American Political Science Review (2025) 1–21
doi:10.1017/S0003055425101184

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

PRIYADARSHI AMAR *University Carlos III Madrid & Instituto Carlos 3 - Juan March, Spain*
SUMITRA BADRINATHAN *American University, United States*
SIMON CHAUCHARD *University Carlos III Madrid & Instituto Carlos 3 - Juan March, Spain*
FLORIAN SICHART *Princeton University, United States*

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

INTRODUCTION

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

A substantial body of research has evaluated the effectiveness of misinformation countermeasures, including fact-checking and corrections (Bowles et al. 2025; Clayton et al. 2019; Porter and Wood 2019), accuracy prompts (Pennycook and Rand 2019), inoculation (Pereira et al. 2024; Roozenbeek et al. 2022), and tip-based information (Guess et al. 2020). While many of these strategies show promise, they are typically one-off, online interventions targeting digitally literate, urban populations and are rarely adapted for offline communities (Blair et al. 2024). In parallel, governments and non-governmental organizations have increasingly turned to classroom-based media and information literacy (MIL) programs aimed at the youth, with a global uptick in such

Priyadarshi Amar, Postdoctoral Research Fellow, Department of Social Sciences, University Carlos III Madrid & Instituto Carlos 3 - Juan March, Spain, priyadarshi.amar@uc3m.es.
Corresponding author: Sumitra Badrinathan, Assistant Professor, Department of Politics, Governance, and Economics, American University, United States. sumitrab@american.edu.
Simon Chauchard, Associate Professor, Department of Social Sciences, University Carlos III Madrid & Instituto Carlos 3 - Juan March, Spain, simon.chauchard@uc3m.es.
Florian Sichart, PhD Candidate, Department of Politics, Princeton University, United States, fsichart@princeton.edu.

Handling editor: Isabela Mares.

Received: December 23, 2024; revised: June 07, 2025; accepted: August 25, 2025.

initiatives since 2016. For example, New Jersey has advanced mandatory K–12 media literacy education (Sitrin 2020), echoing efforts in California, Estonia, and Finland. Theoretically, these programs share important design features: they emphasize repetition, peer-based learning, and delivery by authority figures who can help shape norms. Yet, despite the growing adoption and comparatively high cost of such programs, there is a striking lack of causal evidence assessing their impact. To date, no study has rigorously evaluated the effects of sustained, classroom-based interventions on misinformation outcomes.[1] This reveals a significant disconnect between the academic literature on misinformation interventions and the types of programs being implemented in the real world.

To fill this gap, we conducted a field experiment in 583 villages in Bihar, one of India's least developed states, involving over 13,500 adolescents aged 13–18. India, where misinformation has led to health crises and to political violence (Badrinathan, Chauchard, and Siddiqui 2025; Siddiqui 2020), is a critical case for understanding how falsehoods spread and persist: it represents a combination of low state capacity, shrinking independent media, and elite-driven disinformation in a context of increasing polarization, making misinformation an issue as crucial as it is challenging to address. Our intervention targeted students in grades 8 through 12 and consisted of classroom-based sessions on misinformation. Over a 14-week period, students participated in four 90-minute sessions, held approximately every 3 weeks, with homework assignments between sessions. The curriculum, designed specifically for this context but building on principles of MIL initiatives across the world, focused on health misinformation and aimed to (1) enhance scientific knowledge about health and counter health-related misinformation, (2) equip students with broad critical skills and practical tools to encourage a more responsible consumption of information, and (3) shift norms surrounding misinformation.

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

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

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

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

This study has significant implications not only for the literature on countering misinformation but also for the creation of education policy and public health strategies, and for work on behavioral change in

[1] A partial exception is Apuke, Omar, and Tunca (2023), who report positive effects from a 6-week media literacy course in Nigeria, though concerns about sample size, spillovers, and compliance limit its internal validity.

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

developing countries. Its findings contribute to several academic literatures: to work in American politics advancing knowledge on information and persuasion broadly (Coppock 2023; Huber and Arceneaux 2007); to experimental methods, focusing on theory and practical strategies for communicating scientific ideas (Alsan and Eichmeyer 2024; Andrews and Shapiro 2021); to comparative politics, especially research examining how public infrastructure can strengthen democratic outcomes (Boas and Hidalgo 2011; Gottlieb, Adida, and Moussa 2022; Green et al. 2024); and finally to work focusing on politics in South Asia, exploring effective informational and behavioral interventions to enhance governance and societal outcomes (Banerjee et al. 2014; Cheema et al. 2023; Dhar, Jain, and Jayachandran 2022; Ghosh et al. 2025).

SUSTAINED CLASSROOM EDUCATION AGAINST MISINFORMATION

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

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

Our study addresses this gap. The intervention we design and evaluate in this article is called BIMLI, the *Bihar Information and Media Literacy Initiative*. A multiple-session program held over several months and focused on equipping students with tools to recognize and resist misinformation, BIMLI evaluates, by design, a fundamentally different model from those examined in the existing academic literature: a classroom-based intervention which actively mimics the initiatives policymakers are implementing across the world. We highlight a few design features of the intervention. First, in terms of mode of delivery, we administered the program face-to-face, fostering a peer-based, interactive environment, where respondents encountered key lessons repeatedly over multiple sessions delivered by an instructor. Research suggests that peer interactions in classroom settings can deepen understanding by exposing learners to

TABLE 1. Examples of Media Literacy Interventions

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

diverse perspectives (Dhar, Jain, and Jayachandran 2022), while repeated exposure allows for reinforcement of concepts (Fazio, Rand, and Pennycook 2019), and authority figures promote norm-building (Tankard and Paluck 2016). Second, to mimic the governmental and international organization support for such initiatives across the world, we secured partnership with an agency of the Bihar state government to roll out the program as an official government-offered course.

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

Media literacy treatments in the academic literature have a mixed record, resulting in either weak or null findings (Blair et al. 2024). Studies in the Global North report modest positive effects, but every media literacy intervention conducted in the Global South to date has produced null results, leading some to question whether the optimism around media literacy is warranted (Blair et al. 2024). However, these null results may stem from an insufficient adaptation of experimental designs to local constraints and information environments, or from insufficient attention to best practices around how children learn and internalize information. Taking stock of this, we designed a treatment to mimic how education programs against misinformation are actually delivered on the ground: face-to-face, long-term, and integrated into existing school structures. In a context like India where information sharing is predominantly offline (Gadjanova, Lynch, and Saibu 2022), this delivery model is not only practical but also necessary.

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

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

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

THE POLITICS OF MISINFORMATION IN INDIA

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

In India, such deeply entrenched beliefs are closely tied to social identities and are often exploited by political elites to gain electoral support. Traditional health remedies, many rooted in ancient Hindu culture, have been used to appeal to Hindu voters—particularly under the Hindu nationalist Bharatiya Janata Party (BJP), which currently leads the federal government and portrays itself as a defender of Hindu values (Jaffrelot 2021). One striking example involved a member of parliament hosting a public event promoting cow urine as a COVID-19 cure, which resulted in several hospitalizations (Siddiqui 2020). Research shows that misinformation tied to long-standing identities is especially resistant to correction (Nyhan 2021), and India's enduring Hindu–Muslim cleavages make

religious identity a particularly potent factor in belief formation (Brass 2011; Chauchard and Badrinathan 2025). When elites deliberately reinforce falsehoods to polarize, such misinformed beliefs can be especially persistent: evidence from India suggests that motivated reasoning can impede correction efforts (Badrinathan 2021; Taber and Lodge 2006). Bihar, our study site, is part of a larger northern Indian media ecosystem including neighboring states like Uttar Pradesh, where elite-driven disinformation has sometimes resulted in violence (Badrinathan, Chauchard, and Siddiqui 2025).

For citizens in such contexts, finding ways out of the misinformation trap can be challenging. This is particularly true in Bihar, India's poorest state and home to 127 million people (as of 2023), where one-third live below the poverty line. The state's relative underdevelopment translates into a lack of essential services such as healthcare and education, alongside the failure of many public programs (Sharma 2015). The population we study faces profound structural barriers to learning. Children in Bihar, especially girls, are significantly less likely to attend school compared to those in other states (Muralidharan and Prakash 2017). Students often work for wages instead of attending school, teacher absenteeism is common, infrastructure is lacking (many classrooms lack electricity, seating, or basic materials), and learning suffers: only about half of Indian children enrolled in grade five can read a simple paragraph at the second-grade level (50.1% of children), or solve a two-digit subtraction problem (52.3% of children).[3] These alarming statistics have opened a serious debate on "what works" to improve learning in India, sparking a robust literature on education-based randomized controlled trials (RCTs) to which we contribute (de Barros et al. 2022).

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

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

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

EXPERIMENTAL DESIGN AND DATA COLLECTION

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

The Treatment

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

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

The curriculum emphasized interactive instruction, encouraging engagement between teachers and students as well as among students themselves, approaches notably lacking in many Indian classrooms where rote memorization and passive instruction dominate (Bhattacharya 2022; Kumar 1986). This approach aimed to cultivate analytical thinking

TABLE 2. The BIMLI Curriculum

	Learning objectives and content	Theoretical references	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's (2020) definition of misinformation and several journalistic examples of recent misinformation and its effects.	Examples and illustrations all local: health misinformation leading to vaccine hesitancy in India, falsehoods and doctored images on Indian WhatsApp groups.
Session 2: *"Understanding biases and critical thinking"*	*Objectives*: (1) Develop critical thinking skills. (2) Develop awareness of human biases in info. consumption. (3) Develop awareness of media biases in the production of information. *Modules*: (1) Recap of session 1. (2) Intro to human psychological biases like confirmation bias. (3) News and media system biases. (4) Critical thinking—definitions and strategies to enhance one's critical thinking.	Motivated reasoning from Taber and Lodge (2006), list of psychological biases adapted from Roozenbeek et al. (2022), list of media biases adapted from Ashley, Maksl, and Craft (2013), news framing effects from Druckman and Nelson (2003), fact-opinion discernment from Graham and Yair (2025).	Introduction to the news media environment in India + how biases manifest in the Indian context (e.g., 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 and verify information. (4) How to decide whether to share information.	Concrete examples of tips to spot misinformation (Badrinathan 2021; Guess, Nagler, and Tucker 2019; Vraga, Bode, and Tully 2022), focus on sharing as different from belief (Brashier and Schacter 2020).	Tailored Indian examples focused on WhatsApp such as reverse image search, looking for the "forwarded many times" tag, introduction to Indian fact-checking websites.
Session 4: *"Vaccine importance and talking about misinformation with family"*	*Objectives*: (1) Highlight importance of correcting/combating misinformation. (2) Develop strategies to deal with friends and relatives who spread misinformation. *Modules*: (1) General recap of lessons 1–3. (2) Strategies to fight misinformation at home. (3) Role play and memory games.	Efficacy of social corrections (Badrinathan and Chauchard 2023; Bode and Vraga 2018), 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 deep learning rather than relying solely on passive reception of information, representing a significant departure from the traditional structure of schooling in India (Kumar 1986). For instance, in Session 4 the lesson plan incorporated role-playing exercises in the classroom. In one activity, a student took on the role of a child while another acted as a parent, with the child tasked with employing strategies to engage with a parent that shared misinformation at a family dinner. The scenario aimed to highlight the challenges of addressing health misinformation with adults, particularly when such discussions involve confronting deeply ingrained beliefs in settings where confrontation with adults is discouraged (Malhotra and Pearce 2022).

Finally, a key instructional goal of this program was to focus on fostering critical thinking rather than offering prescriptive tips to spot misinformation. This approach was particularly suited to the Indian context, where much information is shared through friends and family, making source-specific advice (e.g., favoring one TV channel over another) ineffective. Given the decline in mass media credibility amid democratic

backsliding (Mohan 2021), we also avoided endorsing specific media outlets. Instead, we emphasized cues to critically assess information, such as recognizing emotional tone, not relying on shared ethnic identities as a cue to assess information, and identifying appropriate authorities as credible sources for specific topics—for instance, relying on community health workers employed by the government (called ASHA workers) for health-related information. Substantively, the curriculum relied on examples related to health misinformation.

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

Administering Classes

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

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

Recognizing that program success would depend not only on student compliance but also on teacher attendance and quality, we recruited a separate pool of teachers rather than using existing public school staff. Meetings with government officials revealed that public school teachers in Bihar are often overburdened, with high rates of absenteeism among both teachers and students, making compliance difficult. Each recruited teacher visited a given classroom approximately once every 2–3 weeks.[7] The curriculum was designed to be taught fully offline, using face-to-face discussion, printed materials, and minimal digital tools—mirroring the typical learning environment of rural schoolchildren in India.[8]

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

Sampling, Enrollment, and Baseline Data

Figure 1 outlines the timeline and flow of recruitment and roll-out of the study. We sampled $\approx$ 6 villages within a 3 km distance from each of the 100 libraries and

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

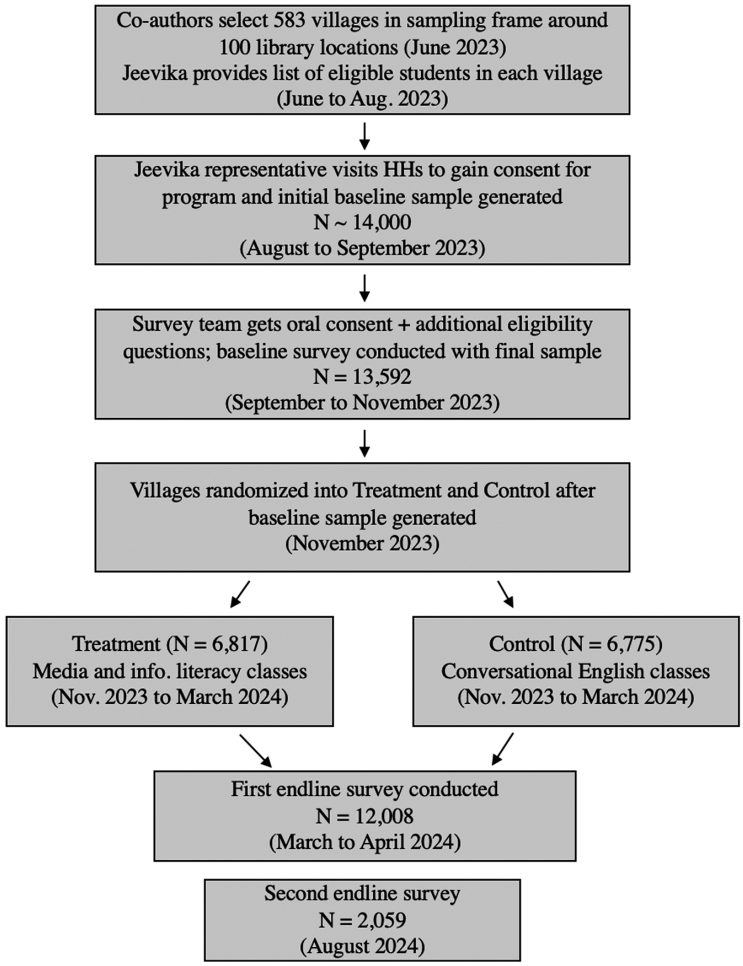
[5] These 100 libraries were located in 100 distinct blocks across the 32 districts.

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

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

[8] See Supplementary Section B1 for a classroom session example.

FIGURE 1. Study Flow and Timeline



randomized roughly 50% of these to receive our treatment; the remaining served as control villages. Our sampling procedures resulted in the selection of 583 villages across the state of Bihar. Before randomization, we categorized each village as having either high or low spillover potential, based on how many sampled villages fell within the same Gram Panchayat (GP)—a local administrative unit encompassing multiple villages. Spillovers were expected to be higher within the same GP, as children from those villages are more likely to attend the same schools or classes. Therefore, GPs with multiple selected villages were classified as high-spillover, while those with only one selected village were considered low-spillover. We then randomly assigned to treatment and control within each library area and spillover stratum (see Supplementary Section A for details).

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

The baseline survey collected demographic, household, and attitudinal data, including items on perceptions of the state, media usage, views on vaccines and traditional medicinal practices, science and reading skills, and social ties. Our baseline sample included

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

13,592 respondents across 583 villages, with 49.9% assigned to treatment and 50.1% to control.[10] In Supplementary Section A, we show balance tables confirming that respondents in treatment and control groups were balanced on key demographics, attitudes, and behaviors. The Supplementary Material also shows that treatment and control villages themselves were balanced on key variables based on census parameters.

Control Condition

Control group units participated in four modules of conversational English language classes, serving as a placebo rather than a pure control. This was done to achieve parity in effort exerted by students, since school attendance is a major problem in Bihar, and since our intervention lasted up to 4 months. We aimed to create comparable classroom dynamics and peer interactions, varying only the content of instruction. We additionally wanted control respondents to benefit from the program and hence chose a topic that fostered engagement without being related to misinformation outcomes. Subjects like math, science, and history were excluded due to overlap with standard curricula or national identity narratives, and nonacademic topics like cooking were discarded due to expected gender biases in their uptake. 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 Supplementary Section B). Topics avoided media, technology, and politics, and the very basic instruction level was unlikely to enable control students to independently navigate new information sources.[11]

Endline Data and Compliance

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

To boost compliance, we implemented a detailed monitoring system. Jeevika staff, women known locally as *didis*, regularly reminded households about upcoming classes. Students were motivated by the promise of a government-issued certificate upon completing the program. External monitors also made random visits to verify teacher presence and adherence to class schedule. Coauthors 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. We detect no significant difference in attendance numbers across treatment and control, with similar proportions attending both sets of classes. However, we do see a significant drop off in attendance for control group respondents during session three, though we note that the difference is substantively small (67% in control group and 74% in treatment) and dissipates during session 4 (see Figure 2). Further, we find that girls were more likely to attend classes compared to boys.[13] Importantly, since we estimate the ITT, lack of differential attrition by condition (Supplementary Section K) is more crucial for the internal validity of our estimates than the minor differential compliance we detect (Supplementary Section C).

Outcome Measures

We hypothesized that the intervention would influence a range of misinformation-related attitudes and behaviors.

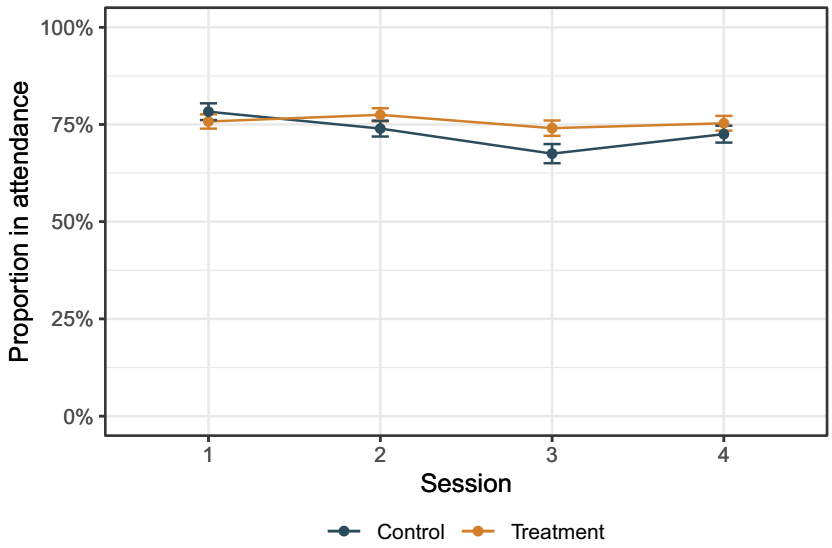
[10] The sample was 58% girls, with respondents ranging from grades 8 to 12 (median grade 10), and 96% enrolled in government schools. It was 91% Hindu and 69% OBC, on par with state census demographics. Language diversity included 43% Hindi-speaking households, 30% Bhojpuri, and 9% Magahi. Fathers' median education was grades 6–9, and mothers' median education was grades 1–5. Socioeconomic 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.

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

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

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

FIGURE 2. Compliance Data Across Treatment and Control



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

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

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:

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

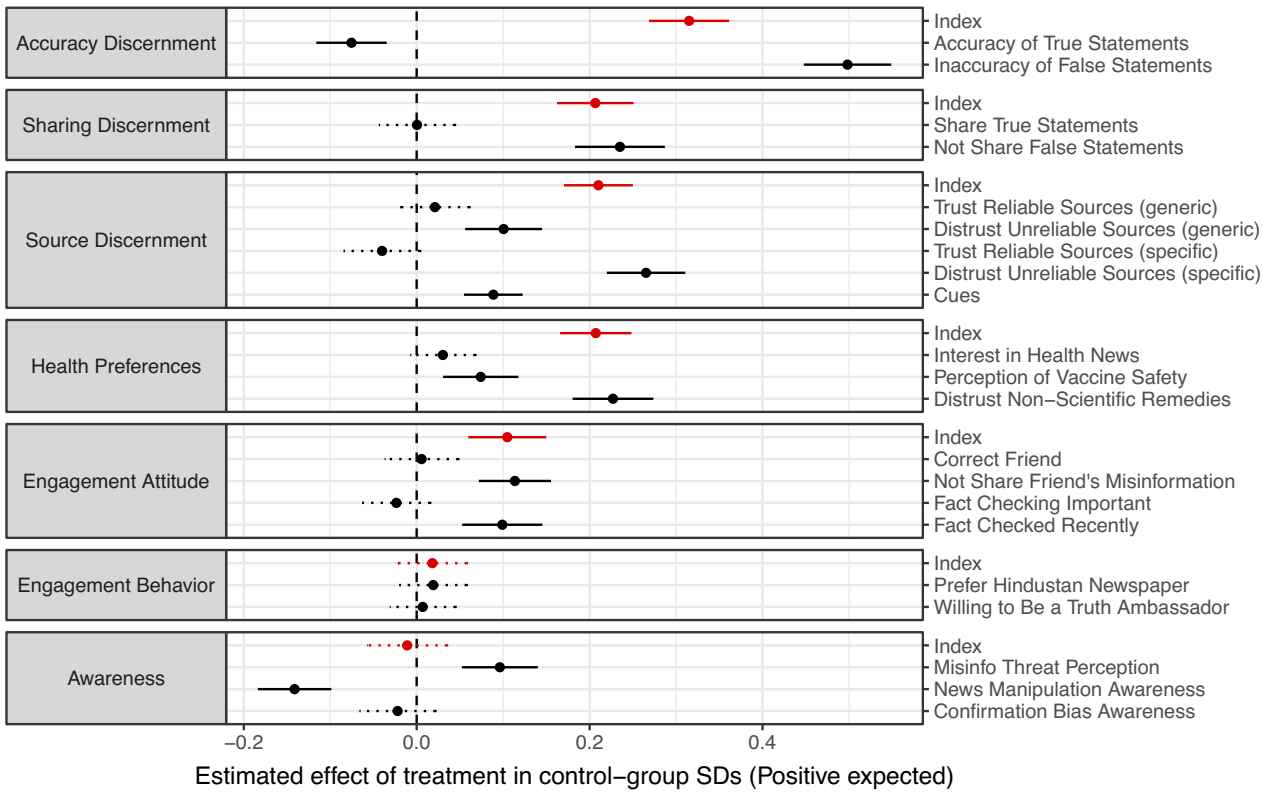
$$Y_{ijk} = \beta_0 + \beta_1 T_{ijk} + \sum_{c} \alpha_c X_{ci} + \sum_{k=1}^{m-1} \gamma_k + \varepsilon_{ijk}, \quad (2)$$

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

[14] Our pre-analysis plan was posted to OSF before endline data collection in February 2024 and is available at: https://osf.io/h43qn.

[15] See Supplementary Section J for CACE specification.

FIGURE 3. Estimated Effect of Assignment to BIMLI Treatment



Note: This figure plots the estimated ITT effect of assignment to BIMLI for seven 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 Supplementary Section G.

First Endline

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

Accuracy and Sharing Discernment

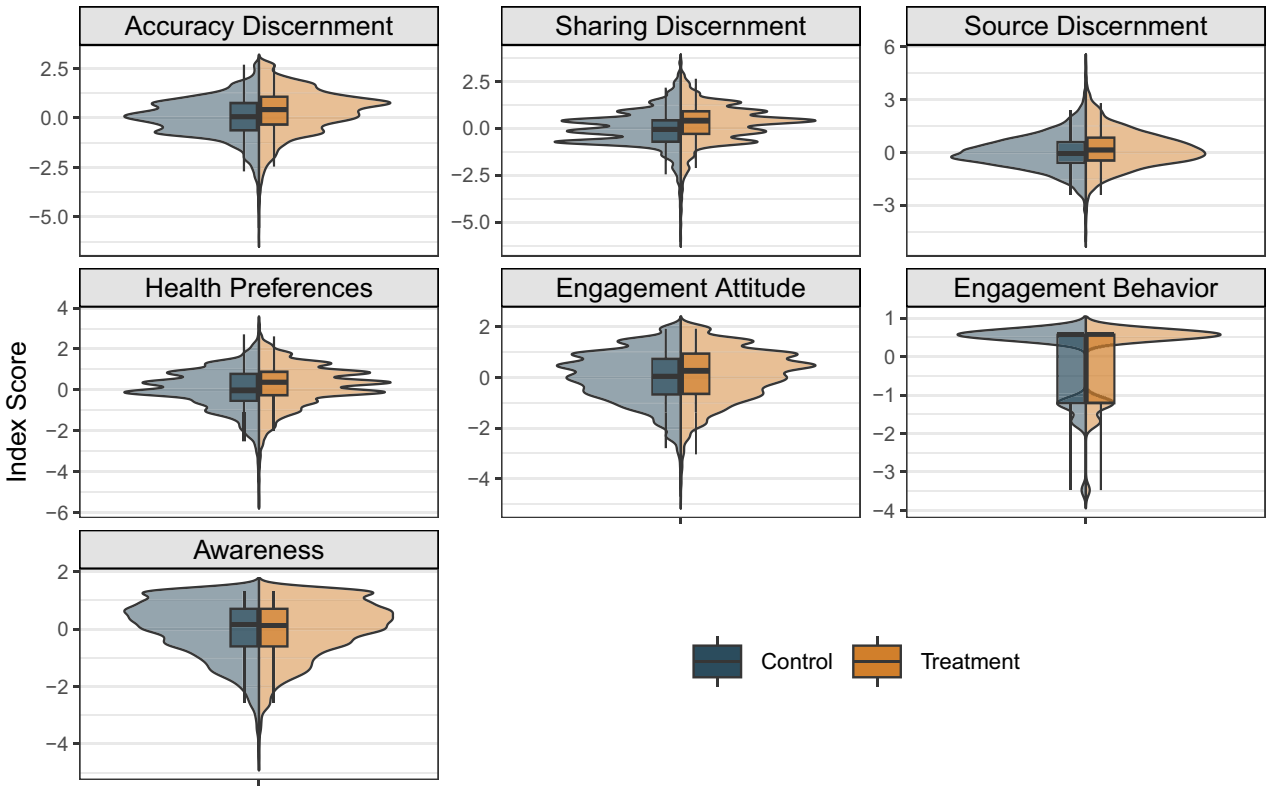
Recent years have seen a growing consensus on testing the efficacy of misinformation interventions through measuring discernment between true and false information. This approach involves (1) rating a mix of true and false content and (2) analyzing ability to discern between them (Guay et al. 2023). Following this, we asked respondents to rate the perceived accuracy of 8 veracity-balanced news stories on a four-point scale. Importantly, only two of these stories were discussed in class, while six were new, meaning that any discernment effects we detect reflect skill application rather than mere recall.[16] We also measured sharing intention using the same items.[17] The selection for these stories was based on extensive fieldwork and piloting to identify the most commonly believed health-related myths, each debunked by at least one fact-checking service in India. Stories were presented to respondents in random order.

With respect to accuracy discernment, ITT estimates show that the treatment significantly helped respondents discern between true and false stories (Figure 3). The magnitude of this effect, a 0.32 SD increase in discernment relative to the control group, is substantively large compared to effects from

[16] We reestimate effects dropping the two items discussed in class and find that results hold (Supplementary Section J4).

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

FIGURE 4. Distribution of Outcome Indices, by Treatment Group



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

comparable contexts.[18] Further, when we compare ITT to CACE estimates, we find that the effect on accuracy discernment is even larger among compliers (see Supplementary Section J).

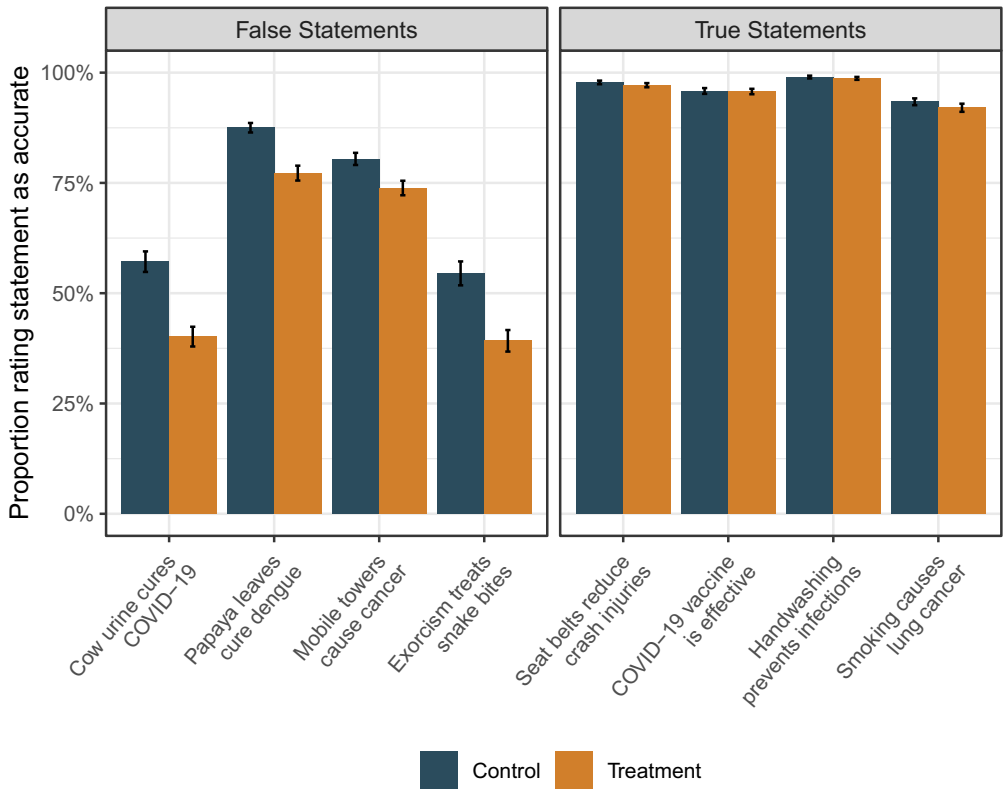
We do see variation in the true and false components of the discernment measure. In Figure 5, we graph perceived accuracy by individual news stories and find that large proportions of respondents believed falsehoods, and the treatment significantly decreased respondents' perceived accuracy of all four false stories, with effect sizes ranging from 0.44 SD (cow urine can cure COVID-19) to 0.18 SD (mobile phone towers cause cancer). With respect to true stories, there is little variation in how treatment and control group respondents rated these stories; on average all respondents were better at discerning true stories relative to false.

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

Empirically, we note that the apparent increase in skepticism on true items should be interpreted with caution: as Figure 5 shows, belief in true statements was already near ceiling at baseline. This limited variance

[18] For example, Guess et al. (2020) find that their digital literacy intervention in India led to a 0.11 SD increase in discernment, while Gottlieb, Adida, and Moussa's (2022) intervention in Côte d'Ivoire produced effect sizes of 0.12 to 0.15 SD. We note that larger and more comparable effect sizes to ours tend to emerge from interventions that are more intensive and of longer duration (see, for example, Bowles et al. [2025] where the intervention was 6 months long) lending further support to our argument that sustained exposure and iterative learning are more effective in shifting outcomes.

FIGURE 5. Accuracy Discernment by News Stories



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

inflates the standardized effect size, giving the impression of a stronger change than is actually the case. In absolute terms, the decrease is small. Finally, our second endline yields null effects on discernment measures for true information. This suggests that while the intervention's positive effect on reducing belief in false information persists over time, the temporary decrease in belief in true information is no longer detectable in the follow-up.

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

Trust in Sources and Source Discernment

To complement accuracy discernment, we introduced measures to evaluate how respondents assess and trust news sources. Recognizing that individuals rarely encounter headlines without accompanying source cues, we incorporated three measures focusing on news sources, including both mediums of news (e.g., platforms and 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 and Facebook), and institutions (e.g., the WHO). The index includes three sources we expect to increase trust in (MBBS doctors, healthcare workers, and government health notices) and three we hope to decrease trust in (ayurvedic doctors, unqualified practitioners, and word of mouth/rumors). Next, we assess situation-specific trust by using a vignette where respondents seek emergency advice for a sick family member and could go to a number of sources. We provide three trustworthy sources (community health center, government materials, and TV doctors) and three untrustworthy ones (family myths, WhatsApp forwards, and TV interviews with ayurvedic doctors). This helps distinguish between general and situation-specific trust and separates transmitters from mediums. Finally, we explore which factors foster trust in specific pieces of information,

examining whether reliance on signals like likes/shares online, shared ethnicity, as well as message tone and emotionality reduces due to the treatment. Our results show that BIMLI, overall, significantly changed how respondents interact with and trust sources for the better, with a notable shift in the index (SD = 0.21).

Health Preferences

We measured health preferences through three components: interest in health news, vaccine safety perceptions, and reliance on alternative medicine. Respondents rated their interest in health news on a scale from very interested to not interested. For vaccine safety, they rated the safety of both the COVID-19 and measles vaccines. To assess reliance on alternative medicine, respondents were asked if they would visit traditional healers and unqualified practitioners, or use home remedies for serious illnesses, and whether they agreed that ayurveda and homeopathy could cure serious diseases.

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

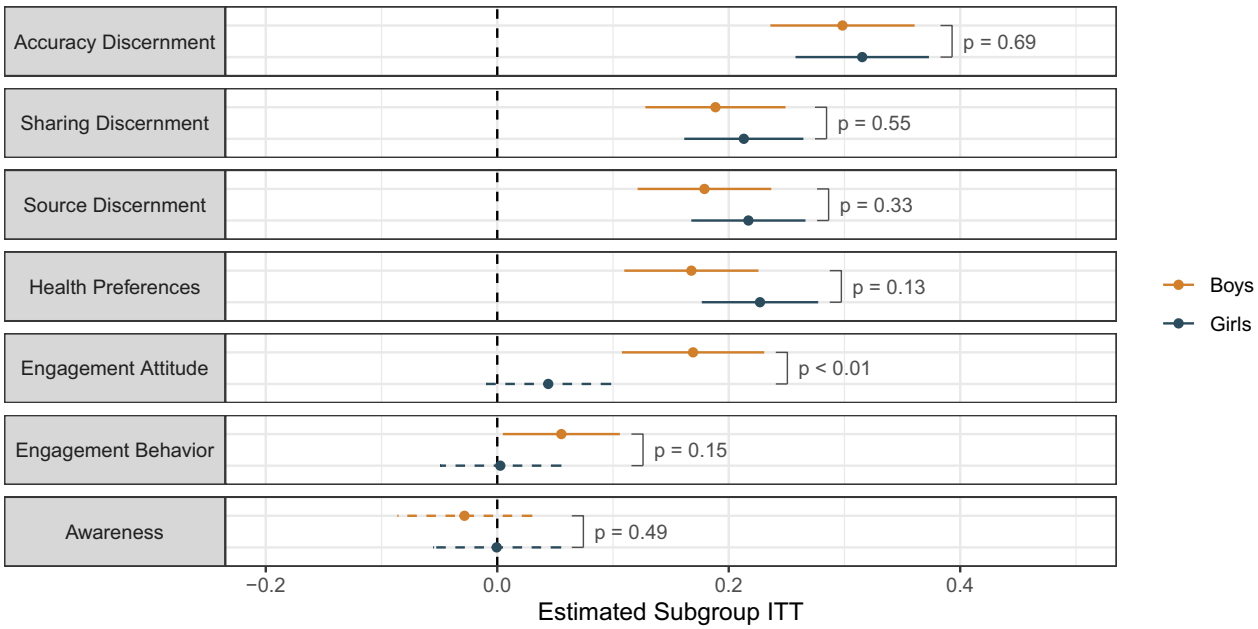
Engagement with Misinformation Countermeasures

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

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

However, the overall null effect on the ITT estimate masks significant gender variation. Analyzing ITT by gender subgroup shows differences in misinformation engagement measures, even though indices for other outcomes show no such variation (Figure 6). Boys are significantly more likely to report intentions to engage in misinformation countermeasures, both in attitudes and behaviors, while the treatment had no effect on girls. Breaking down this result further, control group means for boys are much higher than for girls for both indices. Although point estimates are positive for both groups, boys demonstrate a steeper increase, indicating that updating on these indices is concentrated among those already more amenable to such behaviors (Supplementary Section H2). This result aligns with India's patriarchal context, where strong gender norms condition behavior (Brulé 2020; Heinze, Brulé, and Chauchard 2025; Prillaman 2023). Our indices of behaviors and intentions reflect not only measures on misinformation but also the capacity and willingness to engage in community-based actions, which may require shifts in gender norms (e.g., permission for women to engage publicly) and public safety. For instance, correcting a friend's misinformation demands assertiveness and confrontation, traits not directly targeted by the intervention and particularly challenging to change for women in India. While both girls and boys improved equally in discernment, behavioral changes proved harder where cultural and gender norms created barriers. This suggests that while private preferences can be shifted for all, public behaviors improved only among boys. Achieving similar changes among girls may require interventions that address societal norms alongside misinformation.

FIGURE 6. Effect of Assignment to BIMLI by Gender Subgroup



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

Awareness

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

Heterogeneous Treatment Effects

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

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

TABLE 3. Effect of Assignment to BIMLI Treatment on 4-Month Follow-Up

Outcome	Type	*N*	Estimate	SE	*p*-Value
Accuracy discernment (health)	Index	1,945	0.26***	0.048	< 0.001
Accuracy of true statements	Sub-index	2,053	−0.06	0.040	0.14
Inaccuracy of false statements	Sub-index	1,962	0.33***	0.041	< 0.001
Accuracy discernment (politics)	Index	1,863	0.31***	0.049	< 0.001
Accuracy of true statements	Sub-index	1,992	−0.01	0.041	0.88
Inaccuracy of false statements	Sub-index	1,887	0.31***	0.043	< 0.001
Source discernment	Index	2,028	0.10*	0.044	0.03
Trust reliable sources	Sub-index	2,040	−0.07	0.042	0.08
Distrust unreliable sources	Sub-index	2,056	0.14***	0.041	< 0.001

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

are different as function of being in a high- or low-spillover village.[20] We find that for three outcomes, engagement attitudes, awareness of misinformation, and source discernment, assignment to treatment in a low-spillover village positively affects respondents. This is notable especially with regard to the awareness index, as our main effect was a null result.

Robustness Checks

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

Second Endline

We conducted a follow-up survey with a random subset of 2,059 respondents approximately 4 months after the intervention to assess its long-term effects.[21] The extended time gap is particularly relevant, as India's 2024 general elections occurred between our two endlines—a period when political and partisan attitudes typically become more salient (Michelitch and Utych 2018). The follow-up had three main objectives: (1) to assess whether discernment capacity persisted over time, (2) to evaluate if respondents could apply this skill to political stories—a new and unrelated domain, as the intervention deliberately avoided political topics due to our collaboration with the government, and (3) to examine within-household treatment diffusion to untreated members. To measure this, we interviewed one randomly selected parent or guardian for each of the follow-up households.[22]

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

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

[21] Supplementary Section I describes sampling for the second endline; attrition and compliance are discussed in Supplementary Section K.

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

TABLE 4. Effect of Assignment to BIMLI on Treatment Group Parents/Guardians

Outcome	Type	*N*	Estimate	SE	*p*-Value
Accuracy discernment index	Index	1,786	0.27***	0.054	< 0.001
Accuracy of true statements	Sub-index	2,020	−0.01	0.047	0.88
Inaccuracy of false statements	Sub-index	1,804	0.28***	0.049	< 0.001

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

health), educational interventions can yield transferable benefits across other domains.[23]

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

DISCUSSION AND CONCLUSION

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

These findings are significant given the mixed or null results typically seen in media literacy interventions (Blair et al. 2024). In contrast, our program produced measurable effects in a particularly challenging environment. Bihar, where the study was conducted, has low educational prioritization and a 42% dropout rate before 10th grade (Muralidharan and Prakash 2017). Session compliance in our study reached roughly 70%, a respectable figure given the region's limited state capacity and consistent underperformance on public service delivery (Desai et al. 2019; Jha 2023; Mathew and Moore 2011; Rasul and Sharma 2014). Thus, it was not obvious that a media literacy curriculum like BIMLI would yield positive effects; to the best of our knowledge, this is the first intervention in this context to produce significant belief change in misinformation outcomes. These results suggest that more intensive strategies, featuring peer learning, norm setting, and repeated exposure, may be essential for meaningfully shifting entrenched beliefs, especially where one-off informational interventions have failed.

Despite these encouraging findings, we acknowledge several limitations of the study. First, the intervention was delivered as a bundled, high-dosage program with multiple components, making it difficult to isolate which elements (content, dosage, or delivery format) were most effective, or to tease out mechanisms. Session-wise attendance is not a reliable proxy for variation, as session topics are confounded with peer effects; students attending earlier sessions may form social networks that generate endogenous downstream effects. Moreover, the curriculum involved substantial repetition, with each session revisiting earlier material, further complicating efforts to identify topic-specific impact. Our goal was to design a comprehensive intervention to address the limited success of prior media literacy programs, but future research could unbundle

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

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

the curriculum to assess which elements drive results. A second limitation concerns cost and scalability: implementing such a sustained program required substantial resources. Due to budget and power constraints, we were unable to experimentally vary treatment dosage, but future work could test the minimum intensity required to produce effects.[25] Another important dimension for future exploration involves delivery format. Our treatment combined content, peer learning, and instruction from teachers trained to incorporate interaction and discussion. It is unclear whether the same syllabus, delivered via online modules without peer interaction or a teacher, would yield similar results. Disentangling the roles of content, authority, and peer dynamics will be critical for informing scalable and effective policy design in the future. Lastly, we acknowledge some design limitations. First, we were only able to implement survey-based behavioral measures and behavioral intentions, rather than tracking actual behaviors, due to several logistical constraints. One meaningful outcome we would have liked to track post-treatment is actual vaccine uptake. However, challenges in accessing administrative data and tracking respondents over time made this infeasible. Second, since we had to hire teachers for treatment and control from separate pools, we were unable to implement teacher-level fixed effects to determine if outcomes changed due to teacher quality.

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

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

We attribute these hopeful findings to the setting in which we fielded the study: classrooms and schools have consistently been identified as pivotal sites for knowledge acquisition beyond the household, and public education systems play a crucial role as agents of socialization, especially in contexts where information spread takes place offline. Therefore, our study not only contributes to the literature on persuasion and information processing but also examines the enduring impacts of education and learning. This aligns with existing work exploring the transformative potential of education within schools, investigating education to reshape gender attitudes in India (Dhar, Jain, and Jayachandran 2022) and foster nation-building efforts (Bandiera et al. 2019), along with the potential of interaction with the state via education to shape economic views (Davies 2023). Further, scholars have explored the efficacy of educational tools such as textbooks in persuasion and attitude change (Cantoni et al. 2017), as well as their role in shaping perceptions of representation and marginalization (Haas and Lindstam 2024). 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.

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

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0003055425101184.

DATA AVAILABILITY STATEMENT

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

ACKNOWLEDGEMENTS

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

FUNDING STATEMENT

This research was funded by the Social Science Research Council under their Mercury Project. This project also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme - Grant agreement No. 101002985 (ERC POLARCHATS project).

CONFLICT OF INTERESTS

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

ETHICAL STANDARDS

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

REFERENCES

Ali, Ayesha, and Ihsan Ayyub Qazi. 2023. "Countering Misinformation on Social Media through Educational Interventions: Evidence from a Randomized Experiment in Pakistan." *Journal of Development Economics* 163: 103108.

Alsan, Marcella, and Sarah Eichmeyer. 2024. "Experimental Evidence on the Effectiveness of Nonexperts for Improving Vaccine Demand." *American Economic Journal: Economic Policy* 16 (1): 394–414.

Amar, Priyadarshi, Sumitra Badrinathan, Simon Chauchard, and Florian Sichart. 2025. "Replication Data for: Countering Misinformation Early: Evidence from a Classroom-Based Field Experiment in India." Harvard Dataverse. Dataset. https://doi.org/10.7910/DVN/G2BK1G.

Andrews, Isaiah, and Jesse M. Shapiro. 2021. "A Model of Scientific Communication." *Econometrica* 89 (5): 2117–42.

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): 1325–41.

Badrinathan, Sumitra, and Simon Chauchard. 2023. ""I Don't Think That's True, Bro!" Social Corrections of Misinformation in India." *The International Journal of Press/Politics* 29 (2): 394–416.

Badrinathan, Sumitra, and Simon Chauchard. 2024. "Researching and Countering Misinformation in the Global South." *Current Opinion in Psychology* 55: 101733.

Badrinathan, Sumitra, Simon Chauchard, and Niloufer Siddiqui. 2025. "Misinformation and Support for Vigilantism: An Experiment in India and Pakistan." *American Political Science Review* 119 (2): 947–65.

Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo. 2019. "Nation-Building through Compulsory Schooling during the Age of Mass Migration." *The Economic Journal* 129 (617): 62–109.

Banerjee, Abhijit, Donald P. Green, Jeffery McManus, and Rohini Pande. 2014. "Are Poor Voters Indifferent to whether Elected Leaders Are Criminal or Corrupt? A Vignette Experiment in Rural India." *Political Communication* 31 (3): 391–407.

Bhattacharya, Usree. 2022. ""I Am a Parrot": Literacy Ideologies and Rote Learning." *Journal of Literacy Research* 54 (2): 113–36.

Blair, Robert A., Jessica Gottlieb, Brendan Nyhan, Laura Paler, Pablo Argote, and Charlene J. Stainfield. 2024. "Interventions to Counter Misinformation: Lessons from the Global North and Applications to the Global South." *Current Opinion in Psychology* 55: 101732.

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): 869–85.

Bode, Leticia, and Emily K. Vraga. 2018. "See Something, Say Something: Correction of Global Health Misinformation on Social Media." *Health Communication* 33 (9): 1131–40.

Bowles, Jeremy, Kevin Croke, Horacio Larreguy, Shelley Liu, and John Marshall. 2025. "Sustaining Exposure to Fact-Checks: Misinformation Discernment, Media Consumption, and its Political Implications." *American Political Science Review*, 1–24. https://doi.org/10.1017/S0003055424001394.

Brashier, Nadia M., and Daniel L. Schacter. 2020. "Aging in an Era of Fake News." *Current Directions in Psychological Science* 29 (3): 316–23.

Brass, Paul R. 2011. *The Production of Hindu-Muslim Violence in Contemporary India.* Seattle: 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* 1 (3). https://doi.org/10.37016/mr-2020-028.

Brulé, Rachel E. 2020. *Women, Power, & Property*. Cambridge: Cambridge University Press.
Cantoni, Davide, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang. 2017. "Curriculum and Ideology." *Journal of Political Economy* 125 (2): 338–92.
Carlos, Roberto F. 2021. "The Politics of the Mundane." *American Political Science Review* 115 (3): 775–89.
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): 254–63.
Chauchard, Simon, and Sumitra Badrinathan. 2025. "The Religious Roots of Belief in Misinformation: Experimental Evidence from India." *British Journal of Political Science* 55: e109.
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, 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* 42: 1073–95.
Coppock, Alexander. 2023. *Persuasion in Parallel: How Information Changes Minds about Politics*. Chicago, IL: University of Chicago Press.
Dahlgaard, Jens Olav. 2018. "Trickle-Up Political Socialization: The Causal Effect on Turnout of Parenting a Newly Enfranchised Voter." *American Political Science Review* 112 (3): 698–705.
Das, Anupreeta, and Pragati KB. 2025. "How the Indian Media Amplified Falsehoods in the Drumbeat of War." *The New York Times*. May 17. https://www.nytimes.com/2025/05/17/world/asia/india-news-media-misinformation.html.
Davies, Emmerich. 2023. "The Lessons Private Schools Teach: Using a Field Experiment to Understand the Effects of Private Services on Political Behavior." *Comparative Political Studies* 56 (6): 824–61.
de Barros, Andreas, Johanna Fajardo-Gonzalez, Paul Glewwe, and Ashwini Sankar. 2022. "Large-Scale Efforts to Improve Teaching and Child Learning: Experimental Evidence from India." Unpublished Manuscript.
Desai, Sonalde, Amaresh Dubey, Brij Lal Joshi, Mitali Sen, Abusaleh Sharif, and Reeve Vanneman. 2019. *Human Development in India: Challenges for a Society in Transition*. Oxford: Oxford University Press.
Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2022. "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India." *American Economic Review* 112 (3): 899–927.
Druckman, James N., and Kjersten R. Nelson. 2003. "Framing and Deliberation: How Citizens' Conversations Limit Elite Influence." *American Journal of Political Science* 47 (4): 729–45.
Fazio, Lisa K., David G. Rand, and Gordon Pennycook. 2019. "Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements." *Psychonomic Bulletin & Review* 26: 1705–10.
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 (S1): 127–50.
Gadjanova, Elena, Gabrielle Lynch, and Ghadafi Saibu. 2022. "Misinformation across Digital Divides: Theory and Evidence from Northern Ghana." *African Affairs* 121 (483): 161–95.
Ghosh, Arkadev, Prerna Kundu, Matt Lowe, and Gareth Nellis. 2025. "Creating Cohesive Communities: A Youth Camp Experiment in India." *Review of Economic Studies*: p.rdaf026.
Gottlieb, Jessica. 2016. "Greater Expectations: A Field Experiment to Improve Accountability in Mali." *American Journal of Political Science* 60 (1): 143–57.
Gottlieb, Jessica, Claire L Adida, and Richard Moussa. 2022. "Reducing Misinformation in a Polarized Context: Experimental Evidence from Côte d'Ivoire." Working Paper.
Graham, Matthew H., and Omer Yair. 2025. "Less Partisan but No More Competent: Expressive Responding and Fact-Opinion Discernment." *Public Opinion Quarterly* 89 (1): 7–30.
Green, Donald P., Dylan W. Groves, Constantine Manda, Beatrice Montano, and Bardia Rahmani. 2024. "The Effects of Independent Local Radio on Tanzanian Public Opinion: Evidence from a Planned Natural Experiment." *The Journal of Politics* 86 (1): 231–40.
Guay, Brian, Adam J. Berinsky, Gordon Pennycook, and David Rand. 2023. "How to Think about whether Misinformation Interventions Work." *Nature Human Behaviour* 7 (8): 1231–3.
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): eaau4586.
Guess, Andrew M., and Benjamin A. Lyons. 2020. "Misinformation, Disinformation, and Online Propaganda." In *Social Media and Democracy*, eds. Nathaniel Persily and Joshua A. Tucker, 10–33. Cambridge: Cambridge University Press.
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): 15536–45.
Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, 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. 2024. "My History or Our History? Historical Revisionism and Entitlement to Lead." *American Political Science Review* 118 (4): 1778–802.
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* 25 (1): 110–26.
Harjani, Trisha, Melisa-Sinem Basol, Jon Roozenbeek, and Sander van der Linden. 2023. "Gamified Inoculation against Misinformation in India: A Randomized Control Trial." *Journal of Trial and Error* 3 (1): 14–56.
Heinze, Alyssa, Rachel Brulé, and Simon Chauchard. 2025. "Who Actually Governs? Gender Inequality and Political Representation in Rural India." *The Journal of Politics* 87 (2): 818–22.
Huber, Gregory A., and Kevin Arceneaux. 2007. "Identifying the Persuasive Effects of Presidential Advertising." *American Journal of Political Science* 51 (4): 957–77.
Jaffrelot, Christophe. 2021. *Modi's India: Hindu Nationalism and the Rise of Ethnic Democracy*. Princeton, NJ: 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): 427–50.
Karlan, Dean, and Jacob Appel. 2016. *Failing in the Field: What we Can Learn when Field Research Goes Wrong*. Princeton, NJ: Princeton University Press.
Kosec, Katrina, and Leonard Wantchekon. 2020. "Can Information Improve Rural Governance and Service Delivery?" *World Development* 125: 104376.
Kumar, Krishna. 1986. "Textbooks and Educational Culture." *Economic and Political Weekly* 21 (30): 1309–11.
Lieberman, Evan S., Daniel N. Posner, and Lily L. Tsai. 2014. "Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya." *World Development* 60: 69–83.
Malhotra, Pranav, and Katy Pearce. 2022. "Facing Falsehoods: Strategies for Polite Misinformation Correction." *International Journal of Communication* 16: 2303–24.
Mathew, Athakattu Santhosh, and Mick 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): 412–27.

Mohan, Janani. 2021. "Media Bias and Democracy in India." *The Stimson Center*. June 28. https://www.stimson.org/2021/media-bias-and-democracy-in-india/.
Muralidharan, Karthik, and Nishith Prakash. 2017. "Cycling to School: Increasing Secondary School Enrollment for Girls in India." *American Economic Journal: Applied Economics* 9 (3): 321–50.
Nyhan, Brendan. 2021. "Why the Backfire Effect Does Not Explain the Durability of Political Misperceptions." *Proceedings of the National Academy of Sciences* 118 (15): e1912440117.
Paglayan, Agustina. 2024. *Raised to Obey: The Rise and Spread of Mass Education*. Princeton, NJ: 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): 899–915.
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): zmac004.
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): 590–5.
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): 239–50.
Persily, Nathaniel, Joshua A. Tucker, and Joshua Aaron Tucker. 2020. *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge: Cambridge University Press.
Porter, Ethan, and Thomas J. Wood. 2019. *False Alarm: The Truth about Political Mistruths in the Trump Era*. Cambridge: Cambridge University Press.
Prillaman, Soledad Artiz. 2023. *The Patriarchal Political Order: The Making and Unraveling of the Gendered Participation Gap in India*. Cambridge: Cambridge University Press.
Ramirez, Francisco O, and John Boli. 1987. "The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization." *Sociology of Education* 60 (1): 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: 481–97.
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): 221–39.
Roozenbeek, Jon, Sander Van Der Linden, Beth Goldberg, Steve Rathje, and Stephan Lewandowsky. 2022. "Psychological Inoculation Improves Resilience against Misinformation on Social Media." *Science Advances* 8 (34): eabo6254.
Sen, Somdeep. 2023. "Big Money Is Choking India's Free Press—And Its Democracy." *Al Jazeera.* January 6. 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): 2381–2.
Siddiqui, Danish. 2020. "Hindu Group Offers Cow Urine in a Bid to Ward off Coronavirus." *Reuters.* March 14. 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. 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*, eds. Rich Ling and Scott W. Campbell, 231–50. London: Routledge.
Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3): 755–69.
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): 121–32.
Tully, Melissa, Emily K. Vraga, and Leticia Bode. 2020. "Designing and Testing News Literacy Messages for Social Media." *Mass Communication and Society* 23 (1): 22–46.
Vraga, Emily K., Leticia Bode, and Melissa Tully. 2022. "The Effects of a News Literacy Video and Real-Time Corrections to Video Misinformation Related to Sunscreen and Skin Cancer." *Health Communication* 37 (13): 1622–30.
Washington, Ebonya L. 2008. "Female Socialization: How Daughters Affect their Legislator Fathers' Voting on Women's Issues." *American Economic Review* 98 (1): 311–32.
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): 561–77.