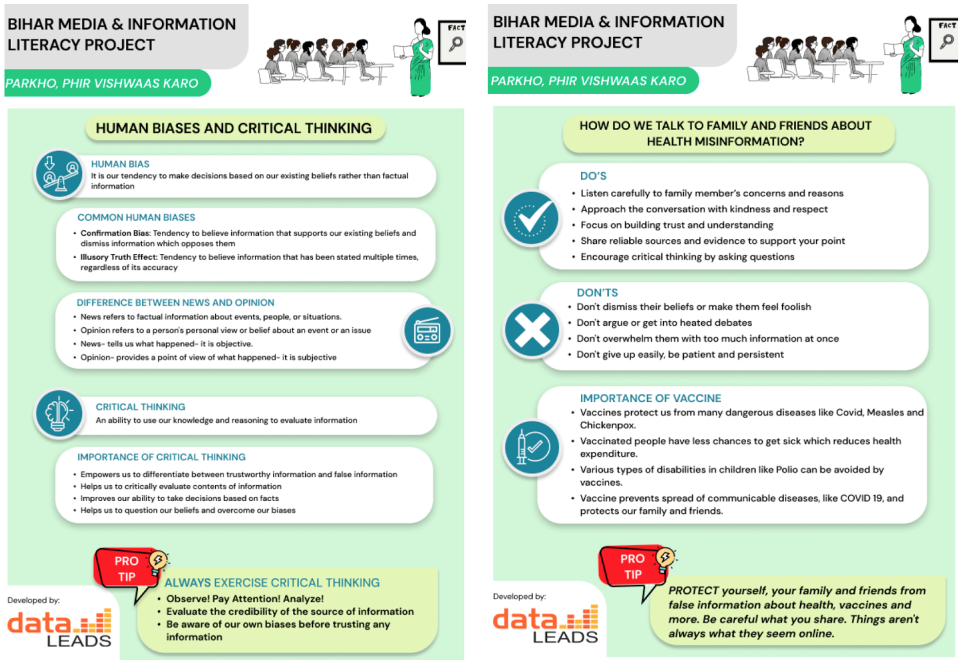
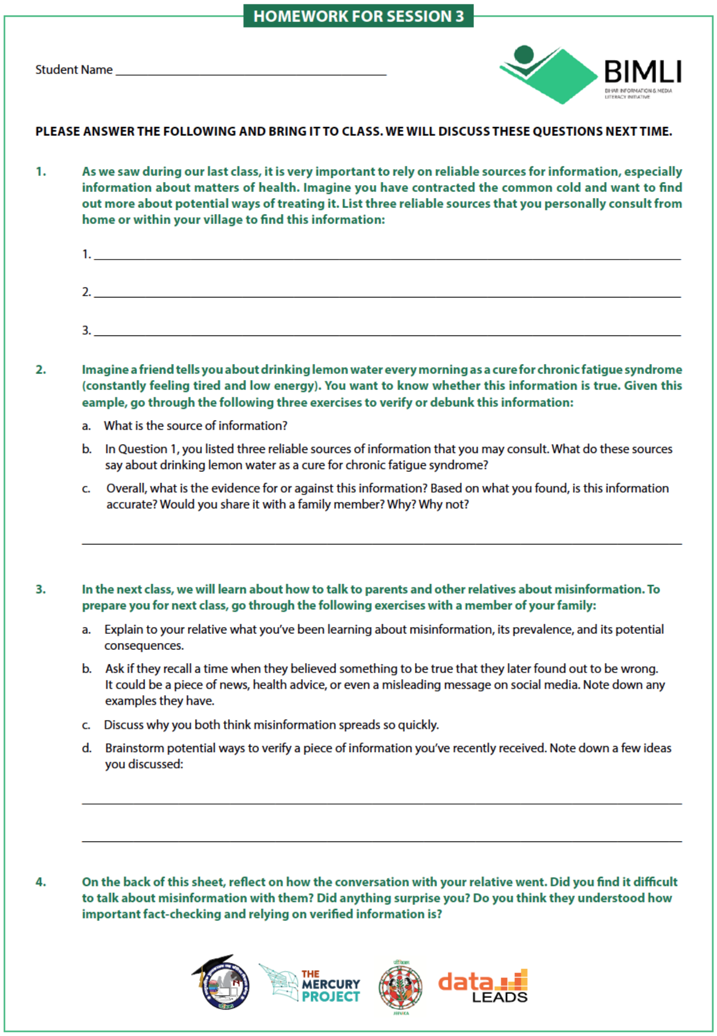


Figure B7: Example of a treatment classroom in session (Vaishali, Bihar)

5. A set of 3 **homework assignment 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 B8 provides an example of a take-home homework assignment.

Figure B8: 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?

THE MERCURY PROJECT

data LEADS

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 B9 provides examples.

Figure B9: Flip chart examples from English classes

C Compliance and attrition

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 a notebook and a pen; 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 continuous measure from 0 to 1 according to how many of the four sessions a particular student attended (we measure this with respondent-level attendance data gathered by teachers during each session). We have three compliance categories in our data:

1. **Non-compliers:** Participants who attended 0 out of 4 sessions.
2. **Partial compliers:** Participants who attended at least 1 out of 4 sessions, measured continuously
3. **Full compliers:** Participants who attended all 4 sessions.

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 continuous measure of compliance from 0-4 sessions attended. 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 that these effects were not different between treatment and control groups. While the coefficient on social media is significant (indicating that among individuals assigned to treatment, social media usage has a significantly larger effect on compliance), this effect is substantively small and barely significant. Moreover this table estimates 27 coefficients; with a 95% CI, we would expect 1.35 estimates to be significant under the null by chance alone.

In terms of attrition, Table C6 highlights that female students and Hindu students attrited at lower rates (i.e., they were more likely to complete the 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 C7 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 C4: Compliance Predictors

Predictor	**N**	**Estimate**	**SE**	**p-value**
Gender - Female	13,590	0.36***	0.030	<0.001
Grade	13,589	-0.12***	0.011	<0.001
Age	13,590	-0.11***	0.009	<0.001
Religion - Hindu	13,590	-0.02	0.062	0.76
Language - Hindi	13,590	-0.04	0.038	0.26
Asset Index	13,590	0.00	0.013	0.72
Father's Education	12,890	0.01*	0.003	0.03
Mother's Education	12,950	0.01**	0.003	0.002
Government School	13,590	0.09	0.070	0.21
Science Knowledge	13,590	0.03*	0.010	0.01
Mobile Internet	13,590	0.03	0.034	0.33
Newspapers	13,590	-0.05	0.039	0.23
Social Media	13,590	-0.02	0.026	0.55
TV	13,590	0.03	0.032	0.39
Friends and Family	13,590	0.02	0.069	0.77
Vaccinated	13,590	0.01	0.033	0.67
Ayurveda	13,590	0.04	0.035	0.32

*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 - Female * T	13,590	0.04	0.060	0.49
Grade * T	13,589	0.02	0.022	0.46
Age * T	13,590	0.01	0.017	0.42
Religion - Hindu * T	13,590	0.07	0.112	0.52
Language - Hindi * T	13,590	-0.06	0.067	0.37
Asset Index * T	13,590	-0.03	0.027	0.32
Father's Education * T	12,890	-0.01	0.006	0.32
Mother's Education * T	12,950	-0.01	0.006	0.18
Government School * T	13,590	0.18	0.142	0.2
Science Knowledge * T	13,590	0.00	0.019	0.82
Mobile Internet * T	13,590	-0.05	0.065	0.44
Newspapers * T	13,590	0.08	0.077	0.28
Social Media * T	13,590	0.12*	0.052	0.02
TV * T	13,590	0.06	0.065	0.33
Friends and Family * T	13,590	0.22	0.131	0.1
Vaccinated * T	13,590	-0.02	0.064	0.72
Ayurveda * T	13,590	0.10	0.072	0.18

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

Table C6: Attrition Predictors

Predictor	**N**	**Estimate**	**SE**	**p-value**
Treatment - Media Literacy	13,590	0.00	0.005	0.42
Gender - Female	13,590	-0.06***	0.006	<0.001
Grade	13,589	0.02***	0.002	<0.001
Age	13,590	0.02***	0.002	<0.001
Religion - Hindu	13,590	-0.03**	0.012	0.005
Language - Hindi	13,590	0.00	0.007	0.82
Asset Index	13,590	0.01	0.003	0.07
Father's Education	12,890	0.00**	0.001	0.006
Mother's Education	12,950	0.00*	0.001	0.04
Government School	13,590	-0.02	0.016	0.34
Science Knowledge	13,590	0.00	0.002	0.16
Mobile Internet	13,590	0.00	0.007	0.68
Newspapers	13,590	0.00	0.010	0.76
Social Media	13,590	0.00	0.006	0.87
TV	13,590	0.00	0.008	0.68
Friends and Family	13,590	0.00	0.017	0.8
Vaccinated	13,590	0.00	0.007	0.59
Ayurveda	13,590	-0.01	0.009	0.34

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

Table C7: Attrition Predictors (*Treatment)

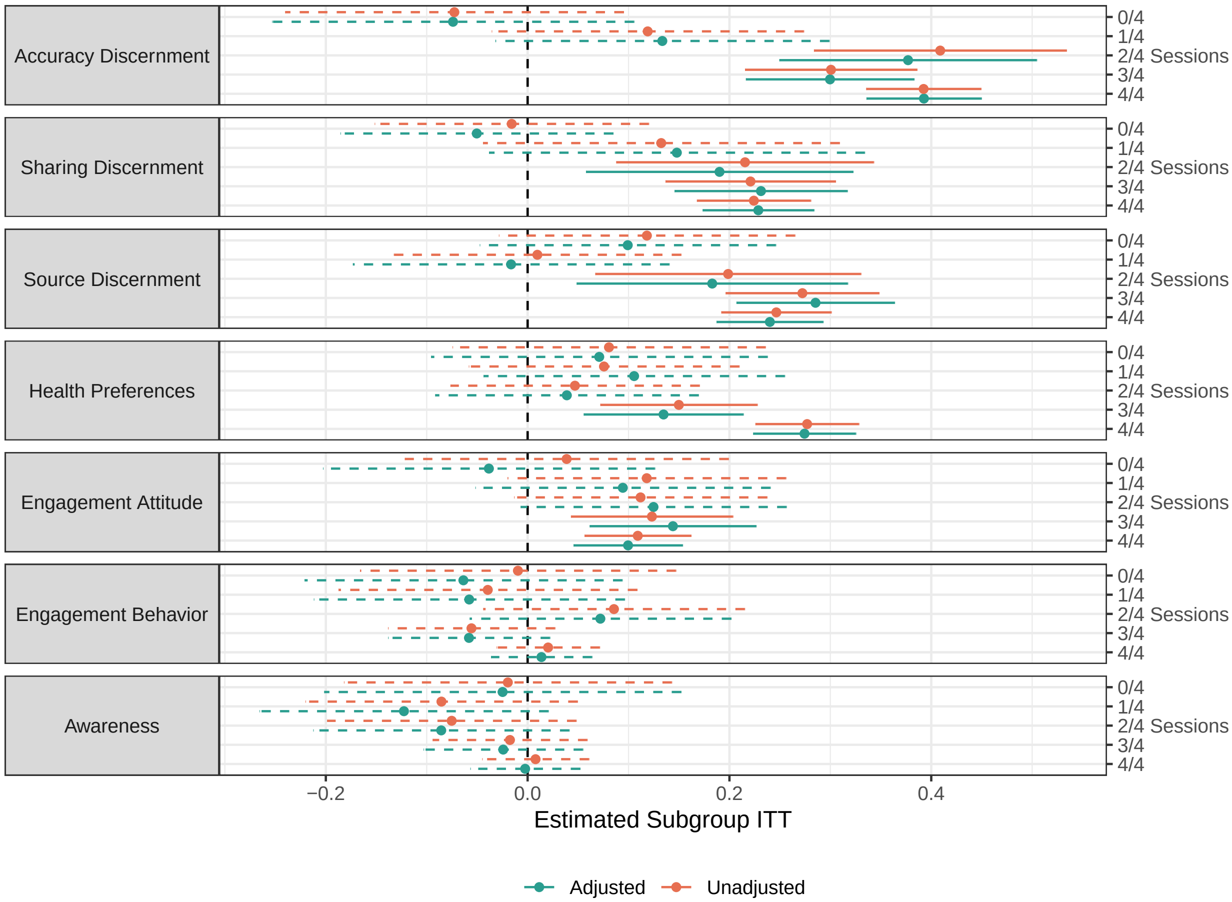
Predictor	**N**	**Estimate**	**SE**	**p-value**
Gender - Female * T	13,590	-0.02	0.012	0.14
Grade * T	13,589	0.00	0.005	0.84
Age * T	13,590	0.00	0.004	0.43
Religion - Hindu * T	13,590	0.01	0.023	0.62
Language - Hindi * T	13,590	0.00	0.011	>0.9
Asset Index * T	13,590	0.01	0.006	0.27
Father's Education * T	12,890	0.00	0.001	0.39
Mother's Education * T	12,950	0.00	0.001	0.06
Government School * T	13,590	-0.02	0.032	0.56
Science Knowledge * T	13,590	-0.01	0.004	0.1
Mobile Internet * T	13,590	0.01	0.014	0.42
Newspapers * T	13,590	-0.01	0.019	0.67
Social Media * T	13,590	-0.01	0.012	0.36
TV * T	13,590	0.00	0.015	0.77
Friends and Family * T	13,590	-0.02	0.034	0.46
Vaccinated * T	13,590	-0.01	0.014	0.35
Ayurveda * T	13,590	0.00	0.017	0.79

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

To assess whether attending more sessions leads to improved outcomes, we re-estimate our main ITT specification (once without adjusting for covariates and once adjusting for covariates listed in Appendix J.2) for subgroups by compliance level for each of our main outcome indices. Figure C10 illustrates that, as expected, the ITT estimates among the subgroup of subjects who attended 0/4 sessions are statistically indistinguishable from 0 at $p < 0.05$.

For the outcomes for which we see the largest overall effect sizes (accuracy and sharing discernment), attending 2/4 sessions seems to be sufficient to move outcomes. Health preferences seem to require extensive engagement with the program; we only see significant effects among students who attended three or more sessions. This fits well with our expectation – informed by existing findings – that health preferences may be especially hard to move in our study context. Engagement outcomes are also moved by attending at least three sessions.

Figure C10: Subgroup ITTs by number of sessions attended



While we consider these analyses useful as suggestive evidence of how effects vary by dosage, we do note that attendance is non-random (see Table C4), so the subsamples with which we estimate the subgroup ITTs are not directly comparable to each other in terms of background characteristics.

D 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, enhances the robustness of the measure. This method 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 is posted to our PAP.

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	(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	(lower values indicate higher levels of believed accuracy)	
3. Sharing discernment	Sharing intentions of false statements	sharing1; sharing2; sharing4; sharing5	(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	(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) — standardized ICW index of bad sources (word of mouth; ayurvedic 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)	

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.

Table E8: Summary statistics of teachers, by treatment condition

	Treatment	Control
Statistic	Mean (SD)/Prop.	Mean (SD)/Prop.
Age	35.864 (9.05)	30.75 (8.66)
Female	.25	.167
Male	.75	.833
Caste - SC/ST	.035	.022
Caste - GEN	.267	.333
Caste - OBC	.156	.351
Caste - Prefer not to say	.556	.281
Religion - Hindu	.583	.952
Religion - Muslim	.25	.048
Religion - Other / Prefer not to say	.167	.
School - Government	.75	.619
School - Private	.25	.333
School - Other	.00	.048
N	45	57

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). Note that in this model we use only district fixed effects, since in most cases the same teacher taught all classes at one library location. We find that for three outcomes where we detect a large main effect, the presence of female teachers appears to significantly impact these outcomes for the better. 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 E9: HTEs by Teacher Demographics

Outcome	**Term**	**N**	**Estimate**	**SE**	**p-value**
Awareness Index	T x Female teacher	11,154	-0.02	0.111	0.85
Awareness Index	T x Muslim teacher	11,154	0.03	0.120	0.82
Awareness Index	T x General caste teacher	11,154	0.01	0.080	0.86
Accuracy Discernment Index	T x Female teacher	10,552	0.17*	0.081	0.03
Accuracy Discernment Index	T x Muslim teacher	10,552	-0.05	0.099	0.6
Accuracy Discernment Index	T x General caste teacher	10,552	0.06	0.067	0.37
Sharing Discernment Index	T x Female teacher	10,694	0.04	0.089	0.68
Sharing Discernment Index	T x Muslim teacher	10,694	0.02	0.096	0.84
Sharing Discernment Index	T x General caste teacher	10,694	0.01	0.066	0.89
Health Preferences Index	T x Female teacher	10,735	0.19*	0.089	0.03
Health Preferences Index	T x Muslim teacher	10,735	0.08	0.111	0.47
Health Preferences Index	T x General caste teacher	10,735	0.05	0.062	0.41
Source Discernment Index	T x Female teacher	10,865	0.16*	0.074	0.03
Source Discernment Index	T x Muslim teacher	10,865	0.02	0.083	0.76
Source Discernment Index	T x General caste teacher	10,865	-0.05	0.066	0.45
Engagement Attitude Index	T x Female teacher	11,235	0.01	0.100	>0.9
Engagement Attitude Index	T x Muslim teacher	11,235	-0.08	0.091	0.35
Engagement Attitude Index	T x General caste teacher	11,235	0.14	0.071	0.05
Engagement Behavior Index	T x Female teacher	11,365	0.01	0.074	>0.9
Engagement Behavior Index	T x Muslim teacher	11,365	-0.03	0.091	0.72
Engagement Behavior Index	T x General caste teacher	11,365	-0.02	0.061	0.69

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

F Ethics

The study was approved by IRB (expedited review). 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 J18).

Figure F12: 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
Index	—	11,122	0.32***	0.024	<0.001
Accuracy of True Statements	Sub-index	11,933	-0.08***	0.021	<0.001
Inaccuracy of False Statements	Sub-index	11,159	0.50***	0.026	<0.001
Inaccuracy: False Statement1	Question	11,571	0.44***	0.028	<0.001
Inaccuracy: False Statement2	Question	11,616	0.30***	0.020	<0.001
Inaccuracy: False Statement3	Question	11,815	0.18***	0.020	<0.001
Inaccuracy: False Statement4	Question	11,972	0.40***	0.031	<0.001
Accuracy: True Statement1	Question	11,996	-0.03**	0.009	0.002
Accuracy: True Statement2	Question	11,973	0.00	0.011	0.72
Accuracy: True Statement3	Question	11,995	-0.03***	0.007	<0.001
Accuracy: True Statement4	Question	11,985	-0.04*	0.014	0.01

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

Table G11: Sharing Discernment

Outcome	Type	N	Estimate	SE	p-value
Index	—	11,264	0.21***	0.023	<0.001
Share True Statements	Sub-index	11,895	0.00	0.023	>0.9
Not Share False Statements	Sub-index	11,320	0.24***	0.026	<0.001
Not Share False Statement1	Question	11,651	0.11***	0.015	<0.001
Not Share False Statement2	Question	11,710	0.11***	0.014	<0.001
Not Share False Statement3	Question	11,862	0.08***	0.015	<0.001
Not Share False Statement4	Question	11,955	0.10***	0.015	<0.001
Share True Statement1	Question	11,991	-0.01	0.008	0.1
Share True Statement2	Question	11,949	0.00	0.010	0.78
Share True Statement3	Question	11,990	0.00	0.008	>0.9
Share True Statement4	Question	11,971	0.01	0.010	0.27

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

Table G12: Source Discernment

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**
Index	—	11,482	0.21***	0.021	<0.001
Trust Reliable Sources (generic)	Sub-index	11,948	0.03	0.020	0.08
Distrust Unreliable Sources (generic)	Sub-index	11,923	0.10***	0.023	<0.001
Generic Trust: MBBS Doctors	Question	11,977	0.00	0.011	0.79
Generic Trust: Health Workers	Question	11,973	0.03*	0.014	0.02
Generic Trust: Ayurvedic Doctors	Question	11,961	0.04	0.021	0.05
Generic Trust: Word of Mouth	Question	11,961	0.08***	0.016	<0.001
Trust Reliable Sources (specific)	Sub-index	11,893	-0.04	0.023	0.08
Distrust Unreliable Sources (specific)	Sub-index	11,900	0.27***	0.023	<0.001
Specific Trust: Health Workers	Question	11,951	0.02	0.013	0.13
Specific Trust: Gov Pamphlets	Question	11,970	-0.05**	0.015	0.002
Specific Trust: AIIMS Doctors	Question	11,969	-0.04*	0.015	0.01
Specific Trust: Traditional Remedies	Question	11,968	0.19***	0.017	<0.001
Specific Trust: WhatsApp	Question	11,962	0.22***	0.018	<0.001
Specific Trust: Ayurvedic Doctors	Question	11,972	0.06***	0.015	<0.001
Cues: Reputable Source	Question	11,881	-0.08***	0.013	<0.001
Cues: Sensational Tone	Question	11,859	-0.09***	0.014	<0.001
Cues: Likes and Shares	Question	11,903	0.14***	0.014	<0.001
Cues: Same Community	Question	11,898	0.13***	0.013	<0.001
Cues	Sub-index	11,657	0.09***	0.017	<0.001

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

Table G13: Health Preferences

Outcome	Type	N	Estimate	SE	p-value
Index	—	11,310	0.21***	0.021	<0.001
Interest in Health News	Question	12,008	0.02	0.012	0.12
Perception of Vaccine Safety	Sub-index	11,472	0.07**	0.022	0.001
COVID Vaccine Safety	Question	11,944	0.02**	0.008	0.002
Chicken Pox Vaccine Safety	Question	11,514	0.03*	0.012	0.03
Distrust Non-Scientific Remedies	Sub-index	11,799	0.23***	0.024	<0.001
Not Use Traditional Remedies	Question	11,922	0.14***	0.013	<0.001
Not Go to Traditional Healer	Question	11,950	0.05***	0.016	<0.001
Ayurveda/Homeopathy Ineffective	Question	11,925	0.11***	0.017	<0.001

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

Table G14: Engagement Attitude

Outcome	Type	N	Estimate	SE	p-value
Index	—	11,869	0.11***	0.023	<0.001
Correct Friend	Question	11,948	0.00	0.012	0.79
Not Share Friend's Misinformation	Question	11,918	0.09***	0.018	<0.001
Fact Checking Important	Question	12,008	-0.01	0.008	0.25
Fact Checked Recently	Question	12,008	0.14***	0.034	<0.001

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

Table G15: Engagement Behavior

Outcome	Type	N	Estimate	SE	p-value
Index	—	12,008	0.02	0.020	0.38
Prefer Hindustan Newspaper	Question	12,008	0.01	0.008	0.34
Willing to Be a Truth Ambassador	Question	12,008	0.00	0.006	0.73

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

Table G16: Awareness

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**
Index	—	11,782	-0.01	0.024	0.64
Misinfo Threat Perception	Question	11,948	0.13***	0.031	<0.001
News Manipulation Awareness	Sub-index	11,910	-0.14***	0.022	<0.001
News Manipulation Awareness1	Question	11,959	-0.09***	0.015	<0.001
News Manipulation Awareness2	Question	11,948	-0.09***	0.018	<0.001
Confirmation Bias Awareness	Sub-index	11,906	-0.02	0.022	0.32
Confirmation Bias Awareness1	Question	11,956	0.00	0.021	0.86
Confirmation Bias Awareness2	Question	11,951	-0.03	0.019	0.14

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

H Heterogeneous treatment effects

Table H17 presents results from the following specification:

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

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 H17, we present the coefficients for the interaction term across multiple pre-registered pre-treatment covariates. These include individual characteristics such as age, grade and gender (measured using an indicator for male), 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 when faced with acute medical symptoms), 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 higher for male students, non-BJP affiliation is associated with a higher effect on awareness, and two outcomes are associated with a higher effect in the low-spillover stratum.

In Figure H13 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 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.

Table H17: 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	0.003	-0.008	-0.003	0.006
	(0.0135)	(0.0119)	(0.0114)	(0.0118)	(0.0116)	(0.0115)	(0.0108)
N	11122	11264	11310	11482	11869	12008	11782
T x Grade	0.012	0.026	0.020	0.004	-0.001	-0.008	0.011
	(0.0161)	(0.0153)	(0.0152)	(0.0137)	(0.0142)	(0.0148)	(0.0132)
N	11121	11263	11310	11481	11868	12007	11782
T x Male	-0.016	-0.010	-0.060	-0.035	0.135***	0.054	-0.024
	(0.0443)	(0.0410)	(0.0378)	(0.0415)	(0.0403)	(0.0393)	(0.0364)
N	11122	11264	11310	11482	11869	12008	11782
T x Non-scientific beliefs	-0.017	-0.010	-0.001	-0.000	0.015	0.015	0.004
	(0.0141)	(0.0139)	(0.0136)	(0.0131)	(0.0136)	(0.0128)	(0.0139)
N	11017	11155	11199	11370	11751	11888	11665
T x Prior media exposure	0.019	0.006	-0.015	0.013	0.034	0.012	0.002
	(0.0213)	(0.0205)	(0.0184)	(0.0192)	(0.0183)	(0.0195)	(0.0194)
N	11045	11189	11239	11406	11789	11928	11705
T x Asset Index	0.027	-0.022	0.013	-0.022	0.037	0.024	-0.013
	(0.0204)	(0.0202)	(0.0203)	(0.0205)	(0.0188)	(0.0186)	(0.0196)
N	11122	11264	11310	11482	11869	12008	11782
T x Party ID (non-BJP)	-0.029	-0.093	-0.010	0.073	0.020	-0.013	0.109*
	(0.0644)	(0.0675)	(0.0596)	(0.0588)	(0.0630)	(0.0574)	(0.0554)
N	6946	7040	7028	7157	7423	7524	7375
T x Low-spillover	0.013	-0.041	0.016	0.081	0.102*	0.035	0.123*
	(0.0470)	(0.0458)	(0.0423)	(0.0427)	(0.0473)	(0.0391)	(0.0479)
N	11122	11264	11310	11482	11869	12008	11782
Library-Spillover FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

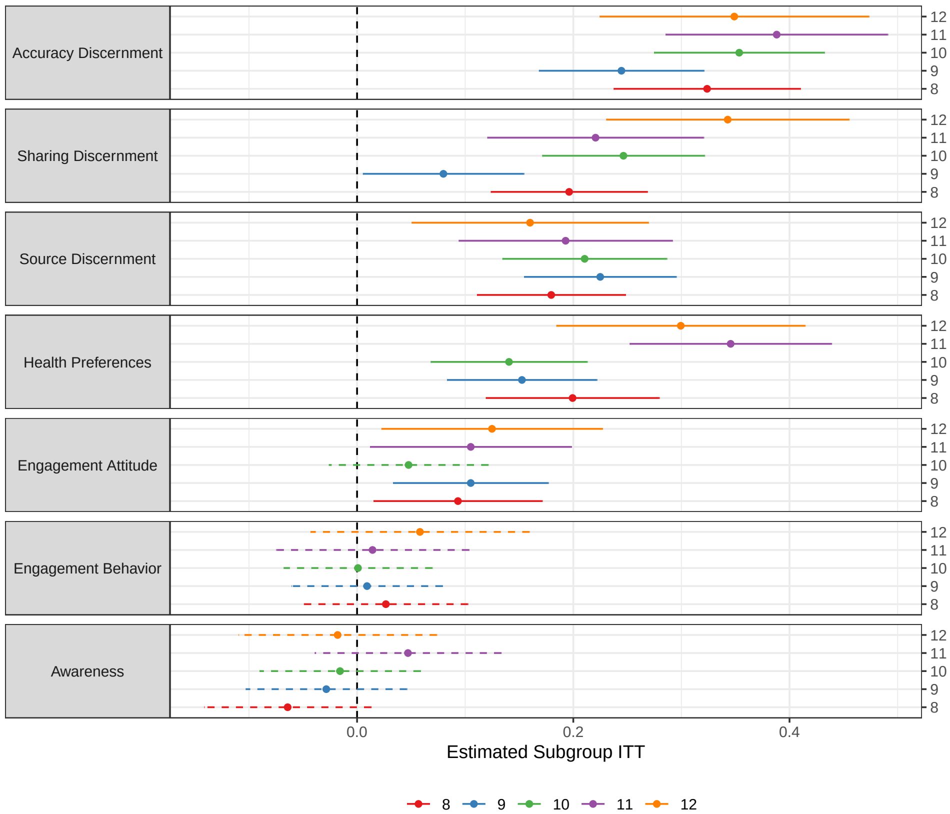
*p<0.05; **p<0.01; ***p<0.001. Models include library-spillover strata FEs.

These informal leaders were village residents and could have one of a number of specific 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

Figure H13: Subgroup ITT by Student Grade



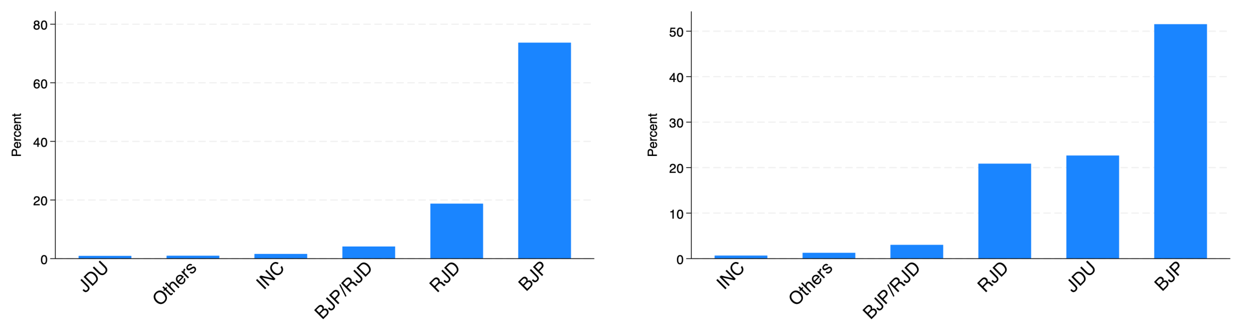
survey and the baseline dataset. Some observations were lost during this process due to incomplete or unstandardized *jati* names; this issue was more prevalent in the baseline data than in the elite survey.

- **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 H14 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 H14: Party Preference Distribution



(a) 2019 National Elections

(b) 2020 State Elections

Figure H15 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 H16 shows the distribution of preferences for the BJP and RJD (the major parties) across 100 libraries, with each bar representing a separate library.

Figure H15: Party Preference of Major Jatis

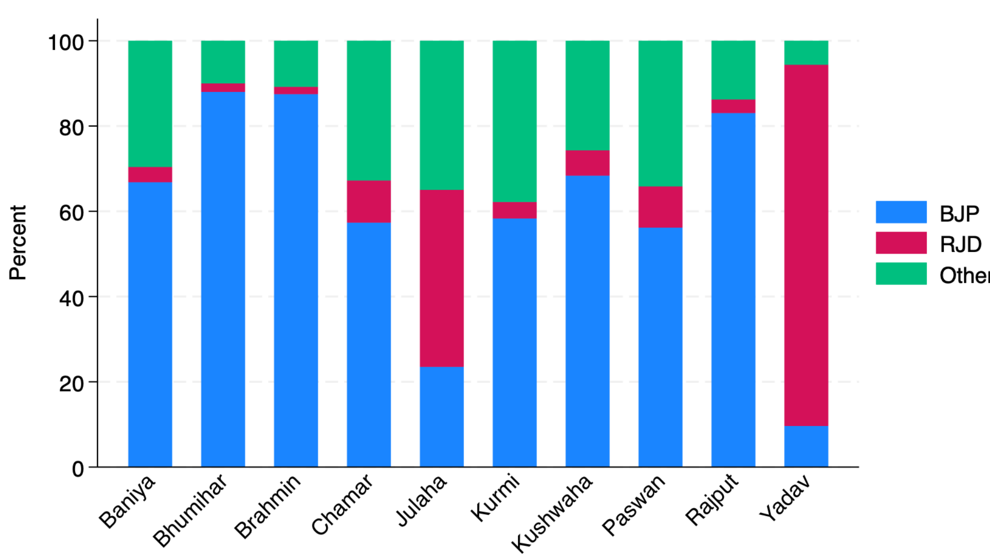
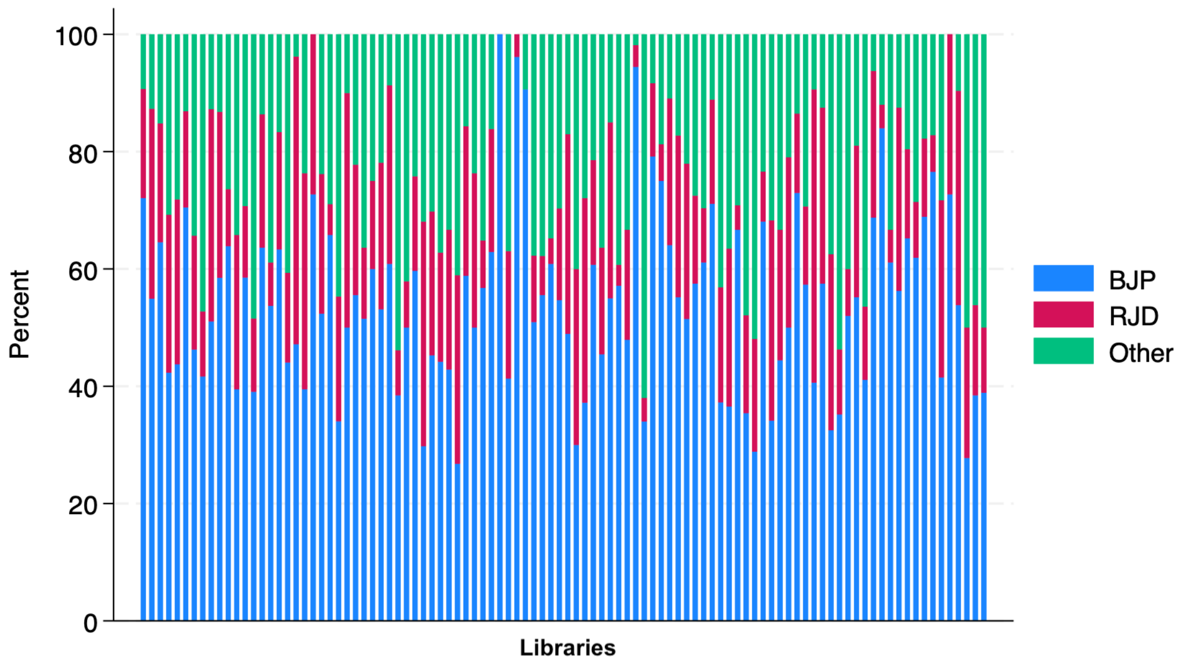


Figure H16: Party Preference by Library



H.2 Gender subgroup ITT

Table H18: Engagement Indices Means by Gender

	Group	Male (mean)	Female (mean)
Engagement Attitude	Treatment	0.200	0.041
Engagement Attitude	Control	0.014	−0.010
Engagement Behavior	Treatment	0.135	−0.030
Engagement Behavior	Control	0.070	−0.049

I Second endline

I.1 Sampling and attrition

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, 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 I19). There is some evidence that students with better effects on sharing discernment and engagement attitudes were more likely to be reached (see Table I21). 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 I20), 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 I19: Follow-Up Sample Characteristics

Characteristic	**Overall** *N = 13590*	**Called, part.** *N = 2059*	**Called, not part.** *N = 2874*	**Not called** *N = 8657*
Treatment				
Spoken English	6,774 (50%)	1,027 (50%)	1,422 (49%)	4,325 (50%)
Media Literacy	6,816 (50%)	1,032 (50%)	1,452 (51%)	4,332 (50%)
Gender				
Male	5,672 (42%)	802 (39%)	1,269 (44%)	3,601 (42%)
Female	7,918 (58%)	1,257 (61%)	1,605 (56%)	5,056 (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%

n (%); Mean (SD); %

Table I20: Follow-Up Attrition Predictors

Predictor	N	Estimate	SE	p-value
Treatment	4,933	-0.01	0.016	0.67
Gender - Female	4,933	0.05***	0.015	<0.001
Grade	4,932	0.01**	0.005	0.01
Age	4,933	0.00	0.005	0.33
Religion - Hindu	4,933	0.08**	0.028	0.004
Language - Hindi	4,933	0.03	0.017	0.06
Asset Index	4,933	0.01	0.008	0.08
Father's Education	4,688	0.00	0.002	0.06
Mother's Education	4,704	0.00**	0.002	0.003
Government School	4,933	0.02	0.040	0.6
Science Knowledge	4,933	0.01*	0.005	0.04
Mobile Internet	4,933	-0.02	0.018	0.18
Newspapers	4,933	-0.03	0.025	0.2
Social Media	4,933	0.00	0.014	0.75
TV	4,933	-0.04*	0.019	0.02
Friends and Family	4,933	0.00	0.043	>0.9
Vaccinated	4,933	-0.02	0.018	0.28
Ayurveda	4,933	-0.02	0.022	0.35
Awareness	4,274	0.00	0.008	0.64
Accuracy Discernment	4,031	0.01	0.007	0.4
Sharing Discernment	4,080	-0.01	0.008	0.29
Health Preferences	4,087	0.01	0.008	0.4
Source Discernment	4,158	0.01	0.008	0.26
Engagement Attitude	4,309	0.00	0.008	>0.9
Engagement Behavior	4,357	0.00	0.008	>0.9

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

Table I21: Differential Follow-Up Attrition Predictors

Predictor	N	Estimate	SE	p-value
Gender - Female * T	4,933	0.03	0.029	0.34
Grade * T	4,932	-0.01	0.011	0.2
Age * T	4,933	0.00	0.009	0.7
Religion - Hindu * T	4,933	0.03	0.058	0.61
Language - Hindi * T	4,933	0.01	0.030	0.72
Asset Index * T	4,933	0.00	0.015	0.85
Father's Education * T	4,688	0.00	0.003	0.12
Mother's Education * T	4,704	0.00	0.003	0.75
Government School * T	4,933	0.08	0.080	0.33
Science Knowledge * T	4,933	0.01	0.010	0.34
Mobile Internet * T	4,933	0.01	0.035	0.84
Newspapers * T	4,933	-0.01	0.050	0.84
Social Media * T	4,933	-0.05	0.027	0.08
TV * T	4,933	0.04	0.038	0.32
Friends and Family * T	4,933	0.08	0.085	0.37
Vaccinated * T	4,933	0.04	0.034	0.28
Ayurveda * T	4,933	-0.04	0.043	0.33
Awareness * T	4,274	-0.01	0.016	0.62
Accuracy Discernment * T	4,031	0.02	0.015	0.2
Sharing Discernment * T	4,080	0.04*	0.015	0.01
Health Preferences * T	4,087	0.02	0.015	0.25
Source Discernment * T	4,158	-0.01	0.015	0.45
Engagement Attitude * T	4,309	0.03*	0.015	0.03
Engagement Behavior * T	4,357	0.00	0.015	0.84

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

I.2 Results

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 I22). 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 I23.

Table I22: Parents' reasons for sending child to classes

	Reason for participation	%
1	Wanted child to learn	70.42
2	Trust Jeevika	26.62
3	Free nature of classes	1.51
4	Wanted child out of the house	1.08

Table I23: Respondent reactions to people sharing misinfo.

	Reaction	Control	Treatment	statistic	p-value
1	Emphasize that it is false	0.35	0.17	-11.42	< 0.001
2	Teach strategy to verify	0.22	0.30	4.87	< 0.001
3	Admonish for spreading false info	0.13	0.13	0.04	0.97
4	Emphasize not sharing	0.30	0.40	5.47	< 0.001

J Robustness checks

J.1 Complier average causal effects

In Table J24, 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 define compliance as a continuous measure from 0 to 1 measuring the number of sessions a student attended (measured with respondent-level attendance data gathered by teachers during each session).

- First stage (predicting treatment using the IV): Treatment uptake (measured as number of classes attended / 4) is regressed on the instrumental variable: treatment assignment. This step estimates the likelihood of receiving treatment based solely on assignment, capturing the behavior of compliers.
- Second stage (estimating the CACE): Predicted values from the first stage (the part of treatment explained by assignment) 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 all 4 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 the share of sessions 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 J24: Complier Average Causal Effects (CACE)

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.397***	0.260***	0.259***	0.269***	0.133***	0.023	-0.014
	(0.0252)	(0.0246)	(0.0243)	(0.0247)	(0.0231)	(0.0232)	(0.0227)
N	11122	11264	11310	11482	11869	12008	11782
R^2	0.160	0.100	0.089	0.105	0.146	0.079	0.146
Library-Spillover FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

J.2 Alternative specifications

In this appendix we present the main results using alternate specifications as robustness checks. First, Table J25 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 J25: Main Results with Library FE

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.314***	0.221***	0.205***	0.212***	0.103**	0.036	-0.008
	(0.0390)	(0.0325)	(0.0304)	(0.0328)	(0.0378)	(0.0296)	(0.0382)
N	11122	11264	11310	11482	11869	12008	11782
R^2	0.028	0.015	0.010	0.010	0.005	0.001	0.004
Library FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

Next, Tables J26 and J27 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.

Table J26: Main Results with District FE

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.312***	0.217***	0.204***	0.212***	0.104***	0.032	-0.008
	(0.0284)	(0.0276)	(0.0255)	(0.0265)	(0.0289)	(0.0247)	(0.0309)
N	11122	11264	11310	11482	11869	12008	11782
R^2	0.113	0.054	0.049	0.059	0.088	0.041	0.076
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

Lastly, Table J28 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 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 female students, an indicator for Hindi as the medium of education, grade, reading skill index, science skill index, prior

Table J27: 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***	0.206***	0.204***	0.105***	0.021	-0.009
	(0.0282)	(0.0271)	(0.0250)	(0.0260)	(0.0288)	(0.0237)	(0.0302)
N	11122	11264	11310	11482	11869	12008	11782
R^2	0.117	0.059	0.054	0.063	0.092	0.047	0.084
District-Spillover FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

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 J28: Main Results (Adjusted Model)

Outcome	Accuracy Discernment	Sharing Discernment	Health Preferences	Source Discernment	Engagement Attitude	Engagement Behavior	Awareness
Treatment	0.320***	0.212***	0.205***	0.211***	0.102***	0.001	-0.022
	(0.0237)	(0.0225)	(0.0215)	(0.0207)	(0.0238)	(0.0196)	(0.0249)
Age	-0.022*	-0.030**	-0.021*	-0.011	-0.012	0.005	-0.014
	(0.0107)	(0.0097)	(0.0100)	(0.0106)	(0.0098)	(0.0090)	(0.0094)
Female	-0.155***	-0.136***	0.013	-0.091***	-0.067**	-0.114***	-0.027
	(0.0236)	(0.0220)	(0.0198)	(0.0220)	(0.0212)	(0.0206)	(0.0196)
Grade	0.056***	0.054***	0.081***	0.067***	0.041***	0.016	0.060***
	(0.0126)	(0.0125)	(0.0133)	(0.0127)	(0.0121)	(0.0116)	(0.0119)
Caste – OBC	0.037	-0.046	-0.079	-0.062	0.004	0.001	-0.042
	(0.0432)	(0.0369)	(0.0406)	(0.0432)	(0.0365)	(0.0362)	(0.0411)
Caste – SC	-0.055	-0.116**	-0.097*	-0.099*	-0.010	0.037	-0.085
	(0.0489)	(0.0433)	(0.0459)	(0.0494)	(0.0422)	(0.0419)	(0.0469)
Caste – ST	-0.127	-0.136	-0.241*	-0.092	-0.091	0.075	-0.194*
	(0.0956)	(0.0776)	(0.0931)	(0.0718)	(0.0894)	(0.0764)	(0.0754)
Hindu	0.002	-0.002	0.068	0.047	-0.012	-0.063	0.007
	(0.0431)	(0.0411)	(0.0408)	(0.0389)	(0.0387)	(0.0376)	(0.0388)
Asset Index	0.085***	0.084***	0.023*	0.026*	0.027**	0.010	0.014
	(0.0113)	(0.0114)	(0.0112)	(0.0118)	(0.0102)	(0.0099)	(0.0105)
Hindi medium	-0.098	-0.117	-0.101	-0.146*	-0.114*	-0.124	-0.076
	(0.0539)	(0.0672)	(0.0648)	(0.0621)	(0.0483)	(0.0634)	(0.0477)
Reading skill index	0.109***	0.133***	0.100***	0.026	0.051**	-0.008	0.010
	(0.0218)	(0.0203)	(0.0217)	(0.0202)	(0.0185)	(0.0181)	(0.0194)
Science skill index	0.062***	0.053***	0.040***	0.059***	0.048***	0.012	0.031***
	(0.0081)	(0.0079)	(0.0083)	(0.0079)	(0.0074)	(0.0071)	(0.0077)
Mobile Internet	0.038	0.023	0.050	0.081**	0.066**	0.109***	0.035
	(0.0298)	(0.0271)	(0.0264)	(0.0285)	(0.0239)	(0.0260)	(0.0263)
BJP+ Vote Share	-0.004	-0.007	0.001	-0.006	-0.005	0.000	0.003
	(0.0050)	(0.0048)	(0.0039)	(0.0043)	(0.0041)	(0.0054)	(0.0047)
Village Nightlights	0.033	0.023	0.104***	0.038	-0.024	-0.016	0.016
	(0.0342)	(0.0354)	(0.0284)	(0.0360)	(0.0302)	(0.0256)	(0.0333)
N	10130	10266	10295	10453	10798	10919	10714
R^2	0.187	0.129	0.106	0.124	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

J.3 Multiple hypotheses corrections

To assuage concerns about multiple comparisons, Tables J29 and J30 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. The FDR adjustment controls the expected proportion of false positives among all rejected hypotheses, ensuring a more balanced approach to identifying significant results while maintaining statistical rigor. Bonferroni corrections, on the other hand, are a more conservative method that adjusts p-values to account for the total number of comparisons, reducing the likelihood of type I errors by making it more difficult to declare statistical significance. 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 J29: Main Results (ITT): Correcting for Multiple Hypotheses

Outcome	**N**	**Estimate**	**SE**	**p**	**p (FDR)**	**p (Bonf.)**
Awareness	11,782	-0.01	0.024	0.64	0.64	>0.9
Accuracy Discernment	11,122	0.32	0.024	<0.001	<0.001	<0.001
Sharing Discernment	11,264	0.21	0.023	<0.001	<0.001	<0.001
Health Preferences	11,310	0.21	0.021	<0.001	<0.001	<0.001
Source Discernment	11,482	0.21	0.021	<0.001	<0.001	<0.001
Engagement Attitude	11,869	0.11	0.023	<0.001	<0.001	<0.001
Engagement Behavior	12,008	0.02	0.020	0.38	0.44	>0.9

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

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

J.4 District subgroup ITT

To determine whether the effects are concentrated in specific districts, we examine ITT estimates by district subgroup. The results reveal no clear pattern. Notably, given significant variation in districts' socio-economic development, we confirm that these findings are not correlated with development outcomes (measured by village night lights data) as well as political outcomes (measured by BJP vote share at the AC level in 2020).

Table J30: Main Results (CACE): Correcting for Multiple Hypotheses

Outcome	**N**	**Estimate**	**SE**	**p**	**p (FDR)**	**p (Bonf.)**
Awareness	11,782	-0.01	0.023	0.54	0.54	>0.9
Accuracy Discernment	11,122	0.40	0.025	<0.001	<0.001	<0.001
Sharing Discernment	11,264	0.26	0.025	<0.001	<0.001	<0.001
Health Preferences	11,310	0.26	0.024	<0.001	<0.001	<0.001
Source Discernment	11,482	0.27	0.025	<0.001	<0.001	<0.001
Engagement Attitude	11,869	0.13	0.023	<0.001	<0.001	<0.001
Engagement Behavior	12,008	0.02	0.023	0.33	0.39	>0.9

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

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

Figure J17: District ITT and village nightlights

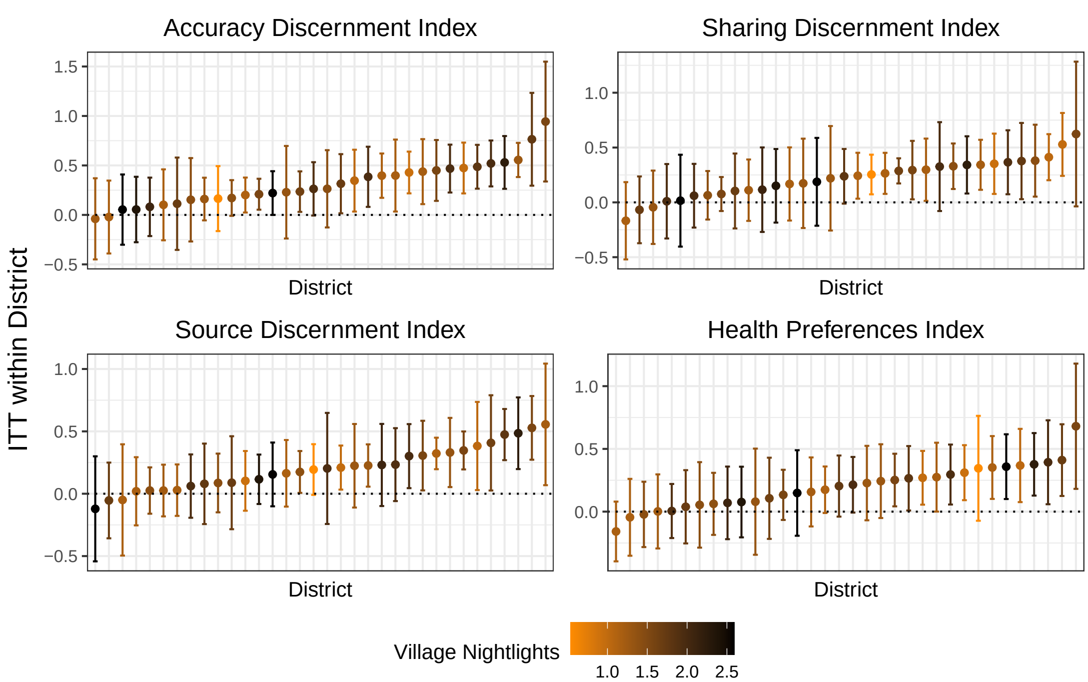
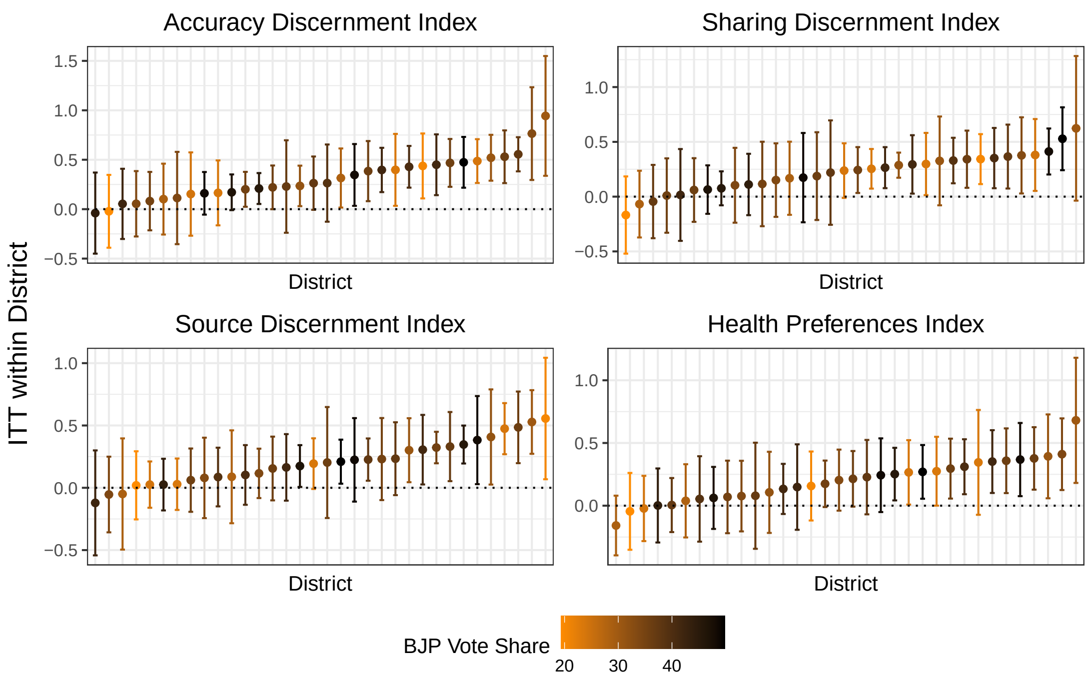


Figure J18: District ITT and BJP vote share



J.5 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 C6 suggests that outcome data is *not* missing completely at random (MCAR), Table C7 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.

Since our outcome indices have wide support and a non-negligible amount of attrition (around 11.6%), using Manski bounds (Manski, 1990) to bound our effects would be overly conservative and, hence, uninformative. Furthermore, the clustered nature of our data further limits our ability to conduct standard sensitivity analyses without changing our estimation methods.

To address potential bias from missing data, we instead 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 J31 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 1.43 control-group SDs, implying that the treated potential outcome for missing individuals in the treatment group would have to be 1.43 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 1.68 SDs/1.93 SDs for the estimate to dip below zero/become negative and significant, respectively.

Table J31: Tipping Point Analysis Results

	Tipping Points (in SD units)		
Outcome	Positive but Not Sign.	Estimate below 0	Negative and Sign.
Accuracy Discernment	1.43	1.68	1.93
Sharing Discernment	1	1.23	1.47
Source Discernment	1.02	1.27	1.53
Health Preferences	1.01	1.25	1.5
Engagement Attitude	0.41	0.73	1.07
Engagement Behavior	–	0.15	0.48
Awareness	–	–	0.22

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.

J.6 Alternative discernment index

Table J32: Discernment (excluding 2 items)

Outcome	**Type**	**N**	**Estimate**	**SE**	**p-value**
Accuracy Discernment	Index	11,150	0.27***	0.024	<0.001
Sharing Discernment	Index	11,313	0.18***	0.022	<0.001

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

K Deviations from PAP

We report two 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 stratum 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 PAP rather than formal hypotheses. The additional variables we considered 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. 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.

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. 2024. “The Religious Roots of Belief in Misinformation: Experimental Evidence from India.” *Working paper.*

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.

Manski, Charles F. 1990. “Nonparametric Bounds on Treatment Effects.” *The American Economic Review* 80 (2): 319–323.

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.