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

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

Word Count: 12,915

Abstract

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

Keywords: Misinformation, Field Experiment, Education, India

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

†PhD Candidate, University of Wisconsin–Madison

‡Assistant Professor, American University

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

¶PhD Candidate, Princeton University

1 Introduction

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

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

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

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

To fill this gap, we conducted a field experiment in Bihar, one of India's least developed states, involving over 13,500 adolescents aged 13–18 across 583 villages. We focus on adolescents because this developmental stage is crucial for moral and identity formation; young people are still open to shaping their attitudes, while adults with entrenched partisan or social identities often resist change when confronted with misinformation corrections (Niemi and Jennings, 1991; Markus and Nurius, 1986; Flynn, Nyhan, and Reifler, 2017). Our intervention targeted students in grades 8 through 12 and consisted of classroom-based sessions on misinformation. Over a 14-week period, students participated in four 90-minute sessions, held approximately every three weeks, with homework assignments between sessions. The curriculum, designed specifically for this study, focused on health misinformation and aimed to (1) enhance scientific knowledge

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

about health and counter health-related misinformation, (2) equip students with broad critical skills and practical tools to encourage a more responsible consumption of information, and (3) shift norms surrounding misinformation.

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

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

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

termeasures, it did result in willingness to change costly behaviors among boys, suggesting that such changes may be more difficult in contexts where conservative gender norms act as barriers for women.

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

Our study contributes to the expanding body of literature addressing strategies to mitigate misinformation on a global scale. But what distinguishes our research from existing work is its departure from the assumption that misinformation predominantly manifests a problem to be addressed online. Prior research has largely operated under this presumption, leading to the development of numerous interventions tailored to online demographics. These interventions include, for instance, online gaming as a strategy to combat misinformation (Roozenbeek and van der Linden, 2019) and the provision of corrective nudges through direct messaging on Twitter (Pennycook et al., 2021). However, such interventions are inherently ill-suited for individuals who do not have regular or sustained access to the internet: in our sample, fewer than 1 in 5 households report access to a mobile phone with internet. Moreover, in contexts like India, those who do have access often use encrypted messaging platforms like WhatsApp, where interventions reliant on platform-based changes are not feasible (Badrinathan, 2021). Indeed,

prior research in developing contexts has underscored the significance of offline sources and communal spaces for news dissemination and consumption, such as interpersonal communication, the role of the family, and face-to-face interactions in community settings like marketplaces (Gadjanova, Lynch, and Saibu, 2022). Operating under this premise, one of our primary contributions is studying adolescents. In doing so, we target a demographic with potentially less entrenched attitudes. Insofar as studies on misinformation consistently indicate that a key impediment to the effectiveness of interventions is motivated reasoning – the inclination to accept or reject information based on pre-existing beliefs and identities (Taber and Lodge, 2006) – we thus potentially engage with a population that has not yet developed resistance to altering entrenched viewpoints (Niemi and Jennings, 1991). Finally, partnering directly with a government agency to implement this experiment as an official program boosted both the legitimacy and scalability of our intervention, reducing non-compliance, and simulating a real-world rollout of a school-based program.

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

2 Sustained Education Against Misinformation

The global rise of misinformation has attracted substantial academic and policy attention (Persily, Tucker, and Tucker, 2020), sparking a surge of experimental studies identifying solutions to the problem. However, many misinformation interventions have limitations. Their effects often

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

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

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

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

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

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

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

BIMLI was, by design, a significantly more intensive and substantively different intervention against misinformation than those typically tested. We introduced several key design and theoretical innovations to aid the successful implementation of the program. First, in terms of mode of delivery, we administered the program face-to-face in classroom settings, fostering a peer-based, interactive environment with iterative learning, where respondents encountered key lessons repeatedly over multiple sessions. Research suggests that peer interactions in classroom

settings can deepen understanding by exposing learners to diverse perspectives (Dhar, Jain, and Jayachandran, 2022), while repeated exposure allows for reinforcement of concepts (Fazio, Rand, and Pennycook, 2019). Additionally, partnering with an agency of the Bihar state government to roll out the program as an official government-endorsed certificate course increased its legitimacy. This official recognition was aimed at incentivizing participation and higher compliance.

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

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

3 The Politics of Misinformation in India

Health-related misinformation is widespread in India. From our own control group data, 55% of respondents reported believing that exorcism can cure snake bites. In other studies from similar contexts (Chauchard and Badrinathan, 2024), over 60% of respondents claimed that cow urine could cure covid-19. While this type of belief may seem harmless, it can have severe consequences

by discouraging citizens from seeking actual medical solutions, such as hospital treatment, leading to potentially fatal outcomes (Bridgman et al., 2020). Such health misinformation may also lead to belief in falsehoods and conspiracies in other domains, highlighting its contagion potential (van Prooijen and Böhm, 2024). The negative consequences of belief in misinformation may be particularly pronounced in regions with lower levels of state capacity and socio-economic development (Badrinathan and Chauchard, 2023a).

In India, such deeply entrenched beliefs are tied to social identities, and are often manipulated by political elites to gain electoral support. Traditional health remedies, many of which are linked to ancient Hindu culture, are used as a tool to garner support from the Hindu voter base. This is especially evident with the rise to power of the Hindu nationalist Bharatiya Janata Party (BJP), the party leading the current federal government in the country, which has positioned itself as a champion of Hindu values (Jaffrelot, 2021). A recent example of politically-driven health disinformation involves a BJP politician hosting a public event where participants drank cow urine, believing it to be a cure for covid – an event that ended with several attendees being hospitalized (Siddiqui, 2020). Politicians from the BJP have even boasted about their ability to make anything go viral, true or false (Kumar, 2019). Previous research demonstrates that misinformation that resonates with long-standing identities can be particularly hard to dislodge (Nyhan, 2021; Chauchard and Badrinathan, 2024). Indian politics has long been dominated by a fundamental cleavage between Hindus and Muslims, and the prominence of religion as a social identity has been central (Brass, 2011). Misinformed beliefs, particularly when propagated by elites seeking to further divide already polarized populations, may thus be very difficult to change. Some research in this context has shown that Indian citizens do hold deeply entrenched misinformed beliefs, with partisan motivated reasoning (Taber and Lodge, 2006) posing a significant barrier to efforts by civil society and other actors to correct falsehoods (Badrinathan, 2021). While Bihar, the context of our study, is not directly under single-party BJP rule, it is governed by a BJP-led coalition. Additionally, its geographical location in northern India places it within the broader media landscape of neighboring states such as Uttar Pradesh — regions that have witnessed multiple instances of elite-driven, high-profile disinformation, at times resulting in violence and fatalities (Badrinathan, Chauchard, and Siddiqui, 2024).

For citizens in such contexts, finding ways out of the misinformation trap can be chal-

lenging. This is particularly true in Bihar, India's poorest state, home to 127 million people, and with over one-third of the population living below the poverty line. The state's relative underdevelopment translates into a lack of essential services such as healthcare and education, alongside the failure of many public programs (Sharma, 2015). Children in Bihar, especially girls, are significantly less likely to attend school compared to those in other states (Muralidharan and Prakash, 2017). Access to the internet is also limited: as per our baseline data, only 11.5% of respondents reported owning a personal cellphone; while many more had access to an internet-enabled smartphone, only 19% of them reported using the internet on that device. With most interactions and sources of information being offline, children primarily rely on their families for information. However, adults may themselves be misinformed, and strong cultural norms around deference to elders and family make such misinformation harder for children to challenge (Malhotra and Pearce, 2022). Even in households with internet access in our sample, it was often through a shared mobile phone, highlighting the stark contrast with Western settings, where internet access is more individualized (Steenson and Donner, 2017). This limited connectivity is compounded by a deteriorating informational environment across India. Independent media and spaces for dissent are shrinking, as state capture of institutions, including news sources, intensifies (Mohan, 2021; Sen, 2023). This aligns with broader observations of democratic decline in the country in recent years, where the space for credible information has narrowed significantly alongside decreasing state capacity (Tudor, 2023).

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

Despite these challenges, public trust in state institutions appears to be notably high in Bihar. During the study's design phase, for example, parents indicated their trust in Jeevika, our government partner organization, as a key reason for their support of the program. Our data also show that a substantial proportion of parents cited trust in Jeevika as their primary motivation for participating. This aligns with previous research highlighting Jeevika's successful

initiatives in Bihar, including lending programs (Hoffmann et al., 2021), promoting women's financial autonomy (Datta, 2015), and supporting agricultural livelihoods (Baul et al., 2024). These findings suggest that citizen interaction with Jeevika and the government is both commonplace and effective in Bihar, potentially fostering trust. This further mirrors research indicating lower socio-economic households and residents of less developed states, like Bihar, may trust state governments more, likely due to reliance on government programs (Kumar, Pratap, and Aggarwal, 2021). Thus, while Bihar's structural challenges may amplify the impact of misinformation, citizen trust in its institutions provides an opportunity to leverage this reliance on both political and educational institutions to counter misinformation effectively. The next sections describe our field experiment that aimed to do so.

4 Experimental Design and Data Collection

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

4.1 The Treatment

The BIMLI program featured four classroom sessions, each about 90 minutes long and approximately 2-3 weeks apart, as well as homework assignments between sessions. We created a custom curriculum and lesson plan for BIMLI for this study. In doing so, our educative curriculum, though bundled, focused on media and information literacy and critical thinking, with the goal of changing norms and providing knowledge and skills. Media and information literacy is broadly defined as the skills and competencies that promote critical and responsible engagement with information and media sources (Jones-Jang, Mortensen, and Liu, 2021). In practice, however, this concept is operationalized in a myriad of ways, with scholars prioritizing different elements of the concept such as critical thinking, knowledge about media industries, or knowledge about psychological biases in thinking. In Table 1 we provide a summary of the treatment

lesson plan, including a description of learning objectives, modules included in each session, key theoretical works on which curriculum design relied, and strategies to tailor the treatment to the local context.

Substantively, our curriculum focused on fostering critical thinking rather than offering prescriptive tips to spot misinformation. Such tips appear to be the way in which many misinformation interventions operationalize "digital literacy" (Guess et al., 2020) or "media literacy" (Hameleers, 2020). Instead, we provided a theoretical framework for evaluating information broadly, regardless of its source. This approach was particularly suited to the Indian context, where much information is shared through friends and family, making source-specific advice (e.g., favoring one newspaper over another) ineffective. Given the decline in mass media credibility amidst democratic backsliding (Mohan, 2021), we also avoided endorsing specific media outlets. Instead, we emphasized cues to critically assess information, such as recognizing emotional tone, not relying on shared ethnic identities as a cue to assess information, and identifying appropriate authorities for specific topics – for instance, relying on community health workers employed by the government (called ASHA workers) for health-related information.

We collaborated with DataLeads, a Delhi-based organization specializing in media literacy, alongside local Bihar educators and Indian media literacy experts, to design this curriculum. The modules consistently incorporated discussion-based activities to ensure engagement, and time-use lesson plans for teachers ensured that their delivery was highly standardized across classrooms. A central component of our teacher training emphasized creating interactive classroom sessions that encouraged engagement between teachers and students, as well as among students themselves. This approach aimed to cultivate analytical thinking, critical reasoning, and deep learning rather than relying solely on passive reception of information from an authority figure. This method represents a significant departure from the traditional structure of schooling in India (Kumar, 1986). For instance, in Session 4 the lesson plan incorporated role-playing exercises in the classroom. In one activity, a student took on the role of a child while another acted as a parent, with the child tasked with employing strategies to engage with a parent that shared misinformation at a family dinner. The scenario aimed to highlight the challenges of addressing health misinformation with adults, particularly when such discussions involve confronting deeply ingrained beliefs in settings where confrontation with adults is discouraged (Malhotra

Table 1: The BIMLI Curriculum

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

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

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

4.2 Administering classes

To bolster the credibility of BIMLI, we signed a memorandum of understanding to secure official collaboration with an agency of the Bihar state government, the Bihar Rural Livelihoods Promotion Society (BRLPS, or as it is commonly known, Jeevika). Despite their governmental affiliation, Jeevika operates autonomously under the leadership of an Indian Administrative Services officer. Our engagement with Jeevika commenced in mid-2022, driven by a shared interest in combating health misinformation, particularly among youth.

In our study, participants were school students in grades 8 through 12, aged between 13 and 18 years old. To dispense the intervention classes, Jeevika made available to us 100 community libraries across 32 districts in Bihar.[4] We ran our classes in these libraries from November 2023 to March 2024. We opted to deliver these classes after school hours and to not use existing public school teachers or school settings for this program. Our meetings with government officials revealed that public school teachers were already overburdened and that absenteeism (in students and teachers) was common in Bihar, which would have made compliance a big challenge. Besides, adding a module or topic in public school curriculum requires extensive bureaucratic authorizations and oversight. Recognizing that the success of such a program is con-

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

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

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

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

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

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

[7]Data from the ASER survey, the Annual Status of Education Report which provides data from annual surveys on children's schooling and learning levels in rural India, highlights some of these issues in public schools. For example, their 2022 report points out that on the days that ASER surveyed schools, only 50% of enrolled children were actually present in public schools in Bihar; that number is the lowest in the 28 states in India that the survey spans.

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

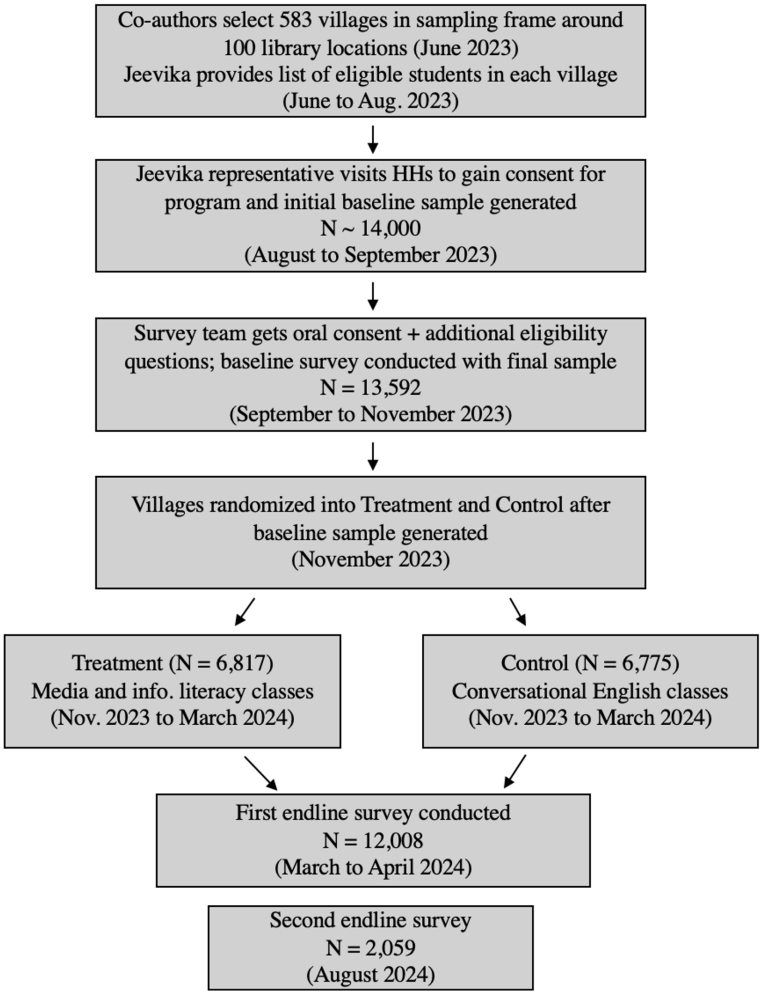
4.3 Sampling, enrolment and baseline data

Figure 1 outlines the timeline and flow of recruitment and roll-out of the study. We sampled $\approx$ 6 villages within a 3 km distance from each of the 100 libraries and randomized roughly 50% of these (typically, 3) around each library to receive our treatment; the remaining served as control villages.[8] Our sampling procedures ultimately allowed us to select 583 villages which served as the sampling frame for this study. We determined pre-randomization whether each village had high- or low-spillover potential, and then assigned to treatment and control within each library and spillover strata (see Appendix A for a detailed note on sampling).

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

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

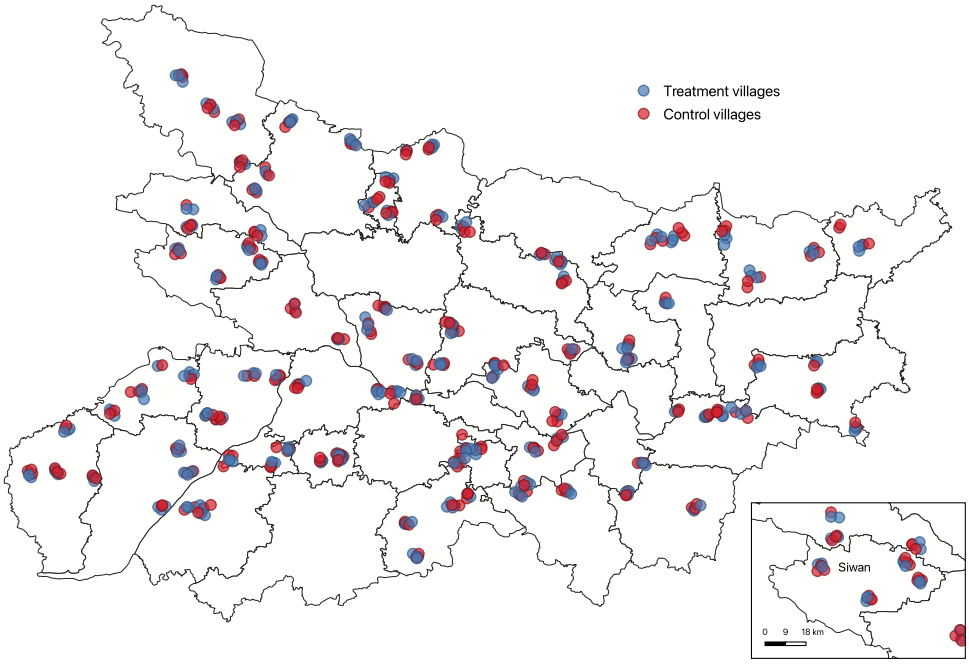
Figure 1: Study Flow and Timeline



Bihar. Crucially, we note that randomization occurred *after* students opted in, avoiding issues with differential opt-in rates between treatment and control.[9] Everyone involved in the study – including teachers, implementation partners, government officials, and coauthors – were blind to treatment status during recruitment and baseline data collection. During household visits, the recruitment pitch stated that students could participate in a free, government-endorsed certificate course with four sessions, designed to benefit their future careers. Students were unaware of their treatment assignment until the first session. Appendix A details sampling procedures.

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

Figure 2: Map of treatment and control villages



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

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

4.4 Control condition

Control group units participated in four modules of conversational English language classes, serving as a placebo rather than a pure control. This was done to achieve parity in effort exerted by students, since school attendance is a major problem in Bihar, and since our intervention lasted 4 months. We aimed to create comparable classroom dynamics and peer interactions, varying only the content of instruction. Encouraged by Jeevika, we additionally wanted control group respondents to benefit from the program and hence chose a topic that fostered engagement without confounding misinformation outcomes. Subjects like math, science, and history were excluded due to overlap with standard curricula or national identity narratives, and several other topics like cooking were discarded due to potential gender biases. We ultimately implemented a curriculum of four sessions on basic conversational English given students had very limited prior exposure. The curriculum focused on spoken skills, covering self-introduction, naming objects, describing activities, and asking questions, using role-playing and group exercises similar to those in the treatment group (see Appendix B). Topics avoided media, technology, and politics, and the very basic instruction level was unlikely to enable control students to independently navigate new information sources. The teacher selection and training differed between the treatment and control groups. DataLeads recruited and trained treatment teachers, while English class teachers were recruited via a local Bihar consultant, resulting in variations in socio-economic characteristics and teaching experiences. Consequently, the treatment effects we measure are influenced by both the treatment content and the teachers' differing backgrounds. Appendix E summarizes teacher demographics by group.

4.5 Endline data and compliance

Our first endline survey was conducted in-person in the weeks following the end of the fourth and last session. Because of the large sample size, the endline took 5 weeks to complete. Although we did not randomize the order in which villages were surveyed at endline, we expect that endline survey timing, while dependent on a number of village-level covariates, is independent of treatment status. At the first endline, we were able to re-contact 12,008 of the total houses sampled at baseline, with an attrition rate of 11.3%. There is no significant difference in

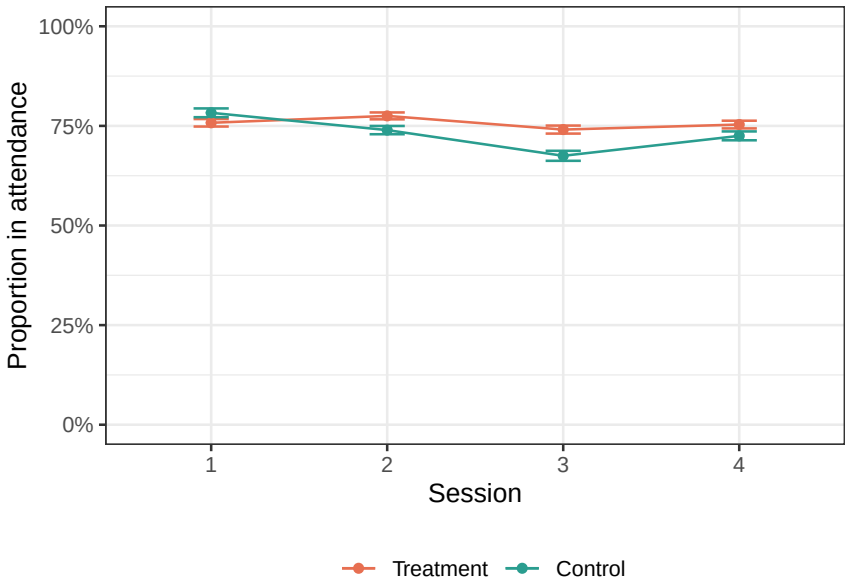
attrition between treatment and control, although we do find that attrition is lower among female students and higher among older respondents (see Appendix C). Moreover, from fieldwork and interviews with enumerators, we note that houses that attrited at endline did so because we were unable to contact them after several tries (in most cases, this was because the respondent or parent was not at home). Crucially, no household refused our survey team entry for the endline survey. We conducted a second endline survey about 4 months, on average, after the intervention, to assess if treatment effects persisted over time. This survey was conducted over the phone with a random subset of 2,059 students and, in each case, one parent or adult guardian.[11]

To boost compliance, we implemented a detailed monitoring system. Jeevika staff, women known locally as *didis*, regularly reminded households about upcoming classes. Students were motivated by the promise of a government-issued certificate upon completing the program. They also received notebooks, pens, flyers, homework, and snacks during each session. External monitors also made random visits to verify teacher presence and adherence to class schedule. Co-authors also visited during initial and final sessions. Teachers were required to upload respondent- level attendance data after each session via an app.

On average, students attended 2.97 classes and 52.7% of the sample attended all four classes across treatment and control. This attendance allows us to measure whether students were fully non-compliant (attended 0 sessions), fully compliant (attended all 4 sessions) or partially compliant (1-3 sessions). Crucially, our definition of attrition does not hinge on compliance with the treatment. Even participants who are (partially) non-compliant with the treatment but continue to engage with the study by completing the endline survey are not classified as having attrited. Compliance data demonstrates that on average, 76% of the sample attended the first class. We detect no significant difference in attendance across treatment and control, with similar proportions attending both sets of classes. However, we do see a significant drop off in attendance for control group respondents during session three, though we note that the difference is substantively small (67% in control group and 74% in treatment) and dissipates during session 4 (see Figure 3). Further, we find that girls were more likely to attend classes (both treatment

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

Figure 3: Compliance data across treatment and control



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

4.6 Outcome measures

We pre-specified and included seven families of outcomes in the first endline survey: accuracy discernment, sharing discernment, health attitudes, trust in sources, engagement with misinformation countermeasures (attitudes), engagement with misinformation countermeasures (behaviors), and awareness of misinformation.[13] Since each outcome family includes multiple items, we compute inverse-covariance weighted (ICW) indices that combine and weight individual items, that are standardized relative to the control group mean and SD. Ultimately our main analysis focuses on seven outcome indices. Appendix D details the rationale for ICW indices and their pre-specified construction. The full endline instrument is posted to our pre-analysis plan. In

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

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

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

5 Estimation and Results

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

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

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

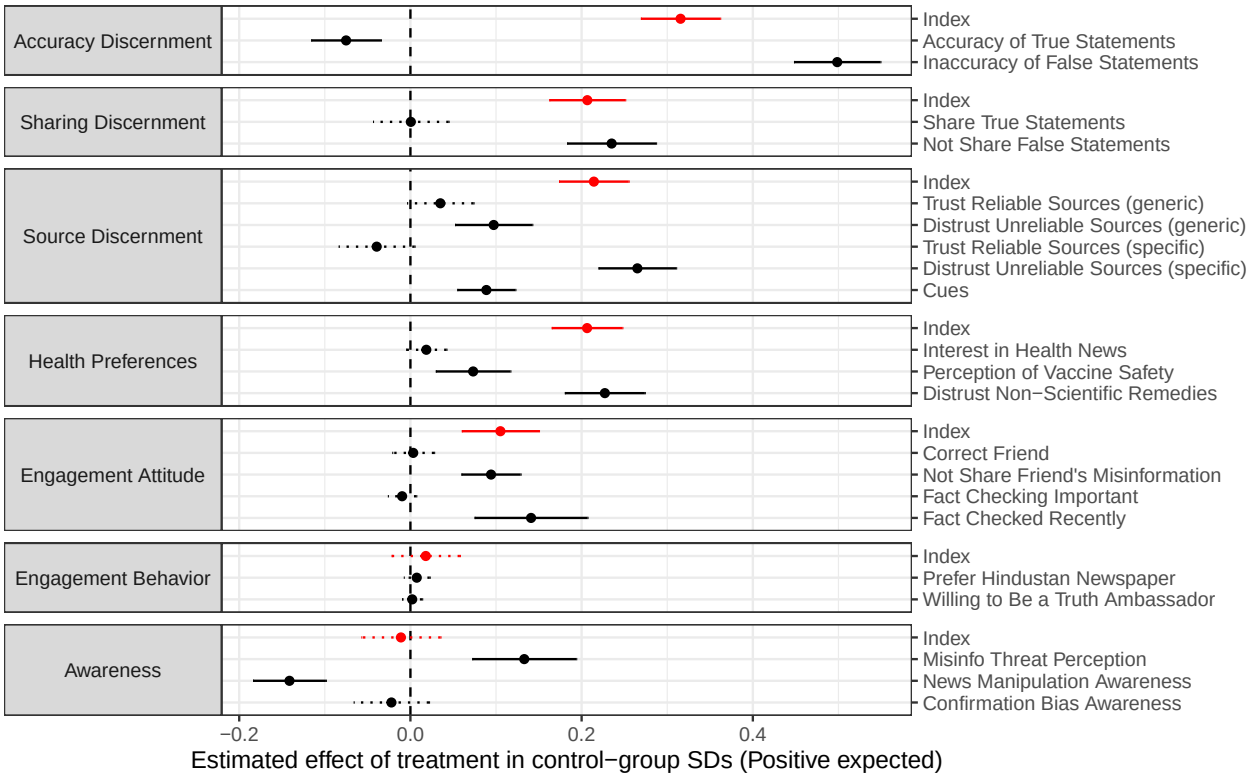
where Y_{ijk} is the primary outcome of interest Y for student i in classroom j and library-spillover strata $k \in \{1, ..., m\}$, β_0 is the intercept, T_{ijk} is a treatment indicator, α_c denotes the coefficient for the control variable X_c, γ_k denotes fixed effects for each library-spillover strata k, and ε_i denotes the random error term for individual i. β_1 denotes the estimated effect of treatment assignment (ITT) on outcome Y. To estimate this equation, we use linear regression with heteroskedasticity-robust standard errors, clustered at the village level. To complement the ITT analysis, we also estimate complier average causal effects (CACE) using a 2SLS model and operationalizing compliance continuously ranging from 0 (0/4 sessions attended) to 1 (4/4 sessions attended).[14]

5.1 First Endline

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

[14] See Appendix J for CACE specification.

Figure 4: Estimated effect of assignment to BIMLI treatment



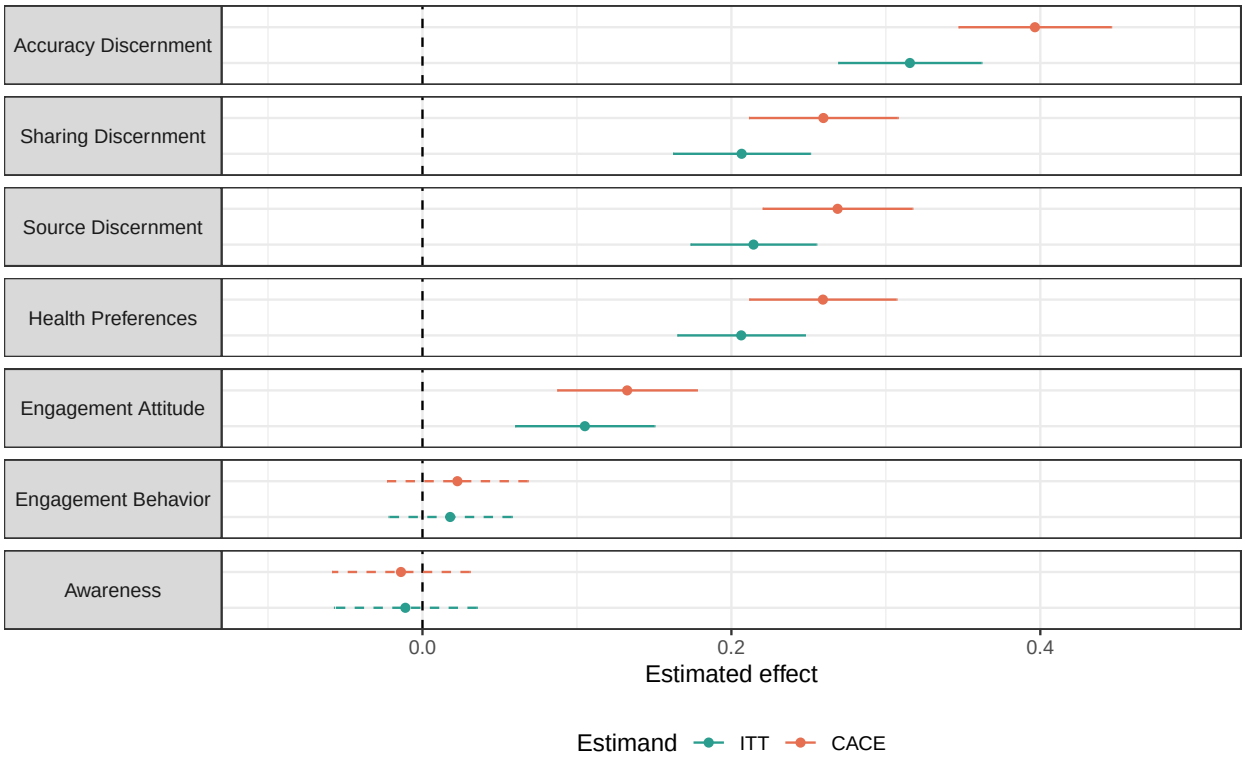
Notes: This figure plots the effect of assignment to BIMLI for 7 outcome families. Each index is an ICW calculation of components within an outcome family. Each component is standardized relative to the control mean and SD. Confidence intervals are at the 95% level and are based on standard errors clustered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

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

5.1.1 Accuracy and sharing discernment

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

Figure 5: Estimated effect of BIMLI treatment on compliers



Notes: This figure plots a comparison of ITT versus CACE effects for 7 outcome families. Each index is an ICW calculation of components within an outcome family. Each component is standardized relative to control mean and SD. Confidence intervals are at the 95% level and are based on standard errors clustered at the village (classroom) level. Tabular results and model specification for CACE estimates are in Appendix J.

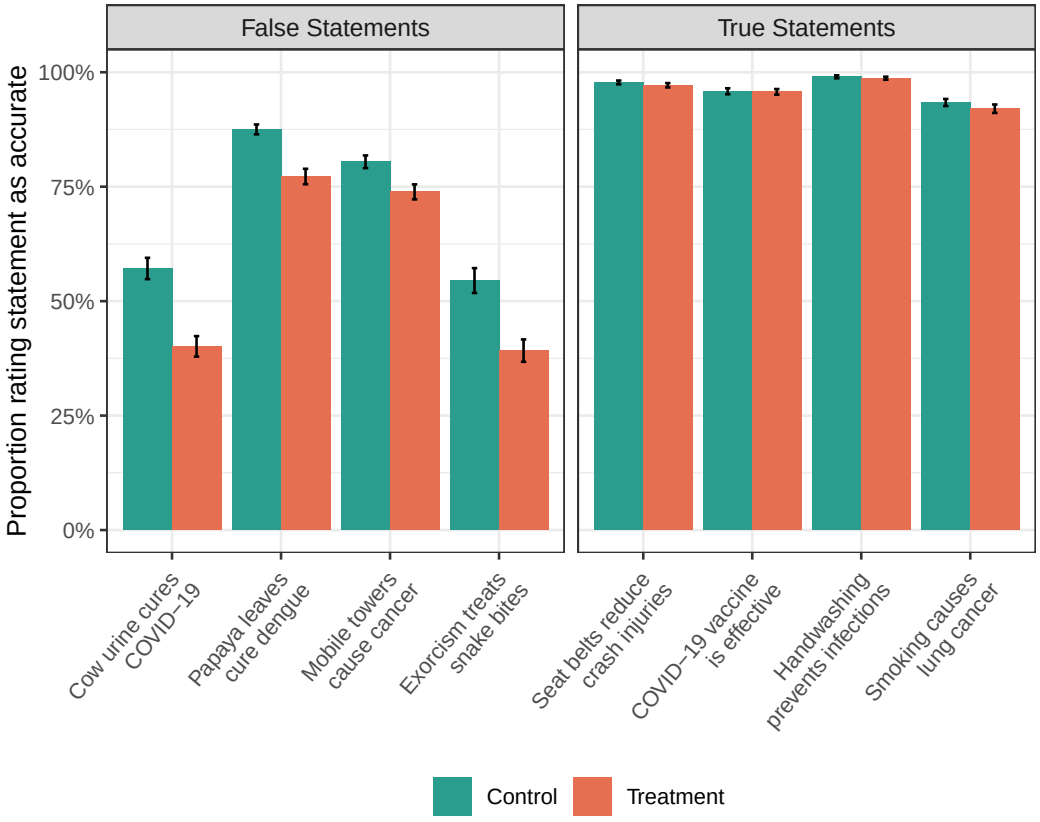
skill application rather than mere recall.[15] We also measured sharing discernment (intention to share each story) using the same items.[16] The selection for these stories was based on familiarity, cultural relevance, and minimizing the introduction of new misinformation. After fieldwork and piloting, we identified the most commonly believed health-related myths, each debunked by at least one fact-checking service in India. Stories were presented to respondents in random order.

With respect to accuracy discernment, ITT estimates show that the treatment significantly helped respondents discern between true and false stories (Figure 4). The magnitude of this effect, a 0.32 SD increase in discernment relative to the control group, is substantively large

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

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

Figure 6: Accuracy discernment by headline



compared to discernment effects detected in research on misinformation in comparable contexts. For example, Guess et al. (2020) find that their digital literacy intervention in India led to a 0.11 SD increase in discernment, while Gottlieb, Adida, and Moussa (2022)'s intervention in Cote d'Ivoire produced effect sizes of 0.12 to 0.15 SD. Further, when we compare ITT to CACE estimates, we find that the effect on accuracy discernment is even larger (0.39 SD) among compliers (see Figure 5). We also see variation in the true and false components of the discernment measure: we find that while the treatment made respondents marginally more skeptical of all news, the magnitude of this effect is much larger for the false statements, and hence the overall effect is a net positive. In Figure 6 we break down the true and false components into individual headlines. The graph demonstrates that large proportions of respondents in the control group thought that false statements were accurate and the treatment significantly decreased respondents' perceived accuracy of all 4 false stories, with effect sizes ranging from 0.44 SD (cow urine can cure covid) to 0.18 SD (mobile phone towers cause cancer). With respect to true stories, there is little variation

in how treatment and control group respondents rated these stories; on average all respondents were better at discerning true stories relative to false.

While we observe varying effects on true and false stories, we attribute this to a ceiling effect for true stories, as belief in these was already high, leaving little room for further increase. Substantively, we do not view the effects on true stories as very concerning, insofar as effects are small, especially before standardization, and arguably caused by these ceiling effects. Moreover, in the current informational environment in India, we view a small degree of skepticism towards all information – especially given high baseline levels of trust in mass media in our sample – as normatively acceptable. Over the past decade, the quality of democracy has significantly declined, affecting the credibility of news sources. Many traditional media outlets have been acquired by government allies, leading to a pro-government bias across these outlets, with ideological diversity in available news sources narrowing (Mohan, 2021). While credible sources do exist, they are predominantly in English and paywalled, making them inaccessible to our sample. In such an environment, generalized skepticism may be warranted. Finally, we note that our second endline yields precise null effects on discernment measures for true information (Table 2). This suggests that while the intervention's positive effect on reducing belief in false information persists over time, the temporary decrease in belief in true information is no longer detectable in the follow-up.

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

5.1.2 Trust in sources and source discernment

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

First, we measure general source discernment by asking respondents to rate their trust of transmitters (e.g., word of mouth), mediums (e.g., radio, Facebook), and institutions (e.g., the WHO). The index includes two sources we expect to increase trust in (MBBS doctors, healthcare workers) and two we hope to decrease trust in (ayurvedic doctors, rumors). Next, we assess situation-specific trust by using a vignette where respondents seek emergency advice for a sick family member and could go to a number of sources. We provide three trustworthy sources (community health center, government materials, TV doctors) and three untrustworthy ones (family remedies, WhatsApp forwards, TV interviews with ayurvedic doctors). This helps distinguish between general and specific trust and separates transmitters from mediums. Finally, we explore which factors foster trust in information, examining signals like likes/shares online, shared ethnicity, as well as message tone and emotionality. Our results show that BIMLI, overall, significantly changed how respondents interact with and trust sources, with a notable shift in the index (SD = 0.21). As seen from Figure 4, these effects are primarily driven by increased distrust in unreliable sources (both generic and specific).

5.1.3 Health preferences

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

Despite the prevalence of health misinformation and reliance on alternative medicine, we show that BIMLI was able to significantly alter respondents' health preferences (index 0.21 SD). Item-wise results indicate that the treatment reduced vaccine hesitancy and stated reliance on alternative forms of medicine. This finding holds significance for several reasons. First, traditional home remedies and the misinformation surrounding them have long existed in India, passed down through generations, suggesting that these beliefs may be deeply ingrained and therefore more resistant to change. Additionally, prior research has indicated that belief in medical misinformation in India is associated with social identities such as religion and partisanship, and given that these identities underpin enduring societal divisions (Chauchard and Badrinathan, 2024), motivated reasoning may impede the effectiveness of misinformation countermeasures (Taber and Lodge, 2006). Despite this, BIMLI had a significant impact on altering health preferences, including preferences regarding which forms of medicine to rely on during crises and beliefs about the efficacy of alternative treatments.

5.1.4 Engagement with misinformation countermeasures

We assessed engagement with misinformation countermeasures using attitudinal and behavioral measures. Attitudinally, we focused on shifting norms around misinformation through four self-reported measures: (1) likelihood of correcting a friend sharing misinformation, (2) likelihood of personally sharing misinformation from friends, (3) perceived importance of verifying information, and (4) frequency of verifying information in the past two months. The treatment significantly influenced respondents' attitudes on this index, but we observed variation across items. Treated respondents were more likely to abstain from sharing misinformation, even from close acquaintances, but were hesitant to correct it, reflecting cultural norms in India that may discourage direct confrontation (Malhotra and Pearce, 2022). While respondents hesitated to correct friends, the shift toward not sharing misinformation suggests that the treatment was effective.

Children in India are accustomed to tests and often excel in educational settings. To ensure our findings were not solely driven by this familiarity, we incorporated two behavioral measures. First, respondents entered a lottery to choose between two subscriptions: a credible Hindi newspaper, *Hindustan*, or a popular magazine, *Manohar Kahaniyan*, featuring horror, mystery, and

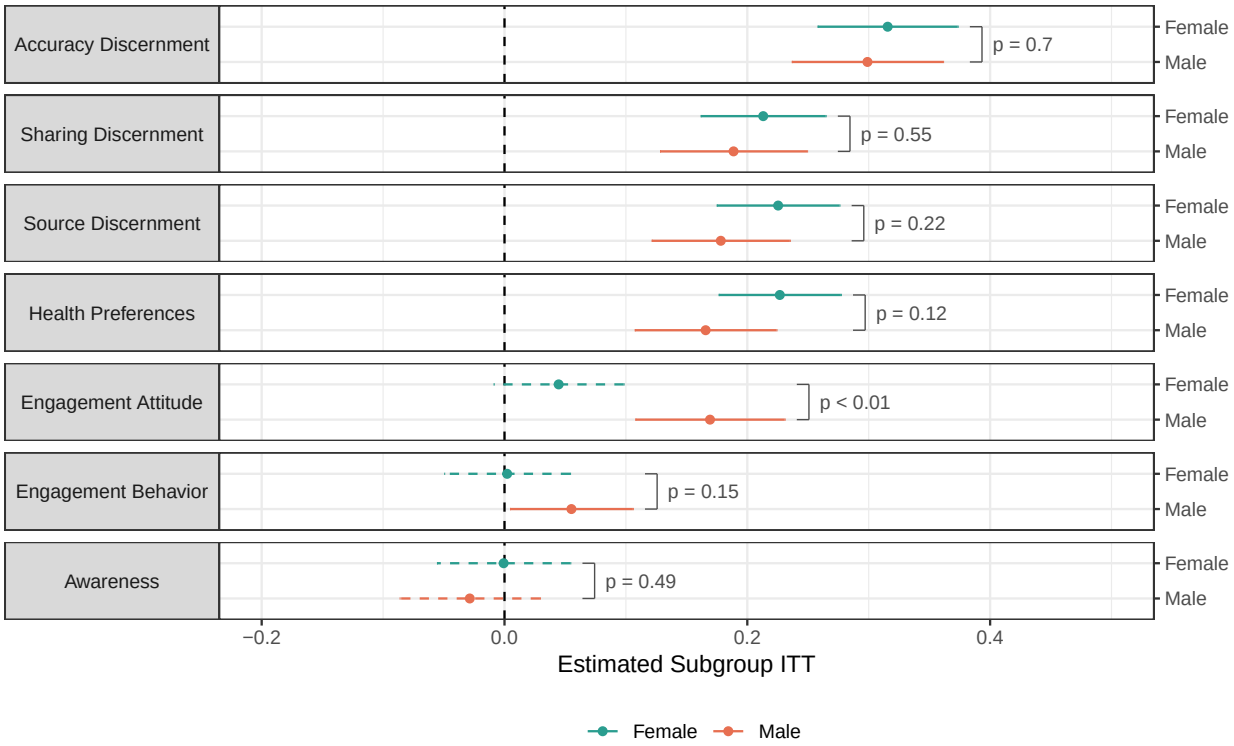
romance stories. We hypothesized greater demand for news among the treatment group, while noting that this measure is unrelated to misinformation directly. Second, we invited respondents to become "truth ambassadors," a community role described as supporting local government by dispelling misinformation during crises, costly in terms of time and effort. We expected higher willingness for this role in the treatment group. ITT results showed no significant impact on these behaviors, with the overall index a null effect.

However, the overall null effect on the ITT estimate masks significant gender variation. Analyzing ITT by gender subgroup reveals notable differences in misinformation engagement measures, even though indices for other outcomes show no such variation (Figure 7). Boys are significantly more likely to report intentions to engage in misinformation countermeasures, both in attitudes and behaviors, while the treatment had no effect on girls. Breaking down this result further, control group means for boys are much higher than for girls for both indices. Although point estimates are positive for both groups, boys demonstrate a steeper increase, indicating that updating on these indices is concentrated among those already more amenable to such behaviors (Appendix H.2). This result aligns with India's patriarchal context, where strong gender norms condition behavior (Brulé, 2020; Prillaman, 2023; Heinze, Brulé, and Chauchard, 2024). Our indices of behaviors and intentions reflect not only measures on misinformation but also the capacity and willingness to engage in community-based actions, which may require shifts in gender norms (e.g. permission for women to engage publicly) and public safety. For instance, correcting a friend's misinformation demands assertiveness and confrontation, traits not directly targeted by the intervention and particularly challenging to change for women in India. While both girls and boys improved equally in discernment, behavioral changes proved harder where cultural and gender norms created barriers. This suggests that while private preferences can be shifted broadly, public behaviors improved only among boys. Achieving similar changes among girls may require interventions that address gender norms alongside misinformation.

5.1.5 Awareness

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

Figure 7: Effect of assignment to BIMLI by Gender Subgroup



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

significantly increased this perception, 78% of respondents already viewed misinformation as a threat, limiting room for further change. Awareness of media and cognitive biases was also measured using items adapted to the Indian context from Ashley, Maksl, and Craft (2013). These definitional items focused on theoretical classroom concepts, and we find no improvement for treated respondents compared to control (p = 0.64). This could be due to (1) the time gap between lessons and the survey: biases were introduced in session 2, at least two months before the endline, (2) the curriculum's focus on application rather than rote learning, and (3) the complexity of these theoretical concepts. Treated respondents were also less likely to express agreement with statements suggesting ways in which the reporting of news in the media can be biased – a point that had been covered in session 2 (see Table 1). Overall we find a null effect on the awareness index. Despite this, we underscore that the significant effects on discernment and other outcomes

suggest respondents were able to successfully retain and apply skills learned in the classroom.

5.1.6 Heterogeneous treatment effects

We pre-registered heterogeneous effects analyses based on a number of variables. Most importantly, to proxy motivated reasoning, we examined interaction effects with partisan identity. While direct questions about party ID were not permitted in the baseline survey due to our collaboration with the government, we estimated household-level partisanship through additional surveys with village-level local elites. We surveyed 1,664 elites across 550 villages and asked questions on sub-caste category-wise party preferences for recent elections. Matching this data back to our baseline, we were able to estimate party ID at the household level.[17] We also analyzed heterogeneous effects for household mobile internet access as previous work indicates that prior exposure to media and the internet can influence how individuals interact with misinformation (Guess et al., 2023). Demographically we looked at socio-economic status (proxied by an asset index), age (student grade) and gender. We also looked at basic science knowledge. Finally, we looked at whether results are different as function of being in a high- or low-spillover village. The results, detailed in Appendix H, show no consistent patterns. Aside from the gender subgroup effects discussed earlier, we found no systematic interaction effects for mobile internet usage, age, or partisanship. This is notable, as past research suggests that partisanship often moderates the impact of misinformation interventions (Flynn, Nyhan, and Reifler, 2017). Our findings indicate that belief change in this context was driven by a model of learning and updating, with no obvious pattern of motivated reasoning (Coppock, 2023), consistent with our argument about adolescents having less fixed political attachments.

5.1.7 Robustness Checks

To test the robustness of our results, we undertake several analyses. First, we re-estimate the baseline model incorporating library fixed effects, district fixed effects, and district-spillover stratum fixed effects. The main results remain unchanged. Second, we run an adjusted model with pre-registered control variables, including demographics (age, gender, grade, caste, religion, lan-

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

guage of schooling), household-level variables (asset index as a proxy for wealth and access to mobile internet), baseline covariates (reading skill and science knowledge indices), and village-level variables (development proxied by nighttime lights data, and partisanship measured by BJP vote share in the last assembly election). Results are robust to these controls. Next, to determine whether the effects are concentrated in specific districts, we examine ITT estimates by district subgroup. The results show no clear pattern. Notably, given significant variation in districts' socio-economic development, we confirm that these findings are not correlated with development outcomes. Following this, we apply multiple corrections across indices, as pre-registered. Results on our main dependent variables remain significant. Finally, to exclude the possibility that our results are driven by differential attrition between treatment and control based on unobservables, we undertake sensitivity analyses using a tipping point method. We show that our analyses are robust to very high levels of differential attrition based on unobservables and find no evidence of differential attrition based on observables. All results on robustness checks are reported in Appendix J.

5.2 Second Endline

We conducted a follow-up survey with a random subset of 2,059 respondents approximately four months after the intervention to assess its long-term effects and potential network spillover impacts.[18] The extended time gap is particularly relevant, as India's 2024 general elections occurred between our two endlines – a period when political and partisan attitudes typically become more salient (Michelitch and Utych, 2018). The follow-up had three main objectives: (1) to assess whether discernment capacity persisted over time, (2) to evaluate if respondents could apply this skill to political stories—a new and unrelated domain, as the intervention deliberately avoided political topics due to our collaboration with government,[19] and (3) to examine spillover learning effects within respondents' networks. Given that students were encouraged to share classroom learnings with families, we interviewed one parent or guardian to evaluate whether treated students' families showed improved discernment abilities.

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

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

Remarkably, our findings indicate that participants in the treatment group continued to exhibit an improved ability to discern truth from falsehood (0.26 SD), as shown in Table 2. Moreover, treated respondents exhibited a significantly higher capacity to accurately assess the veracity of political stories (0.31 SD). This result is particularly striking given that the intervention exclusively targeted health-related content and did not address political claims. Notably, the political stories included in this survey were entirely novel and had not been evaluated during the initial endline assessment. Despite this, treated respondents demonstrated an enhanced ability to differentiate between true and false information on political topics. These findings underscore the broader implications of misinformation education: even when narrowly focused on a specific domain (such as health), educational interventions can yield transferable benefits across other domains. Further, these effects persist over time, highlighting the potential of such interventions to foster broad, enduring resilience against various forms of misinformation. We note that we observe very limited differences between the random follow-up sample that we recontacted versus those who eventually answered (including on outcomes from the first endline), implying that the persistence we observe likely generalizes to the whole sample.

Next, we find that parents/guardians of treated students were also significantly better at discerning true from false health information (0.27 SD), as demonstrated in Table 3. This result is particularly notable as it highlights the potential for "trickle-up" socialization, where children's learning influences their parents (Dahlgaard, 2018). It also suggests that sustained learning may generate valuable network spillover effects. One mechanism for this effect may have been the homework assignments and handouts given to students. Both treatment and control groups received written materials summarizing classroom lessons to take home (see Appendix B). Students worked on assignments at home and had physical copies of handouts and flyers that family members could potentially view or discuss with them. We view this finding as noteworthy, underscoring that educative interventions, even with vulnerable populations such as children, can have effects that spill over to other important members of networks, thereby adding to a literature that identifies change in adults that stem from children's behaviors (Carlos, 2021; McDevitt and Chaffee, 2002; Washington, 2008).

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

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

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

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

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

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

5.3 Mechanisms

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

Several factors suggest that our findings stem not just from recall of classroom content but also from substantive learning and application of new knowledge, and shifts in norms. Many outcome measures, such as willingness to correct friends, involve costly behaviors and intentions. In India, social norms of deference discourage confrontation, even when faced with misinformation (Malhotra and Pearce, 2022; Giles et al., 2003). The observed changes in behavioral intentions thus indicate shifts in norms, making social desirability or demand effects unlikely explanations; culturally, avoiding confrontation is often the socially desirable response (Savani, Morris, and Naidu, 2012). Additionally, the second endline survey included political stories—an entirely new topic not covered during the intervention or first endline. The findings here suggest respon-

dents are not merely recalling information but are learning, retaining, and applying principles across real-world contexts. Finally, our discernment measure included two stories discussed in the classroom and six new ones. To validate the learning mechanism empirically, we recalculated discernment by excluding the two classroom stories, analyzing only the six introduced at the endline (Appendix J.6). The results are still large and significant, underscoring that the observed effects stem from respondents' ability to learn and apply new information rather than mere recall.

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

6 Discussion and Conclusion

In this study, we evaluated the impact of a large-scale, classroom-based intervention aimed at combating misinformation, implemented among over 13,500 adolescents in Bihar, India. In collaboration with a state government agency, we developed a curriculum of sustained education against misinformation that spanned 4 months. ITT estimates showed significant improvements in students' critical thinking and application of classroom lessons to new areas. By the program's

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

end, treatment respondents demonstrated better discernment in evaluating and sharing information, shifted health preferences away from alternative medicine, and enhanced source credibility assessments. We also detected effects on behavioral measures among boys. These effects persisted among a sub-sample interviewed 4 months later. Importantly, follow-up surveys showed that students were able to accurately discern true from false political news, a topic not covered in the program, demonstrating the transferability of the skills learned. Finally, we found that parents/guardians of treated students were significantly better at discernment, indicating that such educational interventions can have spillover effects within social networks, with knowledge trickling upwards through socialization. Several of the outcomes we measure evaluate the acquisition of skills rather than mere recall, reducing the possibility that expressive responding or social desirability alone drove responses.

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

Despite these encouraging findings, we recognize several limitations of the study. The intervention was delivered as a bundled treatment with multiple components, making it challenging to look at treatment effect heterogeneity based on the content of classes. Using session-wise attendance to pinpoint effective elements proves difficult, as session topics are confounded with factors like peer effects. For instance, students attending the first class may form distinct social

networks, leading to downstream effects that are endogenous to the first class they attend. Additionally, each session's content was recapped in subsequent sessions, further complicating efforts to isolate the impact of specific topics. Our goal was to design a comprehensive, high-dosage intervention, addressing the limited success of many media and information literacy programs. Future extensions of this work could unbundle the curriculum to determine where the effects are most concentrated. Another limitation is the high cost and effort required to set up such an intensive and sustained intervention, raising the question of whether this dosage of treatment is necessary. Indeed, many studies use brief fact-checks or primes and nudges and find effects. To provide suggestive evidence of the required dosage to significantly affect outcomes, we computed subgroup ITTs by session attendance (Appendix Figure C.10), and found that attending one session had no significant effects, consistent with other research in this context showing that one-time treatments are insufficient (Badrinathan, 2021). Two sessions significantly improved accuracy, sharing discernment, and source discernment. However, shifting health preferences and attitudes toward counter-misinformation efforts required at least three sessions.[21]

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

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

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

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

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

Despite these limitations, our positive findings offer valuable insights for both academic research on misinformation and policy development. Following the 2016 surge in media literacy

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

initiatives, many were implemented without evidence of their causal effects. To our knowledge, this is the first randomized controlled trial testing the efficacy of such an intervention. The implications are broad: we believe policy-makers and researchers alike should prioritize sustained, iterative treatments, even if intensive and costly. In many settings, these may be the only viable solutions, especially where populations lack internet access, making platform-based solutions like fact-checking unfeasible. From a policy perspective, modules like ours could be integrated seamlessly into school curricula, particularly in contexts with high educational quality. To provide more insight into the applicability of such an intervention as policy, and recognizing that its success is a function of both teaching quality as well as curriculum, we investigated whether certain teacher demographics influenced outcomes. We found that female teachers delivered significantly better results across many measures (see Appendix E). Interestingly, despite significant religious polarization in north India and evidence of discrimination against religious and caste minorities, teacher religion and caste identity did not significantly affect outcomes.

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Bowles, Jeremy, Kevin Croke, Horacio Larreguy, Shelley Liu, and John Marshall. 2024. "Sustained exposure to fact-checks can innoculate citizens against misinformation in the Global South." *Forthcoming, American Political Science Review,*.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Ghosh, Arkadev, Prerna Kundu, Matt Lowe, and Gareth Nellis. 2024. "Creating cohesive communities: A youth camp experiment in india." *Working Paper*.

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

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

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

Graham, Matthew H, and Omer Yair. 2023. Less Partisan but No More Competent: Expressive Responding and Fact-Opinion Discernment. Technical report.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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