NBER WORKING PAPER SERIES

DEBIASING LAW ENFORCEMENT OFFICERS: EVIDENCE FROM AN EXPRESSIVE ARTS INTERVENTION IN INDIA

Sofia Amaral
Girija Borker
Nishith Prakash
María Micaela Sviatschi

Working Paper 35986
http://www.nber.org/papers/w35986

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2026

We thank Ritam Chaurey, Thomas Fujiwara, Dan Keniston, Deepak Saraswat and participants at seminars at the Indian Institute of Management Ahmedabad, Indian Institute of Management Bangalore, Indian Statistical Institute Delhi, Jawaharlal Nehru University, Johns Hopkins University, the LEAP Conference (2026), National Bureau of Economic Research Summer Institute, Economics of Crime (2026), Northeastern University, Northwestern University, University of Notre Dame, the NOVAFRICA Conference on Economic Development (2026), Princeton University, Sciences Po, and the Virtual Development Economics Series. We are deeply grateful to the following individuals and organizations whose support and guidance were indispensable to this project. From the Indian Administrative Service (IAS): Rahul Singh and Amir Subhani. From the Indian Police Service (IPS): K. Suhita Anupam, Kumar Ashish, Rajwinder Singh Bhatti, Manavjeet Dhillon, Amit Kumar Jain, Rakesh Kumar, Vinay Kumar, Ashok Mishra, Kantesh Kumar Mishra, Ranjit Mishra, Gupteshwar Pandey, Alok Raj, Har Kishore Rai, Rajyavardhan Sharma, Sanjiv Kumar Singhal, Praveen Vashishta, Anil Kishore Yadav, and Pramod Kumar Yadav. We also thank Kavita Arora, Durba Ghose (from Mittika), Karla Hoff, Zubair Idrisi, Jaya Iyer, and Vandana Kumari Pandey, and DAI Research and Advisory Services (Implementing Partner). We thank Dibyajyoti Basak, Shubhro Bhattacharya, Asmi Khushi, Neil Oommen, and Aditi Priya for their excellent research assistance, and Bhuvnesh Bajpai, Deepa Imley, Sitaram Mukherjee, Anup Roy, Manas Samanta, and all the field staff at DAI Research and Advisory Services for their invaluable contributions to data collection. This work was supported by the ifo Institute, International Growth Center, Fund for Innovation in Development (FID), J-PAL Crime and Violence Initiative, J-PAL Gender Initiative, The World Bank, and Princeton University. The views expressed here do not necessarily represent the views of our partners. The study was registered with the AEA RCT Registry (ID AEARCTR-0013270), and ethics approval was obtained from Princeton University and DAI Research and Advisory Services Private Limited. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Sofia Amaral, Girija Borker, Nishith Prakash, and María Micaela Sviatschi. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Debiasing Law Enforcement Officers: Evidence from an Expressive Arts Intervention in India
Sofia Amaral, Girija Borker, Nishith Prakash, and María Micaela Sviatschi
NBER Working Paper No. 35986
November 2026
JEL No. O10, P0

ABSTRACT

Frontline state agents exercise substantial discretion over citizens' access to public services, yet little is known about whether the attitudes and beliefs shaping this discretion can be changed among incumbent bureaucrats. Partnering with the Bihar Police - one of India's largest states that operates under comparatively severe policing capacity constraints, we conduct a cluster randomized controlled trial across 417 police stations and roughly 2,500 active duty senior field police officers to evaluate a three-day expressive arts intervention, grounded in the Theatre of the Oppressed, designed to change police responses to gender-based violence (GBV). Approximately 7 months after baseline, treated officers exhibit more favorable attitudes toward GBV, greater empathy, and stronger legal knowledge. These changes translate into observed policing behavior: in standardized victim audits conducted one year later, officers in treated stations are 15 percentage points less likely to blame victims, dismiss fewer complaints, and provide more referrals, while administrative records show a 41 percent reduction in registration delays and, over a longer horizon, increased GBV reporting by citizens. Exploiting variation generated by officer transfers, we find suggestive evidence that these effects persist when treated officers relocate and diffuse to untreated officers working alongside treated colleagues. The effects also extend to the workplace: junior female officers working with treated senior officers report lower harassment and anxiety. These findings show that the attitudes shaping bureaucratic discretion are malleable and that changing how incumbent state agents exercise authority can improve both citizen-facing service delivery and the broader workplace environment.

Sofia Amaral
The World Bank
s.amaral@worldbank.org

Girija Borker
The World Bank
gborker@worldbank.org

Nishith Prakash
Northeastern University
n.prakash@northeastern.edu

María Micaela Sviatschi
Princeton University
Department of Economics
and NBER
msviatschi@princeton.edu

1 Introduction

Improving the performance of frontline bureaucracies is central to building effective states (Besley and Persson, 2009; Page and Pande, 2018). Existing research has emphasized the role of personnel selection, incentives, and accountability in determining the quality of public services (e.g., Duflo et al., 2012; Das et al., 2016; Finan et al., 2017). However, even when bureaucrats operate under the same formal rules and incentives, frontline interactions often require substantial discretion, and differences in officials' attitudes, beliefs, and decision-making can shape the quality of services citizens receive. Whether these attributes can be changed among incumbent state agents and whether doing so improves their interactions with citizens, remains less well understood.

Law enforcement provides a particularly consequential setting in which to study this question. Police officers routinely make high-stakes decisions about whether to take complaints seriously, initiate formal procedures, and provide citizens with access to legal protections (Dube et al., 2025; Amaral et al., 2025). Moreover, unlike many other public services, policing has no close private substitute: the state retains a monopoly over the application of the law. Biases in the exercise of police discretion can therefore restrict access to justice and weaken citizens' willingness to engage with the state.

These concerns are especially salient for gender-based violence (GBV), which is characterized by stigma, fear of backlash, and forms of victimization that are frequently concealed from public scrutiny (Murvartian et al., 2024). In this case, dismissive responses, victim-blaming, and procedural barriers at police stations reduce reporting and perpetuate inequality in legal protection.[1] In response to these challenges, policymakers have turned to gender quotas which have increased reporting crimes against women (Miller and Segal, 2019; Sukhtankar et al., 2022; Amaral et al., 2021; Sviatschi and Trako, 2024; García-Hombrados et al., 2024; Asik et al., 2026; Bernhardt et al., 2026). Yet policing remains a male-dominated and highly hierarchical profession, in which senior officers exercise substantial discretion and tenure protections limit the scope for personnel reallocation.[2] These institutional features raise a central question: can the attitudes and behavior of

[1] Despite high prevalence of GBV, women's reporting rates remain extremely low, reflecting both social stigma and institutional failure. Prior work shows that police officers' attitudes toward GBV, such as beliefs about victim credibility or whether sexual harassment merits police action, strongly shape officers' behavior and affect crime incidence (Amaral et al., 2025)

[2] In India, several states have adopted gender quotas in law enforcement. However, female recruitment remains low, and quota thresholds are often not met (Tata Trusts, 2023). Qualitative evidence further suggests that female officers face workplace harassment and hostile working conditions, potentially discouraging women from entering or

existing state agents be reshaped, and does doing so improve the services citizens receive?

This paper studies whether reshaping GBV attitudes and skills of senior police officers through a novel expressive arts intervention can improve policing performance and the workplace environment in the presence of entrenched gender bias. To study this question, we partner with the Bihar Police in India, to implement a clustered randomized controlled trial covering 417 police stations across twelve districts and roughly 2,500 active duty senior field police officers serving 42 million citizens. Bihar is the poorest state in India, and provides a particularly relevant context: despite recent legislative improvements, gender violence remains highly prevalent, reporting rates are extremely low, and women remain underrepresented in decision-making roles within police stations.[3] Our intervention targets key decision-makers— Station House Officers, Inspectors, Sub-Inspectors, and Assistant Sub-Inspectors—who shape both citizen-facing behavior and internal workplace norms.[4]

Motivated by evidence that immersive and experiential learning can be more effective than traditional instruction in shifting beliefs, norms, and decision-making in high-discretion settings (Andries et al., 2024; Dube et al., 2025; Ashraf et al., 2024; Chaudhuri et al., 2025), we design a novel expressive arts-based training tailored to address police biases towards GBV in India. The intervention was informed by 18 months of qualitative fieldwork, including sustained observation in police stations and interviews with officers, GBV survivors, and civil society organizations. This fieldwork highlighted two features of the police organization that shaped the program design: substantial discretion in officers' responses to GBV and a highly hierarchical workplace in which senior officers influence both case handling and norms of appropriate conduct.

The intervention therefore targets senior active duty field officers and combines perspective-

remaining in the police force. Moreover, most quotas apply to lower-ranking positions within police stations, limiting women's representation in decision-making roles. Indeed, Bernhardt et al. (2026) show experimentally that assigning female constables to police stations increases GBV registration while, at the same time, male colleagues become more likely to express biased beliefs about female complainants. This tension points to a distinct challenge: improving representation may expand women's access to the state without changing—and may in some cases reinforce—the beliefs through which incumbent male officers exercise discretion.

[3] According to the Bihar Economic Survey 2025-26, Bihar's per capita remains the lowest in India with a per capita GDP at current prices of USD 925. Bihar's GDP is comparable to those of other low-income countries such as Sierra Leone, Uganda, or Gambia.

[4] Reforming the police has been a persistent challenge in Bihar. GBV remains widespread: 40 percent of women aged 18–49 report having experienced spousal physical or sexual violence, yet only 2 percent of survivors seek help from the police (NFHS-5, 2022, 2021). This gap persists despite the existence of formal legal protections against domestic violence and a 35 percent gender quota in police recruitment (Tata Trusts, 2023). Moreover, women remain underrepresented in decision-making roles within police stations, limiting their influence over frontline responses to GBV.

taking, experiential learning, and collective reflection. It draws on Freirean conscientization (Freire, 1970) and Boal's *Theatre of the Oppressed* (Boal, 2014), adapting these methods to policing and officers' interactions with GBV survivors. Through role play and simulated police-station scenarios, officers assume the perspectives of survivors and responding officers, confront competing interpretations of the same situation, and collectively reflect on appropriate responses. Rather than prescribing a single behavioral rule, the training exposes officers to context-rich scenarios in which they must exercise judgment, with the aim of improving how they interpret situations and use discretion. Guided reflection, peer feedback, and explicit commitments to station-level practice changes reinforce these experiences and create shared expectations about acceptable conduct (Bicchieri, 2016).

We hypothesize that the intervention operates through two reinforcing channels. First, it changes how officers interpret situations and exercise discretion by fostering perspective-taking and empathy (*individual attitudes channel*). Second, because the training is delivered collectively and publicly to senior officers from multiple stations, it creates shared experiences and expectations about appropriate conduct, potentially reshaping workplace norms and diffusing through the police hierarchy (*workplace channel*). We use plausibly exogenous officer transfers to explore the importance of these channels.[5]

Our analysis draws on a rich set of newly collected data that captures officer attitudes, behavior, and workplace dynamics using surveys, random victims' audit data, and administrative police records. We measure police officers' mindsets, attitudes, and technical skills related to GBV using detailed novel survey measures administered to senior male officers. We make use of self-reported measures, along with vignettes, thus, taking into account possible measurement concerns (Bursztyn et al., 2025). To measure real-world policing behavior, we implement a standardized victim audit in which trained decoys visit police stations to lodge complaints and systematically record officers' responses, and we complement these measures with administrative data on crime registration and handling drawn from approximately 500,000 official police records. To assess changes in the workplace environment and potential spillovers, we conducted confidential surveys of junior female constables at treated and control police stations. These officers were

[5]Separating these channels experimentally would require independently varying the individual perspective-taking component and the collective exposure of officers within stations. Such a factorial design was not feasible because the training was designed and delivered as a single integrated program rather than as separable modules. We therefore rely on plausibly exogenous officer transfers to distinguish patterns consistent with persistent individual-level change from those consistent with workplace norm diffusion.

not targeted by the intervention, allowing us to measure whether the intervention affected their experiences of harassment, anxiety, and perceptions of senior officers. In addition, we use data on officer transfers to examine the mechanisms underlying the intervention's effects and the extent to which those effects spillover to untreated officers. Together, these data allow us to link changes in officer attitudes to downstream changes in workplace culture and service delivery.

We find that immersive expressive-arts training significantly improves police attitudes and thus, performance, workplace environment, and interactions with survivors of GBV. In particular, we find that treated senior officers exhibit greater empathy, stronger beliefs in the credibility of survivors' complaints, and greater legal and procedural knowledge. Approximately one year after the intervention, standardized victim audits show that these changes are accompanied by substantial improvements in real-world police behavior: officers are less likely to dismiss cases or blame victims and more likely to provide appropriate guidance and referrals.

Consistent with these behavioral changes, administrative data show that officers in treated police stations register GBV cases more quickly following reported incidents, indicating improved responsiveness in formalizing complaints. We also find that treated officers devote more time to investigating and following up on GBV cases. Furthermore, after one year, we observe an increase in the overall volume of GBV First Information Reports (FIRs). Importantly, this increase emerges only after earlier improvements in officer attitudes, behavior, and responsiveness, a sequence consistent with improved service delivery increasing victims' willingness to seek help and engage with the justice system.

We also provide suggestive evidence on the mechanisms underlying these effects. We exploit officer transfers across police stations, which are frequent in the Indian bureaucracy (Iyer and Mani, 2012), to examine whether the effects persist with treated officers and diffuse to untreated colleagues.[6] Treated officers who move to control stations retain improvements in GBV attitudes, consistent with a persistent individual-level change that survives a change in workplace environment. At the same time, untreated officers who move into treated stations exhibit sizable improvements, consistent with exposure to treated peers and station-level norm diffusion. Taken together, these patterns suggest that the intervention operates through a combination of durable

[6]Transfers are not randomly assigned. They are common in the Indian bureaucracy and often occur for administrative reasons, but they may also reflect political considerations, disciplinary actions, officer requests, or responses to major incidents. We therefore interpret the transfer analysis as suggestive evidence on persistence and diffusion rather than as a source of exogenous variation.

individual change and changes in workplace norms.

The effects also extend beyond directly treated officers to the broader workplace environment. Eight months after the training, junior female officers stationed with treated senior officers report significantly lower levels of workplace harassment and anxiety, suggesting that changes in senior officers' behavior affect not only their interactions with GBV survivors but also how senior officers exercise authority within the police organization.

Several pieces of evidence make alternative explanations based on changes in formal incentives, sample composition, social desirability, or short-run monitoring less likely. First, the intervention did not alter officers' formal responsibilities, compensation, promotion prospects, or enforcement intensity, making explanations based on changes in formal incentives less likely. Second, results are robust to controlling for the share of female officers, and social desirability measures.[7] The fact that we also detect changes in performance using audit data, and among non-treated officers (females and non-treated transferred officers) further reassures that results are not driven by these measurement concerns. Third, we also find that treatment effects do not vary with officers' rank or hierarchical position, which would be expected if changes were primarily driven by senior officers rewarding visible compliance. Moreover, key outcomes are measured well after the training—approximately one year later in the standardized victim audits and over an even longer horizon in the administrative records. The audits capture behavior in settings where officers were unaware of being evaluated, while the administrative records are generated independently of the research team, reducing concerns about short-run signaling or experimenter demand effects. Together, these patterns support an interpretation in which the intervention operates through shifts in officer attitudes and decision-making rather than through mechanical channels.

This paper contributes to the literature on public service delivery (e.g., Finan et al., 2017; Rasul and Rogger, 2018) and frontline workers (Maynard-Moody and Musheno, 2003) by providing causal evidence that an intervention targeting attitudes, and decision-making among state agents can improve performance in a high-discretion institutional setting. Prior work on policing shows how officers' decision making is affected by their attitudes and prevalent social norms (Amaral et al., 2025), but there is limited evidence on whether interventions can generate persistent be-

[7] As we will explain, our novel measures include a mix of direct-questioning and vignettes – tools designed to mitigate concerns over desirability bias (Bursztyn et al., 2025).

havioral change among frontline bureaucrats or service providers – key agents in GBV response. We show that an immersive, experiential training targeting senior police officers significantly improves both attitudes and real-world behavior toward victims of GBV. Thereby establishing a direct link between bias reduction and service quality in a core state function. Unlike reforms that rely on changing personnel composition or creating parallel institutions, our intervention operates within the existing hierarchy, targeting decision-makers who shape both case outcomes and workplace norms. The results suggest that attitudes, norms and institutional culture are integral components of public sector performance. More broadly, by targeting senior officers within hierarchical police stations, we complement recent work on the role of public-sector managers in shaping the performance of government agencies (Fenizia, 2022). Our findings suggest that managers may influence not only organizational performance and worker effort, but also workplace norms, discretionary behavior, and the treatment of vulnerable citizens.

Second, we contribute to the experimental literature on police training (Dube et al., 2025; Bernhardt et al., 2026; Harris et al., 2022; Banerjee et al., 2021; Canales et al., 2025). Closely related, Dube et al. (2025) show that experiential training can induce a "cognitive pause" in high-stress policing environments and reduce use of force. We build on this work by studying a different source of discretionary failure: gender bias in officers' interactions with survivors of GBV. We show that experiential training can generate persistent changes in attitudes and service quality. We further combine survey measures with standardized victim audits and administrative records, allowing us to trace effects from individual attitudes to real-world police behavior as well as quantifying spillover effects on untreated officers.

Finally, we contribute to the literature on gender norms, perspective-taking, and experiential learning (e.g., Dhar et al., 2022; Sharma, 2024; Chaudhuri et al., 2025; Banerjee et al., 2019; Andries et al., 2024; Ashraf et al., 2024). This literature shows that narratives, curricula, and participatory interventions can shift attitudes. Most closely related, Chaudhuri et al. (2025) study Jana Sanskriti's community-based participatory theater in West Bengal and find changes in the acceptability of domestic violence and women's voice. We bring these insights into a markedly different institutional setting: a hierarchical state bureaucracy in which participants themselves exercise substantial authority over citizens. Rather than seeking to change norms among citizens, our intervention seeks to change how incumbent state agents interpret situations and exercise discretion. Our setting allows us to study whether experiential methods designed to challenge deeply rooted social norms can alter the behavior of state agents, whether these changes persist,

and whether they diffuse through the organizations in which they work.[8]

2 Background

2.1 Gender Based Violence and the Police in Bihar

GBV is pervasive in Bihar and remains severely under-reported. Findings from the National Family Health Survey show that 40 percent of women aged 18–49 have experienced spousal physical or sexual violence (NFHS-5, 2022), and more than 20,000 cases of violence against women were officially registered in 2022 alone (National Crime Records Bureau, 2023). Yet, official figures substantially understate the true scale: only 2 percent of women who experience physical violence approach the police (NFHS-5, 2021). The gap between prevalence and reporting reflects both the social stigma surrounding gender violence and institutional failures that deter women from seeking help, including the tendency of officers to question the validity of complaints, to promote reconciliation over formal registration, and to resist recording cases despite a legal obligation to do so (Sukhtankar et al., 2022). The importance of officer attitudes to GBV outcomes is not merely correlational. Experimental evidence from Hyderabad demonstrates that uniformed police patrols reduced severe forms of sexual harassment by 27 percent, and that officer attitudes toward sexual harassment are a key mechanism explaining this effect (Amaral et al., 2025). This finding implies that improvements in officer views towards GBV can independently improve outcomes for survivors, providing direct motivation for interventions that target the attitudinal roots of institutional failure.

Bihar has made significant institutional efforts to increase the representation of women in the police force. Following a landmark policy decision in 2013, the Chief Minister introduced a 35 percent reservation for women in police recruitment from the rank of constable to Deputy Superintendent of Police (Tata Trusts, 2023). By 2023, Bihar had the highest share of female police personnel among Indian states, with women comprising approximately 24 percent of total strength, nearly double the national share of 12 percent (Bureau of Police Research and Development, 2023).

These aggregate figures, however, obscure a pronounced concentration of female personnel

[8] A recent systematic review of gender-norm interventions emphasizes the limited evidence on adult men outside student populations, long-run effects, and behavioral outcomes beyond self-reported attitudes (Stewart et al., 2021).

at the lowest rank. According to our baseline data, women hold only 6 of 417 Station House Officer positions—the officer in charge of each police station and the primary decision-maker for case registration—none of the Inspector positions, and 230 of 1,779 Sub-Inspector positions (12.9 percent) across the twelve districts in our study; at the constable rank, by contrast, women comprise 27.3 percent of strength (see Table A1). Gender quotas have therefore not materially changed the composition of the tier that exercises the most consequential discretion over case registration, investigation, and the treatment of survivors. The officers who occupy that tier remain overwhelmingly male, and the institutional biases embedded in their attitudes and practices can shape women's access to justice.

2.2 Police Hierarchy and Discretion in India

The Bihar Police operates as a strictly hierarchical institution, much like police forces elsewhere in India and around the world. At the state level, the police force is headed by the Director General of Police (DGP). At the district level, Superintendents of Police (SPs) and Deputy Superintendents of Police (DSPs) oversee operational and administrative matters. The basic unit of policing is the police station, each headed by a Station House Officer (SHO), typically an Inspector or, in some cases, a Sub-Inspector. Sub-Inspectors (SIs) and Assistant Sub-Inspectors (ASIs) serve under the SHO's supervision and bear primary responsibility for registering First Information Reports (FIRs), conducting investigations, collecting evidence, and executing arrests. Below this tier, head constables and constables perform duties under close supervision and do not ordinarily hold decision-making authority, though they frequently serve as the first point of contact between citizens and the state. In cases involving GBV, a female constable is required to be present when a woman files a complaint, giving junior female officers routine exposure to how senior officers receive and handle such complaints.

Our intervention design was informed by eighteen months of qualitative fieldwork, including sustained observation across police stations and interviews with officers, survivors of GBV, and civil society organizations. This fieldwork highlighted two features of the organization that are particularly consequential for GBV cases. First, senior officers act as gatekeepers to the formal justice system, exercising substantial discretion over case registration, investigation, the treatment and referral of complainants. The formalization of a complaint as an FIR, the treatment of a survivor, and whether a case proceeds depend substantially on decisions made by these offi-

cers. Second, this discretion is exercised within a highly hierarchical workplace: senior officers supervise junior personnel, so their conduct can also shape station-level practices and norms. Our qualitative work further revealed persistent gender biases in the handling of GBV complaints and a widespread perception among officers and administrators that conventional classroom-based training rarely translated into changes in behavior.

These qualitative patterns are also captured in our baseline survey. Gender-biased views were widespread among senior male officers prior to the intervention. Fifty-three percent endorsed views attributing harassment or rape to women's failure to conform to socially acceptable behavior, while 46 percent placed the burden of proof on the victim in cases of violence against women. Such beliefs were not confined to a small subset of stations: in more than 90 percent of police stations, officers expressed views consistent with blaming women who travel alone for incidents such as molestation or chain-snatching. Moreover, consistent with Amaral et al. (2025), these patterns are particularly salient when officers evaluate GBV relative to non-GBV crimes, suggesting that gender bias may distort how officers assess these complaints rather than simply reflecting a general tendency toward skepticism or blame.

These institutional features motivated us to target SHOs, Inspectors, SIs, and ASIs. Rather than altering the composition of the police force or constraining officers' discretion through rigid rules, the intervention targets incumbent decision-makers and seeks to change how they interpret and respond to situations in which discretion must be exercised. Targeting senior officers also creates the potential for effects to extend beyond treated individuals through supervision and station-level norms.

3 Experimental Design

3.1 The Intervention: A Theater of the Oppressed Approach

Building on the qualitative evidence described in Section 2, we designed an expressive arts-based intervention to address the specific barriers that emerged in police responses to GBV. Our fieldwork pointed to three closely related gaps: persistent gender-biased interpretations of complaints, limited perspective-taking with respect to survivor experience and incomplete application of legal procedures. Importantly, officers' attitudes and technical responses were intertwined in practice:

how officers assessed the credibility and seriousness of a complaint shaped whether and how legal procedures were applied.

These observations motivated a departure from conventional classroom-based instruction. Motivated by evidence that immersive and experiential learning can shift attitudes, perspective-taking, and decision-making in ways that conventional information provision often does not (Andries et al., 2024; Ashraf et al., 2024; Canales et al., 2025; Chaudhuri et al., 2025; Dube et al., 2025). Rather than only providing information or prescribing a fixed behavioral rule, the intervention combines experiential components, like perspective-taking, role play, collective reflection, and scenario-based exercises, with guided instruction, practical exercises, and peer feedback. This enables officers to clearly see how changes in attitudes translate into appropriate case handling. The curriculum is based on Paulo Freire's pedagogy of *conscientização* (Freire, 2020), which holds that changing an oppressive social reality first requires the party enacting it to develop a critical, experiential awareness of that reality, rather than an abstract or externally imposed one. Augusto Boal's Theatre of the Oppressed (Boal, 2014) operationalizes this insight for group settings: it replaces the passive spectator with the *spect-actor*, a participant who enters a scene, interrupts it, and rehearses an alternative response before having to enact that response in real life. Applied to policing, this means officers do not discuss, in the abstract, how a GBV complaint should be handled; they physically occupy the role of the survivor, the responding officer, and the bystander, and revise a scene collectively until the group agrees it reflects appropriate conduct.

Through simulated police-station scenarios, officers alternate between the roles of survivors and responding officers, confront competing interpretations of the same interaction, and collectively reflect on appropriate responses. Rather than teaching a single rule of conduct, the training exposes officers to real life context-rich situations that they have experienced before in which they must exercise judgment, with the aim of improving how they interpret complaints and use discretion. See Appendix B for a detailed description of the intervention and the curriculum.

3.2 Curriculum and Delivery

To design the curriculum, we collaborated with Mittika, a Delhi-based NGO specializing in applied theatre and gender-focused training.[9] The research team was closely involved in curriculum development to ensure that the training reflected the institutional constraints identified during the qualitative fieldwork and remained aligned with the intervention's theory of change.

Training was delivered over three consecutive days in batches of approximately 15–30 senior officers drawn from multiple police stations within the same district, including Station House Officers (SHOs). To preserve operational capacity, only a subset of eligible officers from each station participated in any given batch. This structure allowed officers from different stations to interact while limiting disruption to routine police operations, an important consideration given that participants remained active field-duty officers responsible for day-to-day policing responsibilities.

The curriculum followed a structured progression from reflection on gender stereotypes and professional roles, to perspective-taking in GBV cases, and finally to legal and procedural practice. The first phase focused on officers' perceptions of their professional role and the sources of gendered judgments. Officers reflected on their vision of an "ideal police station," distinguished aspects of police performance within and outside their own locus of control, and participated in exercises designed to make implicit gender stereotypes explicit. For example, participants categorized traits and professional roles as "male" or "female" and subsequently discussed the extent to which these associations reflected social norms rather than job requirements. GBV-specific scenarios were introduced only after these initial exercises, allowing the curriculum to first establish a common framework for discussing professional responsibility and discretion.

The second phase introduced GBV-specific scenarios using role play and participatory theatre. Facilitators enacted realistic police-station interactions in which a complaint was handled incorrectly—for example, by dismissing a survivor or encouraging informal compromise rather than following appropriate procedures. Consistent with the Theatre of the Oppressed approach, participants could interrupt these scenes, identify problematic responses, and enter the scenario themselves to rehearse alternative actions. Officers subsequently alternated between the roles of

[9] Mittika's applied-theatre work spans India, Afghanistan, and Tibet. The curriculum also drew on continuous input from senior Indian Police Service (IPS) officers, lawyers, and civil society members to keep it aligned with operational realities on the ground. The lead trainer and the wider facilitator team, along with their credentials, are detailed in Appendix B.4.

survivors and responding officers while peers observed and provided feedback. These exercises combined perspective-taking with practice in interpreting context and exercising discretion.

The final phase linked these experiential components to concrete police procedures. Officers reviewed relevant GBV legislation and procedural requirements, completed scenario-based quizzes, and received a practical handbook summarizing key legal provisions, procedures, and referral pathways. Sessions covered practices such as private intake, non-judgmental communication, and appropriate referrals. Each officer concluded the training by identifying specific actions that could be implemented upon returning to the police station.

Interactive games reinforce this material throughout the three days: in a *Snakes and Ladders* exercise, officers advance or regress on a game board according to how they respond to hypothetical complaint-handling scenarios, making salient how small procedural shortcuts erode institutional credibility; in a *Web of Relations* exercise, participants physically map interdependencies within a station, illustrating how one officer's conduct shapes the norms of the workplace as a whole. The full session-by-session curriculum, including objectives and design rationale for every session across all three days, is reported in Appendix B.3.

The curriculum's design translates directly into its delivery. Facilitators were trained through a six-day residential Training of Trainers (ToT) in September 2022, led by Mittika together with the same lead trainer and a retired Director General of Police, and screened on their command of GBV content, delivery quality, energy, teamwork, and process flow; 14 of 19 candidate facilitators (74 percent) were retained for the main rollout following the ToT and a pilot conducted in a non-sample district (full names, recruitment criteria, and ToT structure are detailed in Appendix B.4).

Implementation proceeded district-by-district in coordination with a nodal officer designated by each district Superintendent of Police, to secure institutional buy-in while minimizing disruption to routine station operations. Across the study, take-up among invited officers averaged 84 percent, with average attendance of 2.61 sessions of the three core sessions; together with the follow-up summary session, participating officers received approximately 35 hours of curriculum exposure on average. Full implementation and monitoring protocols, including the standard operating procedure used to sequence the rollout across twelve districts and the attendance-tracking and fidelity-tracking instruments used throughout, are reported in Appendix B.5 and B.6; selected facilitator field notes and translated participant reflections are reported in Appendix B.7.

3.3 Conceptual Framework

The intervention was designed to affect police behavior through two potentially reinforcing channels: *individual attitudes* and *workplace* channel. Both channels are particularly relevant in policing, where officers operate under substantial discretion and appropriate responses often depend on the context of an interaction rather than on the mechanical application of a single behavioral rule.

First, the intervention may change how officers interpret situations and exercise discretion (*individual attitudes channel*). Perspective-taking, role reversal, and repeated exposure to realistic GBV scenarios were designed to increase empathy, reduce victim-blaming, and change how officers assess the credibility and needs of complainants. Rather than prescribing a uniform response, these exercises required officers to engage with alternative perspectives and rehearse decisions across different situations. If these changes become internalized, their effects should persist beyond the training environment and remain with officers even when they move to stations where their colleagues were not trained.

Second, the intervention may affect behavior through workplace norms and peer exposure (*workplace channel*). Training was conducted collectively among senior officers, who enacted scenarios, observed one another's responses, provided feedback, and collectively discussed appropriate conduct. These interactions may change officers' beliefs about how colleagues and supervisors expect GBV cases to be handled. The performative and interactive nature of the training may also create salient shared experiences that officers can recall and reinforce through subsequent workplace interactions, helping sustain common expectations about appropriate conduct over time. Because treated officers occupy supervisory positions within a hierarchical organization, these expectations may also diffuse to officers who were not directly trained. If this channel is important, we should observe changes among untreated officers exposed to treated colleagues and, potentially, in the workplace experiences of junior female officers.

While our experimental design does not independently randomize these components, we exploit administrative data on officer transfers across stations and survey data on untreated officers to provide suggestive evidence on whether treatment effects persist with directly treated officers and diffuse to untreated colleagues.

3.4 Sample Selection, Randomization, and Timeline

We identify the effects of the intervention using a clustered randomized controlled trial with a single treatment arm: police stations are randomly assigned to treatment or control, and officers within treatment stations are invited to the training described above, while officers in control stations receive no exposure to the curriculum, materials, or facilitators. The unit of randomization is the police station rather than the individual officer, since senior officers within the same station interact daily and a within-station comparison would otherwise be contaminated by exposure to treated peers; the unit of analysis is correspondingly either the station or the officer, depending on the outcome.

Sample selection. Our study takes place in 12 districts in Bihar: Bagaha, Bettiah, Bhojpur, Gopalganj, Motihari, Muzaffarpur, Nalanda, Patna, Saran, Sitamarhi, Siwan, and Vaishali (see Figure 1). This area, the number of stations and police officers required for the study, were selected based on consultations with the Additional Director General of Police (Weaker Section) and the Special Secretary to the Home Department. We approached the Superintendents of Police of each district and obtained permission letters to conduct the survey and intervention activities in their jurisdiction. The study covers 417 police stations and outposts across these districts, and Bihar has a total of 1,094 police stations and outposts (National Herald India, 2022), so our sample covers nearly 38% of the state total.[10] The officer sample consists of the senior officers posted in these stations at baseline: we completed the baseline survey for 3,526 police officers across the 417 stations.

Randomization. Of the 417 stations, 217 were randomly assigned to treatment and 200 to control, stratified by district, the number of SHOs, Inspectors, SIs, and ASIs posted at the station at baseline, and station-level characteristics from the Police Station Facility Survey (PSFS) we collected in January 2022. Within treatment stations, all male officers serving as Assistant Sub-Inspectors (ASIs), Sub-Inspectors (SIs), Inspectors, and Station House Officers (SHOs) were invited to attend via a letter from the Superintendent of Police (SP) and through phone calls from the district nodal officer. This recruitment strategy mirrors how officers are invited for other duties, such as election duty or additional training. Figure 1 maps the stations in our sample and their treatment assignment.

[10] Outposts serve as an extension of the main police station: they are typically located far from the main station and staffed by a few personnel responsible for maintaining law and order in their area.

Following randomization, we conduct balance tests using both officer-level and police-station–level baseline characteristics. Table 1 reports that treatment and control groups are statistically similar across the vast majority of observable characteristics.

Figure 1: Study sample and distribution of police stations.



Note: The left panel highlights the study districts involved in the study. The right panel plots the police stations in the study and its treatment assignment.

Between baseline and endline, we successfully re-survey 73% of the original sample. A total of 955 officers attrit from the study sample—487 from the control group and 468 from the treatment group. The primary reason for attrition is administrative transfers of officers to districts outside the study area, a common feature of policing in this context. A small number of officers left the sample due to suspension, retirement, or ill health. We show that across treatment and control stations, there is a small and economically meaningless difference in attrition in Table A2 but, the characteristics of officers being transferred do not differ across treatment arms. As an additional robustness check, we will compute Lee bounds (see Table A3) to assess the sensitivity of our main estimates to differential attrition (Kremer et al., 2009; Baird et al., 2011; Drexler et al., 2014; Fiala et al., 2022).

Table 1: Balance at Baseline

	Control Mean	Treatment Mean	Difference (2)−(1)	SE	P-value
	(1)	(2)	(3)	(4)	(5)
Panel A: Demographics					
Age	47.241	47.506	0.289	0.312	0.354
Years of service (force)	22.505	22.218	-0.250	0.384	0.515
Years of service (station)	1.257	1.135	-0.115**	0.046	0.013
Married	0.916	0.907	-0.009	0.011	0.405
Caste: SC	0.183	0.194	0.012	0.015	0.430
Caste: ST	0.029	0.033	0.002	0.007	0.751
Caste: OBC	0.404	0.369	-0.036*	0.019	0.065
Caste: General	0.347	0.368	0.022	0.020	0.278
Education: 10th	0.341	0.326	-0.017	0.017	0.315
Education: Plus 2	0.177	0.187	0.010	0.015	0.501
Education: Diploma	0.009	0.009	-0.000	0.004	0.959
Education: Some college	0.001	0.003	0.002	0.002	0.179
Education: B.A.	0.392	0.410	0.020	0.017	0.232
Education: Post Graduate	0.079	0.065	-0.014	0.010	0.173
Rank: ASI	0.482	0.477	-0.008	0.018	0.675
Rank: SI	0.370	0.379	0.012	0.020	0.563
Rank: PSI	0.059	0.055	-0.005	0.009	0.587
Rank: Inspector	0.001	0.003	0.002	0.002	0.242
Rank: SHO	0.088	0.086	-0.001	0.008	0.889
Panel B: Police Station Characteristics					
PSFS General Index (raw average)	1.755	1.819	0.072	0.061	0.240
PSFS Female Infrastructure Index (raw share)	0.474	0.472	0.007	0.031	0.817
PSFS M-F Segregation Index (raw average)	0.187	0.173	-0.014	0.015	0.349
Panel c: Baseline Outcome Indices					
GBV Skills Index (item average)	2.225	2.230	0.006	0.011	0.582
Attitudes Index (item average)	2.912	2.917	0.006	0.015	0.689
Technical Skills Index (item average)	0.643	0.642	0.000	0.007	0.947
Number of officers	1189	1382	2571		
Number of stations			417		

Notes: Standard errors clustered at the station level. Each row reports results from a separate OLS regression of the baseline characteristic on the treatment indicator, absorbing strata fixed effects. The difference column reports the coefficient on treatment with stars denoting significance of the p-value. Sample restricted to officers who completed both baseline and endline surveys. Raw outcome indices are equal-weighted averages of the original, direction-aligned components and are not standardized; their means therefore describe the score in its original units. The GBV Skills, Attitudes, and Technical Skills indices are calculated as simple averages of their constituent, direction-aligned items (item averages). Because these indices are not standardized, their means are reported in the original units of the underlying survey scales. Source: Baseline survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Timeline. Following the PSFS in January 2022, we completed the baseline survey of male officers in September 2022. Randomization was conducted for each district following the completion of its baseline. Training began in November 2022 and was rolled out on a staggered basis across the twelve districts. The endline survey was also conducted on a staggered basis starting in June 2023, following the completion of training in each district.

4 Data Collection and Outcome Measures

4.1 Data Collection

Our study relies on three main datasets. First, we use survey and decoy data to measure officer attitudes, technical knowledge, and complaint-handling behavior. Second, we collect administrative data to further measure police performance and GBV reporting. Third, we collect survey data on female constables to assess how changes in police officer attitudes translate into changes in workplace culture.[11]

Officers survey. The baseline officers' survey was conducted in our sample districts, following a pilot survey conducted in the district of Arwal in September 2022.[12] Participation in this survey was motivated as part of a professional task that the officers had to complete. Given that we had received permission from the respective SPs of the districts, the officers were comfortable sharing their responses with our survey team. The survey module was comprehensive, and it took, on average, 45 minutes to administer.

The survey instrument collected: (i) demographic and socio-economic characteristics of the officers, (ii) novel measures of victim-blaming, (iii) truthfulness perception of GBV reports, (iv) acceptance towards gender and GBV, (v) empathy, (vi) a scale to measure behaviors associated with externalization of responsibilities by the officers, (vii) technical and legal knowledge of the officers regarding GBV, (viii) well-being of officers (GAD and PHQ score), and (ix) a personal rigidity scale (Rehfisch, 1958).

[11] The study was registered with the AEA RCT Registry (ID AEARCTR-0013270) and received ethics approval from the Institutional Review Board of Princeton University, with additional approval from DAI Research and Advisory Services Pvt Ltd, which led data collection in the field.

[12] Arwal is a non-sample district that was chosen as a pilot district given its proximity to the base location of our study, i.e., Patna, and the fact that it is a relatively small district with a total of 12 police stations and outposts as of September 2022.

We also measure social desirability bias based on Crowne and Marlowe (1960). Finally, the instrument also included time-use related questions within the workplace. We compute indices from the different scales. The indices are described in Appendix C.

Decoy. To complement the self-reported officer survey and address concerns about social desirability bias and Hawthorne effects, we implement a standardized victim audit modeled on the design of Banerjee et al. (2021) and Das et al. (2016). Trained standardized victims (SVs) visit sample police stations and attempt to lodge a complaint under conditions designed to closely mimic those faced by an actual complainant. Responding officers are unaware that the visitor is a trained actor or that their response is being evaluated. The audit covers 183 stations across six districts (Bhojpur, Gopalganj, Muzaffarpur, Nalanda, Sitamarhi, and Siwan). Each station receives three visits over a pre-agreed three-month window. The specific case presented at each visit and the timing of each visit within the window are randomized by the research team, and visits are staggered across waves to limit cross-station communication before a visit occurs. Within an hour of each interaction, the SV completes a structured questionnaire recording the officer's response and the quality of the interaction, and a subset of the interview is voice-recorded. We conduct this audit approximately one year after the conclusion of the intervention in a district, allowing sufficient time for durable behavioral change to emerge. Table A4 shows that characteristics are balanced across groups.[13]

Administrative police records. We obtained all First Information Reports (FIRs) registered at the sample police stations from the State Crime Records Bureau of the Bihar Police. The data contain around 500,000 complaints. Because these records are generated independently of the research team as part of routine police record keeping, they provide an objective measure of case registration and processing that is not subject to biases from self-reporting by officers.

Workplace survey. To assess whether the intervention alters workplace culture for officers who were not themselves eligible for training, we survey female constables stationed at each sample police station. The survey is administered exclusively by female enumerators and collects demographic and socioeconomic characteristics together with the workplace-culture items analyzed in Section 4.2. We survey 1,810 female constables across 366 of the 417 sample stations, approximately eight months after the conclusion of the intervention in a district. The full survey

[13]Full detail on the sampling, training, and ethical safeguards for the standardized victim audit, including protocols to prevent fraudulent case registration and to protect the well-being of SVs, is reported in the Appendix D.

items are listed in Appendix C.3.3. Table A5 shows that the female constables in our survey are broadly representative of the larger population of female constables in the study area.

Additional data. Alongside the datasets above, we fielded four additional surveys to probe enumerator and facilitator effects and to validate instruments and balance.

Enumerators survey. To study potential enumerator effects, we conducted an enumerator survey before the start of the baseline survey in September 2022. The survey was administered after the enumerators were selected to be part of the survey team. It was self-completed on tablets and was voluntary. We surveyed the enumerators' team on socio-economic characteristics, including age, education, birthplace, number and gender of children, life satisfaction, gender norms, and personality traits following Di Maio and Fiala (2020).

Reflection survey. To understand immediate reactions of participants to the training, we collected a reflection survey in both treatment and control stations (median duration: 13 minutes). We use this information to evaluate the differential effects on officers' empathetic concern and perspective taking. In the treatment group, the surveys were conducted in the respective training venues upon the conclusion of the final day of training. Additionally, a set of training feedback-related questions were also asked in the reflection survey for the treatment group in order to capture an immediate self-reported measure of satisfaction with the training curriculum. The reflection surveys for the control group were administered at the police stations, and our team of enumerators visited the individual stations during the same month in which the training sessions were conducted for the respective districts.

Facilitators survey. To study potential effects of the individual differences in the characteristics of the facilitators, we conducted a facilitator survey before the start of our pilot training in September 2022. The survey was administered after the facilitators were selected to be part of the training team, following the Training of Trainers (ToT), which was conducted between September 12 and 17, 2022, and Pilot Training Sessions in Arwal, which were conducted between September 18 and 21, 2022. Participation in this survey was voluntary, and the surveys were self-completed on tablets or smartphones via a link that was sent to the facilitators.

We surveyed the team of facilitators on socio-economic characteristics, including age, education, birthplace, marital status, and number of children. We also asked specific questions related to their work experience and familiarity with the context of working with the Bihar Police. The survey instrument also had questions to capture the perceptions of the facilitators on the current

state of policing towards GBV crimes. Finally, the instrument included questions to capture the individual differences in personality traits.

Police Station Facility Survey. We collect information on station infrastructure and construct a police station facility index, defined as a simple count of the services available at each station. We use this index in the stratification of the randomization and to assess balance across treatment and control stations. The survey also measures the extent to which station infrastructure meets the needs of female officers.

4.2 Outcome Variables

Our outcomes are organized around the two skill domains the training targets, officers' attitudes toward GBV and their technical and legal knowledge, and are reported in standard deviation units. Primary outcomes are self-reported and are combined into a single GBV Skills Index; secondary outcomes capture real-world behavior and organizational spillovers that the officer survey does not observe. All indices follow Anderson (2008) and are constructed by inverse-covariance weighting of their standardized components (Appendix C.1); the construct and full item list for each index are given in Appendix C.

Primary Outcomes. The GBV Skills Index is the pre-specified primary outcome. It combines an Attitudes Index and a Technical Skills Index, reflecting the intervention's joint aim of reshaping how officers regard GBV complainants and strengthening their command of the relevant law and procedure. The Attitudes Index aggregates five component indices, for victim-blaming, empathy, acceptance towards GBV, externalizing of responsibilities, and skepticism about the truthfulness of complaints; the Technical Skills Index aggregates a GBV identification index and a legal knowledge index. We report the GBV Skills Index as our primary summary measure and analyze its components separately to interpret mechanisms. Each component index, the construct it captures, and its underlying survey items are defined in Appendix C; aggregating correlated items into indices also limits the number of hypotheses we test and improves power.

Secondary Outcomes. Secondary outcomes measure behavior and organizational spillovers that self-report cannot capture, drawn from the standardized victim audit, administrative FIR records, and the female constables' survey described in Section 4.

Real-world behavior. From the standardized victim audit we record the responding officer's

observed conduct toward an unannounced complainant, measured along the same constructs the officer survey captures by self-report, which lets us compare stated attitudes with behavior. Because conduct is observed rather than reported, these measures are less exposed to social desirability bias than the survey (audit items are listed in Appendix C.3.1). From administrative FIR records we construct two station-level outcomes: the delay between a reported incident and its formal registration, and the volume of GBV FIRs filed over time. The first measures officer responsiveness; the second, measured over a multi-year window, reflects citizens' willingness to report.

Workplace environment. To assess whether shifts in senior officers' attitudes reach officers who were not eligible for the training, we measure three outcomes among female constables: an anxiety score, reported harassment, and perceptions of senior male colleagues. Because constables did not participate in the intervention, we read these as evidence on organizational spillovers rather than as a direct target of the training.

5 Empirical Specification

To estimate the causal effect of the intervention on officer-level outcomes, we estimate the following ANCOVA specification:

$$Y_{is} = \beta_0 + \beta_1 T_s + \delta Y_{is}^0 + \gamma_r + \epsilon_{is}, \tag{1}$$

where Y_{is} is the endline GBV skills index for officer i in station s, T_s is an indicator equal to one if station s was assigned to treatment, and Y_{is}^0 is the baseline value of the respective outcome. The coefficient β_1 is the intent-to-treat effect of assigning a station to treatment. All specifications include stratification fixed effects, γ_r, which absorb district given the staggered implementation across districts. Standard errors are clustered at the station level. The overall GBV skills index combines two sub-indices, attitudes toward GBV and technical skills, and we also estimate effects on each separately. To better understand mechanisms, we analyze the components of each sub-index. The attitudes index comprises victim-blaming, empathy, views of GBV, externalizing police responsibilities, and truthfulness of complaints. The technical skills index comprises the ability to identify GBV and legal knowledge.

In robustness tests, we add to the main specification baseline officer controls (age, caste, and education) and station-level controls: the police station facility index defined above, the remoteness of the station, and the presence of conflict near the station. We also show that the results are robust to including enumerator fixed effects and, in a separate specification, facilitator fixed effects.

To examine whether the intervention improves workplace culture for junior female officers, we measure (i) workplace well-being, using an anxiety score, (ii) perceptions of male officers, and (iii) the incidence of harassment.

We pre-specify heterogeneity analyses by baseline levels of the primary outcomes, strata, urban location, and transfer status. In particular, we exploit officer transfers across police stations to shed light on the mechanisms underlying the treatment effects. In our setting, transfers are generally driven by administrative and law-and-order considerations, including responses to major criminal cases, episodes of civil unrest, and election-related policing needs. Importantly for our analysis, treatment assignment does not predict whether an officer subsequently moves across treatment arms. We examine this directly in Table A6, where we find no relationship between treatment assignment and cross-arm movement. Table A7 further examines whether baseline officer characteristics predict such moves. Cross-arm movement is related to officers' tenure at their baseline station, but is not systematically associated with baseline outcome indices.

We use the resulting changes in officers' workplace exposure to analyze the mechanisms underlying the treatment effects. First, we examine whether treated officers who were transferred to control stations retain improvements after leaving the original treatment environment. Persistence among these officers would be consistent with an *individual attitudes channel*, whereby the intervention generates individual-level changes that travel with the officer. Second, we examine untreated officers who were transferred into treated stations. Improvements among these officers, despite not having received the training themselves, would be consistent with a *workplace channel*, whereby exposure to treated colleagues and the station environment diffuses changes in expected and accepted conduct.

6 Results

In this section, we present four main findings. First, using a short reflection survey administered on the final day of the training, we show that the intervention increased officers' empathy and perspective-taking, the first-stage channel through which the Theatre of the Oppressed approach is meant to reshape attitudes. Second, we show that the expressive-arts intervention improved officers' attitudes toward GBV and strengthened their technical and procedural skills. Third, we demonstrate that these changes translated into meaningful improvements in police performance. Evidence from standardized victim audits indicates that treated officers responded more appropriately to survivors, exhibiting less victim-blaming, fewer dismissive responses, and greater use of referrals and guidance. Consistent with these behavioral changes, administrative FIR records show that treated police stations registered GBV complaints more quickly following reported incidents, indicating greater responsiveness in formalizing cases. Over a longer horizon, we also observe an increase in GBV reporting, suggesting that improved police responses may have increased victims' willingness to engage with the justice system.

Fourth, we provide suggestive evidence on the mechanisms underlying these effects and their diffusion within police organizations. We find that improvements persist among treated officers who transfer to control stations, consistent with a persistent *individual attitudes channel* in which changes in attitudes and decision-making travel with the officer beyond the original treatment environment. At the same time, untreated officers who transfer into treated stations exhibit improvements in GBV-related skills and attitudes, consistent with a *workplace channel* through which norms and practices diffuse via peer exposure and day-to-day interactions. These effects extend beyond directly treated officers. Junior female constables working in treated stations report lower levels of harassment and anxiety, suggesting that changes among senior officers shape not only interactions with GBV survivors but also the broader workplace environment within police stations.

6.1 First Stage: Empathy and Perspective Taking

Using a short reflection survey administered on the final day of the training, we examine whether the experiential learning component increased empathy and perspective-taking, two central channels through which the Theatre of the Oppressed approach seeks to reshape attitudes and norms.

We construct a standardized index from validated items capturing empathetic concern (e.g., feeling protective when someone is taken advantage of, feeling touched by others' misfortune, describing oneself as soft-hearted) and perspective-taking (e.g., looking at both sides of a disagreement, imagining another person's viewpoint before criticizing them). The index also incorporates reverse-coded items reflecting emotional detachment or resistance to considering others' perspectives.[14]

Table 2 shows that officers in treated stations exhibit statistically significant increases in this index. These effects reflect a greater willingness to understand others' emotional states and stronger affective responses to unfair treatment. The magnitude is comparable to effects in the perspective-taking literature, where brief interventions shift targeted attitudes by roughly 0.04 to 0.16 standard deviations (Broockman and Kalla, 2016; Kalla and Broockman, 2020), while intensive multi-week curricula at schools move measured perspective-taking substantially more (Alan et al., 2021). We next examine whether these immediate shifts translated into durable improvements in officers' GBV-related attitudes and technical skills.

Table 2: Expressive Arts Training: Immediate Effects on Officers' Reflection Index

	Reflection Index	
	(1)	(2)
Expressive Arts	0.090**	0.085**
	(0.043)	(0.043)
Control mean	0.000	0.000
Strata FE	Yes	Yes
Station controls	No	Yes
Number of Officers	2830	2830

Notes: Standard errors are clustered at the police station level in parentheses. The dependent variable is a standardized reflection index constructed from 14 items (5 items reverse-coded). All specifications include strata fixed effects. Column (2) adds station controls (general infrastructure, female infrastructure, and male-female segregation indices). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[14]See Appendix C.2.7 for the construction of the Reflection Index. In particular, we construct a single standardized Reflection Index that combines the 14 items of the Empathic Concern and Perspective-Taking subscales of the Interpersonal Reactivity Index (Davis et al., 1980). The items capture respondents' ability to place themselves in another person's position, understand and share others' emotions, and appreciate alternative viewpoints. Higher values indicate greater empathic concern and perspective-taking.

6.2 Male Officers' Gender-Based Violence Attitudes and Technical Skills

Table 3 presents the effects of the intervention on an index of GBV skills which combines an attitudes index and a technical skills index. We find that the expressive arts intervention significantly improves officers' GBV skills and attitudes by 0.082 standard deviations and their technical skills by 0.085 standard deviations relative to the control group. All specifications control for the baseline value of the outcome and include strata, district, and officer-strength fixed effects, with standard errors clustered at the station level. These results are robust to social desirability, controls, enumerator fixed effects (See Table A8 and Table A9).

Tables A10 and A11 in the Appendix decompose the two indices into their components. These component-level estimates are exploratory and less precisely estimated than the index-level tests, since aggregating correlated measures into indices improves statistical power (Anderson, 2008). The point estimates are positive for every component, but the statistically significant improvements are concentrated in empathy and the truthfulness of complaints within the attitudes domain. These effects are meaningful in context. The gender attitudes of adult men have proven difficult to shift durably: intensive exposure can move them in the short run, but the gains often fade once the intervention ends (Dahl et al., 2021), and the largest prior evaluation of gender-targeted police reform in India finds no shift in male officers' gender attitudes at all (Sukhtankar et al., 2022). Against that backdrop, the improvement we find is substantive: it moves both officers' attitudes about GBV and their technical handling of cases, and it survives correction for multiple testing.[15]

[15]The magnitude of these effects is smaller than those documented in interventions targeting gender attitudes among adolescents in India (Dhar et al., 2022). The comparison needs to be considered keeping in mind key differences wherein, our intervention targets adult, relatively senior male police officers, a population whose gender attitudes may be more established than those of adolescents.

Table 3: Effects of Expressive Arts Training on Officers' GBV Attitudes and Skills

	GBV Skills Index	Attitudes Index	Technical Skills Index
	(1)	(2)	(3)
Expressive Arts	0.082**	0.082**	0.085**
	(0.040)	(0.042)	(0.043)
Control mean	0.000	0.000	0.000
Strata FE	Yes	Yes	Yes
Baseline control	Yes	Yes	Yes
Number of clusters	417	417	417
Number of officers	2571	2571	2571
Adjusted p-values			
Unadjusted p-value		0.042	0.048
Romano-Wolf		0.006	0.006
Bonferroni		0.083	0.096

Notes: Standard errors clustered at the station level (in parentheses). Regressions include only officers who completed both baseline and endline surveys, including those who switched between control and treatment stations. All regressions include district and officer strength fixed effects and the baseline value of the outcome. All indices created using inverse covariance weighting (swindex) and normalized to the control-group mean. Control means in this table are computed in the regression estimation sample, so they can differ slightly from zero when the estimation sample differs from the normalization sample. The GBV Skills Index is excluded from the correction family as it is a composite of the Attitudes and Technical Skills indices. Romano-Wolf p-values control the familywise error rate using 1,000 bootstrap replications (seed 12345). Bonferroni p-values multiply the unadjusted p-value by $K = 2$. Source: Baseline and endline surveys; police station facility survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

In the Appendix, we also examine whether treatment effects vary with officers' baseline characteristics, including their attitudes, rank, and education. We find little evidence that the effects of the intervention are concentrated among particular types of officers. Treatment effects do not vary systematically across these dimensions. More flexibly, applying the heterogeneity approach of Chernozhukov et al. (2018) using a rich set of baseline officer characteristics yields a similar pattern (see Figures A1 and A2). While predicted treatment effects vary across officers, differences between those with the highest and lowest predicted effects are imprecisely estimated. We also examine heterogeneity across station characteristics, including the share of female constables and whether the station is located in a rural area, and again find little evidence of systematic differences in treatment effects.

6.3 Male Officers' Performance in Gender-Based Violence Cases

To assess whether the observed changes in attitudes translated into changes in real-world policing behavior, we analyze outcomes from two main data: i) standardized decoy visits conducted approximately one year after the intervention, and; ii) administrative records of individual-level crime record fillings over two and half years since the intervention. The intervention may improve performance not because officers mechanically follow new rules, but because the training changes how officers interpret ambiguous situations, evaluate complainant credibility, and allocate effort in high-discretion settings.

Evidence from standardized decoy visits. Table 4 reports estimates of the training's effect on observed officer conduct during standardized decoy visits conducted approximately one year after the intervention. Treated officers are significantly less likely to dismiss a complaint as a waste of time (column 1), to assert that such cases are false (column 2), to blame the complainant (column 3), or to ask irrelevant, insinuating questions of the "what were you wearing" type (column 6). They are also more likely to refer complainants to a Women's Police Station (column 4) and to explain available options and next steps (column 5). The largest effects fall on the dimensions that reflect the most deeply rooted gender biases: victim-blaming declines by 15 percentage points against a control mean of 0.72, and unnecessary questioning by 14 percentage points against a control mean of 0.63, reductions of just over a fifth relative to the control group. These are the margins through which officers exercise discretion when deciding whether a complaint is credible, merits attention, and should be advanced through the formal justice system. Directing a complainant toward a Women's Police Station rather than turning her away informally moves a case toward formal registration, the gatekeeping step that governs access to justice (Sukhtankar et al., 2022).

These findings are informative because they capture observed behavior rather than self-reported attitudes. Officers were unaware that the interactions formed part of an evaluation, so the estimates record how they respond in realistic complaint-handling situations rather than in a survey. The results therefore provide direct evidence that the intervention changed not only what officers report believing about GBV, but also how they act when a potential complainant is in front of them. Measured approximately one year after training, the effects indicate persistent changes in frontline case handling rather than short-run compliance, and they are consistent with a shift in the exercise of bureaucratic discretion that reduces behaviors discouraging reporting while in-

creasing the guidance and referrals that support access to justice.[16]

Table 4: Effects of Expressive Arts Training on Police Responses to GBV Decoy Victims

	Waste of time	Cases false	Blames victim	Referred to WPS	Explained alternative	Asks unnecessary questions
	(1)	(2)	(3)	(4)	(5)	(6)
Expressive Arts	-0.07***	-0.08***	-0.15***	0.07***	0.08*	-0.14***
	(0.02)	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)
Control mean	0.06	0.15	0.72	0.09	0.21	0.63
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	183	183	183	183	183	183
Number of decoy visits	549	549	549	549	549	549

Notes: Standard errors clustered at the station level (in parentheses). The unit of observation is a decoy visit. Regressions include district and strata fixed effects. Control means are computed in the regression estimation sample. Source: Decoy survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Evidence from administrative records. Next, we also corroborate the decoy findings using administrative data, which allow for a granular examination of the temporal dynamics of police responsiveness. Event study estimates in Figure 2 in panel (a) indicate a sharp improvement in procedural efficiency following the expressive arts intervention - as measured by the number of days taken to file a crime report since its incident. The combined effect post-intervention shows the latency between a reported GBV incident and its formal registration decreased by 41%.

This marked reduction in registration lag suggests a fundamental shift in the "gatekeeping" behavior of male officers. In the policing of all crimes, including GBV, the registration of a FIR represents the critical juncture of state recognition and the foundation of all legal proceedings. The observed reduction in delays implies a decline in officer hesitation and a move away from informal mediation toward a more survivor-centered responsiveness.

To further understand these behavioral changes, we conduct exploratory analysis using administrative records on police officers' time use and task assignments. We find that treated officers devote more time to investigating GBV cases and to follow-up activities with victims (see

[16]Results are robust to controlling for the pre-treatment volume of GBV FIRs at the police-station level, suggesting that differences in baseline GBV caseloads do not account for the observed improvements in officers' responses to standardized complainants (see Table A12 in the Appendix).

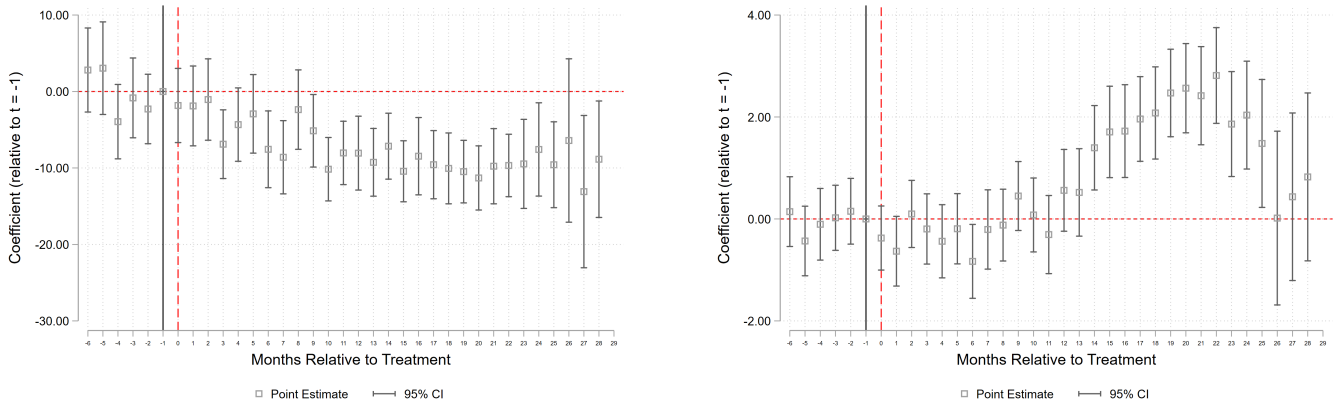
Table A13). In particular, treated officers report spending approximately four additional minutes investigating crimes against women and two additional minutes on follow-up, relative to control means of approximately six and four minutes, respectively. We find no statistically significant change in time spent assisting victims.[17] These patterns suggest that the intervention shifted officers' effort toward the discretionary margins of investigation and case follow-up, consistent with the observed reduction in FIR registration delays.

We also find gains on the demand side. Consistent with improvements in overall GBV skills, and job performance, we find that in the long-run, GBV victims' adjust their reporting behavior. In panel (b) of Figure 2 we show a gradual increase in the overall number of FIR filled. This results suggests that by signaling of a safer reporting environment to the public, citizens - mostly women - are more likely to report GBV to the police.

The persistence and strengthening of these effects approximately one year after the intervention are consistent with the emergence of new station-level norms and survivor-centered practices among police officers. By changing both individual attitudes and the workplace environment in which officers operate, the intervention appears to have increased victims' willingness to seek justice. Taken together, these findings suggest that immersive behavioral interventions can generate durable improvements in the quality of justice delivery for female survivors of violence.

[17]The average effect on time devoted to GBV cases is modest in the full sample, in part because the measure contains a large mass of zeros: officers were asked about time spent on GBV regardless of whether they were currently handling a GBV case. To provide a sense of the magnitude among officers, we also examine time use among those reporting an active GBV case. In this subsample, treated officers spend approximately 20 additional minutes on GBV-related work, relative to a mean of approximate 100 minutes.

Figure 2: Treatment Effects on Job Performance and GBV Victim's Reporting



(a) Effect on days to file a crime.

(b) Effect on crime reports.

Notes: Figures display event-study estimates. The coefficients represent post-intervention effects. All regressions include strata and district fixed effects. Standard errors are clustered at the police-station level. Confidence intervals at the 95 percent level.

6.4 Mechanisms: How the Training's Effects Persist and Spread

Two questions bear on how the intervention produced the changes documented above. The first is whether it generated persistent individual-level change in officers' attitudes and decision-making, or whether its effects depend on continued exposure to treated peers. The second is whether training senior officers reshaped the station as a workplace, reaching colleagues who were never trained. We take these in turn, exploiting post-intervention officer transfers for the first and the experiences of untrained junior women and senior untreated officers for the second.

6.4.1 Persistence Across Transfers

It is important to understand if the intervention generated persistent individual-level changes in officers' attitudes and decision-making, or if the effects also depend on continued exposure to treated peers and the emergence of new norms within police stations. To shed light on these channels, we exploit officer transfers across police stations following the intervention. Officer transfers after the intervention separate two groups: trained officers who moved to stations where no other colleague had been trained, and untrained officers who moved into treated stations. The comparison distinguishes change that officers internalize and carry with them from behavior that depends on the treated environment.

Table 5 reports the treatment effects separately for officers who remained in treated stations throughout (treatment stayers), officers initially assigned to treated stations who transferred to control stations (leavers), and untrained officers who transferred into treated stations (newcomers), with non-moving control officers as the omitted category. Two main findings emerge. First, treated officers who are transferred to new stations where no other peer received the training continue to exhibit significant improvements in GBV skills and attitudes. Importantly, these effects are statistically indistinguishable from those of treated officers who remained in treated stations. This pattern is consistent with the individual-level channel, suggesting that the training altered officers' internal attitudes and decision-making rather than merely changing behavior in response to local peer environments.

Second, untreated officers who transfer into treated stations also exhibit sizable improvements, particularly in overall GBV skills and technical skills, with weaker evidence on attitudes. These patterns are consistent with a workplace diffusion channel: officers who were not directly treated nonetheless improve when exposed to treated peers, potentially through learning, imitation, or shifts in perceived expectations within stations.

Taken together, the transfer-based heterogeneity analysis suggests that the intervention operates through a combination of persistent individual-level changes and station-level norm diffusion. To further examine the role of workplace environment, we next estimate spillover effects on junior women constables who were not trained.

Table 5: Differential Treatment Effects by Transfer Status

	GBV Skills Index	Attitudes Index	Technical Skills Index
	(1)	(2)	(3)
Treatment stayer	0.132***	0.103**	0.156***
	(0.047)	(0.051)	(0.051)
Leavers	0.129*	0.118	0.070
	(0.075)	(0.081)	(0.086)
Newcomers	0.157**	0.096	0.170**
	(0.075)	(0.080)	(0.082)
p-value: T Stayer vs Leavers	0.959	0.859	0.298
p-value: T Stayer vs Newcomers	0.038	0.227	0.039
Control mean	0.000	0.000	0.000
Fixed effects	Yes	Yes	Yes
Baseline outcome	Yes	Yes	Yes
Officer controls	Yes	Yes	Yes
Number of clusters	415	415	415
Number of officers	2262	2262	2262

Notes: Standard errors clustered at the station level (in parentheses). Regressions include only officers who completed both baseline and endline surveys, and examine heterogeneity by transfer status. The sample here contains only officers with no movement at all, or officers who moved between arms. The omitted category is officers in control stations at baseline who did not move stations at all between baseline and endline. Treatment stayers are officers in treatment stations who did not move stations at all between baseline and endline. Leavers are officers who were initially assigned to treatment stations but transferred to control stations between baseline and endline. Newcomers are officers who were initially assigned to control stations but transferred to treatment stations between baseline and endline. The bottom rows report p-values testing whether these effects differ significantly. All regressions include district and officer strength (at the station level) fixed effects, the baseline value of the outcome, and officer controls (age, years of service, marital status, caste, education, and rank). Source: Baseline and endline surveys; police station facility survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

6.4.2 Organizational Spillovers and Workplace Culture

If the intervention altered station-level norms or supervisory practices, its effects may extend beyond treated officers to colleagues who did not participate in the intervention. We examine this possibility using outcomes for female constables, none of whom received the intervention directly. Whether training programs can generate broader changes in organizational culture remains far from established, particularly in public-sector settings.[18] Table 6 shows a test of whether the intervention affected the workplace environment experienced by non-treated officers.

We find that eight months after the intervention, female officers in treated stations report significantly improved workplace environments, with the overall workplace anxiety index declining

[18]The closest evidence, from Azulai et al. (2020), finds that a training module for mid-level bureaucrats in Ghana's civil service improved organizational culture six to eighteen months later.

Table 6: Effects of Expressive Arts Training on Female Officers' Workplace Environment

Panel A			
	Anxiety Score	Perceptions Towards Male Officers	Reported Harassment
	(1)	(2)	(3)
Expressive Arts	-0.39^{**}	0.09	-0.06^{*}
	(0.15)	(0.06)	(0.04)
R^2	0.050	0.027	0.013
Control mean	2.53	0.00	0.00
Strata FE	Yes	Yes	Yes
Number of clusters	366	366	366
Number of officers	1810	1810	1803
Panel B			
	Anxiety Score	Perceptions Towards Male Officers	Reported Harassment
	(1)	(2)	(3)
Share Trained Officers	-0.42^{*}	0.18^{**}	-0.07
	(0.22)	(0.09)	(0.05)
R^2	0.048	0.029	0.013
Strata FE	Yes	Yes	Yes
Number of clusters	366	366	366
Number of officers	1810	1810	1803

Notes: Standard errors clustered at the station level (in parentheses). All regressions include district and officer-strength fixed effects. All indices are constructed using inverse covariance weighting and normalized so that the control-group mean equals zero. Treatment is an indicator equal to one if the female police officer surveyed at endline worked in a police station assigned to the treatment group, and zero otherwise. *Share Trained Officers* is the proportion of senior police officers in a station who received the treatment. On average, 38.43% of senior officers across all police stations were trained. Source: Female Constables' Survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

by 0.39 standard deviations, indicating more favorable conditions. Reports of harassment also decrease by 0.06, while perception towards male officers measures show no statistically significant change. Importantly, stations with a higher share of trained officers exhibit additional improvements: greater training saturation is associated with lower anxiety, and more positive perceptions of male colleagues. These findings suggest that the intervention produced institutional spillovers that improved organizational climate for women, consistent with a mechanism in which training top officers shifts norms within stations rather than affecting only directly treated officers. In particular, the fact that outcomes for female constables improve despite not being direct recipients of the intervention points to a station-level channel operating through supervisory practices and peer interactions.[19]

[19]Given the emotionally intensive nature of the intervention, we also examine whether the training affected officers'

This spillover is economically consequential. Because workplace harassment and violence are themselves drivers of women leaving their jobs (Adams-Prassl et al., 2024) and of the occupational segregation and pay gaps that follow (Folke and Rickne, 2022). A reduction in anxiety and harassment among junior women is not merely a marker of improved culture but a plausible channel for retaining women in a heavily male-dominated institution, fostering a more gender-equal police force both in the immediate term but also in terms of the pipeline of officers.

7 Cost and Benchmarking

The training curriculum and its delivery were funded by the Fund for Innovation in Development (FID), separately from the accompanying research and evaluation activities (officer surveys, standardized victim audits, and administrative data collection), which were funded through the J-PAL Crime and Violence Initiative and the J-PAL Gender Initiative.[20] The cost of delivering the intervention was EUR 136,560 (USD 151,600), approximately USD 108 per officer trained and USD 699 per treated station. The 217 treated stations serve an estimated 21.8 million residents; spread across that population, the one-time delivery cost is under one US cent per resident. The training is a one-time investment in officers, but its potential benefits extend to the much larger population these officers serve.

At USD 108 per officer and an estimated effect of 0.082 standard deviations on the GBV attitudes index, the implied cost is approximately USD 1,320 per standard deviation of attitude change. This is a cost-effectiveness metric rather than a welfare metric: it relates program cost directly to measured impact without requiring a monetary valuation of changes in GBV outcomes. If these effects persist, this one-time cost is more appropriately expressed on an annualized basis: USD 108 per officer if the effects last only one year, falling to as little as USD 22 per officer per year if they persist for five. Appendix E reports these calculations under alternative persistence assumptions.

To our knowledge, no published study reports both the cost and the impact of a police-focused

psychological well-being. We find no detectable effects on either depression (PHQ-9) or anxiety (GAD-7), suggesting that improvements in GBV-related attitudes and behavior were not accompanied by deterioration in these dimensions of officers' mental health.

[20] See Appendix E for the full cost breakdown by component and category, cost shares, and the underlying calculations.

gender-based-violence training program specifically: the closest police intervention, the Women's Help Desk reform evaluated by Sukhtankar et al. (2022), is a bundled reform rather than a stand-alone training and does not report costs on a comparable basis. This limits the scope for benchmarking on a common cost basis. Our intervention produces effects at three levels: officers' attitudes and technical skills, their case-handling behavior in standardized audits, and administrative case registration. No single external study speaks to all three, so we benchmark each against the closest available evidence (Appendix E reports the full comparison).

On cost, the program is inexpensive in absolute terms, approximately USD 108 per officer for a three-day training with follow-up and monitoring. The closest comparator that is itself a behavioral training for serving police and reports both cost and field impact, the 16-hour Sit-D program in Chicago, which cut uses of force and discretionary arrests by about 23 percent, cost USD 807 to 864 per officer trained (Dube et al., 2025); we do not read the gap as a like-for-like efficiency comparison, since cross-country training costs are dominated by local wage levels. On attitudes, the training's effect on officers' GBV attitudes was on the order of half the standardized effect of a 2.5-year, 28-session gender-attitudes curriculum for adolescents in India, a population generally taken to be more malleable and reached over far more contact time (Dhar et al., 2022); because the two indices measure different attitudes, this comparison is only approximate. On the behavioral and administrative outcomes of most direct policy relevance, the closest comparator is again Sukhtankar et al. (2022), which raised the number of women's cases formally registered. Our training moved that same registration margin, raising GBV reporting over the longer run, and additionally reduced the delay between a reported incident and its registration by 41 percent, a margin that study did not examine.

Beyond cost, the intervention's significance lies in where its measured effects sit in the broader chain from policing to women's welfare. The attitudes, case-handling, and registration outcomes we document are the proximate margins through which frontline policing shapes whether women seek and obtain redress. More effective redress can, in turn, alter women's perceived risks from participating in public life and reduce the need for costly avoidance responses. A companion police experiment in Hyderabad makes this downstream link concrete: where uniformed patrols reduced severe public-space sexual harassment, women became roughly 30 percent less likely to move elsewhere in response to it, easing constraints on physical mobility that are linked to women's schooling (Borker, ming) and labor force participation. Such changes can affect labor supply on both the extensive margin—whether women work at all—and the intensive mar-

gin—where, when, and how much they work. In Hyderabad, police visibility and officers' attitudes toward harassment were central to the effect (Amaral et al., 2025), the same attitudinal margin targeted by our intervention. A training that improves these attitudes and the quality of redress at modest cost may therefore have consequences that extend beyond police performance to women's willingness to take risks, participate in public spaces, and pursue economic opportunities, even where we do not attach a dollar value to these downstream benefits.

8 Conclusion

Public institutions frequently fail citizens not only because of limited resources or weak incentives, but also because of the beliefs and biases of the frontline agents who deliver the state's services. Policing makes this especially consequential. Officers exercise substantial discretion over which complaints are registered, how survivors are treated, and whether the law is applied consistently. Unlike most public services, policing has no private substitute. When this discretion is shaped by entrenched gender norms, access to justice becomes unequal precisely for those who depend on the state most.

This paper asks whether the performance of an existing, hierarchical, and male-dominated police force can be improved by reshaping the attitudes and skills of the officers who already staff it. Partnering with the Government of Bihar, we implemented a clustered randomized evaluation of a three-day expressive arts curriculum, grounded in the Theatre of the Oppressed, delivered to senior officers in 417 police stations. Approximately 7 months after the baseline, treated officers exhibit significant improvements in attitudes toward GBV, higher empathy, and stronger legal and procedural knowledge. These gains translate into real-world behavior. Standardized victim audits conducted roughly one year later show that treated officers are less likely to dismiss complaints or blame survivors and more likely to provide appropriate guidance and referrals, and administrative records show that stations register cases substantially more quickly after an incident occurs. Over a longer horizon, reports of GBV rise in treated stations, consistent with improved police responsiveness increasing survivors' willingness to seek help.

Beyond documenting that bias among state agents is malleable, the paper provides evidence on how change emerges and why it persists. Exploiting plausibly exogenous officer transfers, we show that treated officers continue to improve after moving to stations where no peer was trained,

while untreated officers improve after moving into treated stations. Together with lower reported harassment and anxiety among junior female constables, who were never targeted by the intervention, these patterns point to two reinforcing channels: durable individual-level change in how officers interpret situations and exercise discretion, and the diffusion of new norms through the station hierarchy. The sequence of effects is itself informative. Improvements in officer attitudes, conduct, and responsiveness appear first, while increases in citizen reporting emerge only with a lag, consistent with better service delivery gradually rebuilding trust rather than with mechanical compliance or short-run experimenter demand.

These findings speak to a central question in the study of state capacity. Whereas many reforms seek to improve service delivery by changing who provides it, through female recruitment, gender quotas, specialized courts, or dedicated women's police stations, we show that the behavior of the officers already in place can be reshaped from within. This distinction matters in settings where governments struggle to recruit and retain women despite formal quotas, and where tenure protections limit personnel reallocation. By targeting senior officers who shape both case outcomes and workplace norms, the intervention operates on the decision-makers whose discretion is most consequential for survivors.

More broadly, our results suggest that improving governance may require addressing the social and organizational foundations of bureaucratic behavior. Citizens do not experience the state in the abstract; they experience it through a single interaction with a frontline official, and the quality of that interaction can determine whether the law reaches them at all. Interventions that cultivate perspective-taking and reshape workplace norms can change how those interactions unfold, generating durable improvements in the treatment of vulnerable citizens and offering a scalable path to strengthening state capacity in high-discretion institutional settings.

References

Abramsky, T., Devries, K., Kiss, L., Nakuti, J., Kyegombe, N., Starmann, E., Cundill, B., Francisco, L., Kaye, D., Musuya, T., Michau, L., and Watts, C. (2014). Findings from the SASA! study: A cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda. *BMC Medicine*, 12:122.

Adams-Prassl, A., Huttunen, K., Nix, E., and Zhang, N. (2024). Violence against women at work. *The Quarterly Journal of Economics*, 139(2):937–991.

Alan, S., Baysan, C., Gumren, M., and Kubilay, E. (2021). Building social cohesion in ethnically mixed schools: An intervention on perspective taking. *The Quarterly Journal of Economics*, 136(4):2147–2194.

Amaral, S., Bhalotra, S., and Prakash, N. (2021). Gender, crime and punishment: Evidence from women police stations in india. IZA Discussion Paper 14250, Institute of Labor Economics (IZA).

Amaral, S., Borker, G., Fiala, N., Kumar, A., Prakash, N., and Sviatschi, M. M. (2025). Sexual harassment in public spaces and police patrols: Experimental evidence from urban india. *The Quarterly Journal of Economics*, 140(4):3191–3231.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Andries, M., Bursztyn, L., Chaney, T., Djourelova, M., and Imas, A. (2024). In their shoes: Empathy through information. Technical Report Working Paper 32569, National Bureau of Economic Research.

Ashraf, N., Bryan, G., Delfino, A., Holmes, E. A., Iacovone, L., Meyer, C. J., and Pople, A. (2024). Learning to see the world's opportunities: Memory, mental experiencing, and the economic lives of the vulnerable. Working Paper.

Asik, G., Erten, B., Tekin, E., Tumen, S., and Ulucan, H. (2026). Shielding women from vio-

lence: The effects of violence prevention and monitoring centers on intimate partner violence in türkiye. NBER Working Paper 34886, National Bureau of Economic Research.

Azulai, M., Rasul, I., Rogger, D., and Williams, M. J. (2020). Can training improve organizational culture? Experimental evidence from Ghana's civil service. Working paper, University College London; World Bank; University of Oxford.

Baird, S., McIntosh, C., and Özler, B. (2011). Cash or condition? evidence from a cash transfer experiment. *The Quarterly journal of economics*, 126(4):1709–1753.

Banerjee, A., La Ferrara, E., and Orozco, V. (2019). Entertainment, education, and attitudes toward domestic violence. In *AEA Papers and Proceedings*, volume 109, pages 133–137. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Banerjee, A. V., Chattopadhyay, R., Duflo, E., Keniston, D., and Singh, N. (2021). Improving police performance in rajasthan, india: experimental evidence on incentives, managerial autonomy, and training. *American Economic Journal: Economic Policy*, 13(1):36–66.

Baron-Cohen, S. and Wheelwright, S. (2004). The empathy quotient: an investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, 34(2):163–175.

Bernhardt, A., Kruks-Wisner, G., Mangla, A., and Sukhtankar, S. (2026). Beyond recruitment: Supporting gender integration in policing. Working Paper.

Besley, T. and Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American economic review*, 99(4):1218–1244.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Blatner, A. and Blatner, A. (1988). Foundations of psychodrama. *New York*.

Blattman, C., Jamison, J. C., and Sheridan, M. (2017). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review*, 107(4):1165–1206.

Boal, A. (2014). Theatre of the oppressed. In *The Improvisation Studies Reader*, pages 97–104. Routledge.

Borker, G. (Forthcoming). Safety first: Perceived risk of street harassment and educational choices of women. *American Economic Review*.

Broockman, D. E. and Kalla, J. L. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282):220–224.

Bureau of Police Research and Development (2023). Data on police organizations. Technical report, Ministry of Home Affairs, Government of India.

Bursztyn, L., Haaland, I. K., Röver, N., and Roth, C. (2025). The social desirability atlas. Technical report, National Bureau of Economic Research.

Canales, R., Santini, J. F., González Magaña, M., and Cherem, A. (2025). Shaping police officer mindsets and behaviors: Experimental evidence of procedural justice training. *Management Science*, 71(11):8995–9013.

Chaudhuri, S., Hoff, K., Jalan, J., and Santra, S. (2025). Participatory theater empowers women: Evidence from west bengal, india. Working paper, McGill University; Columbia University; Centre for Studies in Social Sciences, Kolkata; Centre for Training and Research in Public Finance and Policy, Kolkata. June 8, 2025.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernandez-Val, I. (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. *National Bureau of Economic Research*.

Crowne, D. P. and Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, 24(4):349–354.

Dahl, G. B., Kotsadam, A., and Rooth, D.-O. (2021). Does integration change gender attitudes? the effect of randomly assigning women to traditionally male teams. *The Quarterly Journal of Economics*, 136(2):987–1030.

Das, J., Holla, A., Mohpal, A., and Muralidharan, K. (2016). Quality and accountability in health care delivery: audit-study evidence from primary care in india. *American Economic Review*, 106(12):3765–3799.

Davis, M. H. et al. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*.

Dhar, D., Jain, T., and Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. *American economic review*, 112(3):899–927.

Di Maio, M. and Fiala, N. (2020). Be wary of those who ask: a randomized experiment on the size and determinants of the enumerator effect. *The World Bank Economic Review*, 34(3):654–669.

Drexler, A., Fischer, G., and Schoar, A. (2014). Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2):1–31.

Dube, O., MacArthur, S. J., and Shah, A. K. (2025). A cognitive view of policing. *The Quarterly Journal of Economics*, 140(1):745–791.

Duflo, E., Hanna, R., and Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–1278.

Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica*, 90(3):1063–1084.

Fiala, N., Garcia-Hernandez, A., Narula, K., and Prakash, N. (2022). Wheels of change: Transforming girls' lives with bicycles. Technical report, CESifo.

Finan, F., Olken, B. A., and Pande, R. (2017). The personnel economics of the developing state. *Handbook of economic field experiments*, 2:467–514.

Folke, O. and Rickne, J. (2022). Sexual harassment and gender inequality in the labor market. *The Quarterly Journal of Economics*, 137(4):2163–2212.

Freire, P. (1970). *Pedagogy of the Oppressed*. Herder and Herder, New York. Translated by Myra Bergman Ramos.

Freire, P. (2020). Pedagogy of the oppressed. In *Toward a sociology of education*, pages 374–386. Routledge.

García-Hombrados, J., Martínez-Matute, M., and Villa, C. (2024). Specialised courts and the reporting of intimate partner violence: Evidence from spain. *Journal of Public Economics*, 239:105243.

Harris, D., Borcan, O., Serra, D., Telli, H., Schettini, B., and Dercon, S. (2022). Proud to belong: The impact of ethics training on police officers. Technical report, University of Oxford.

Hoeffler, A. and Fearon, J. D. (2014). Benefits and costs of the conflict and violence targets for the post-2015 development agenda. Conflict and violence assessment paper, Copenhagen Consensus Center.

Hoeffler, A. and Fearon, J. D. (2026). *Worse than War: The Global Costs of Violence*. Princeton University Press.

Iyer, L. and Mani, A. (2012). Traveling agents: political change and bureaucratic turnover in india. *Review of Economics and Statistics*, 94(3):723–739.

Kalla, J. L. and Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, 114(2):410–425.

Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *The Review of Economics and statistics*, 91(3):437–456.

Maynard-Moody, S. and Musheno, M. (2003). *Cops, Teachers, Counselors: Stories from the Front Lines of Public Service*. University of Michigan Press, Ann Arbor.

Michaels-Igbokwe, C., Abramsky, T., Devries, K., Michau, L., Musuya, T., and Watts, C. (2016). Cost and cost-effectiveness analysis of a community mobilisation intervention to reduce intimate partner violence in Kampala, Uganda. *BMC Public Health*, 16:196.

Miller, A. R. and Segal, C. (2019). Do female officers improve law enforcement quality? effects on crime reporting and domestic violence. *The review of economic studies*, 86(5):2220–2247.

Murvartian, L., Matías-García, J. A., Saavedra-Macías, F. J., and Crowe, A. (2024). A systematic review of public stigmatization toward women victims of intimate partner violence in low- and middle-income countries. *Trauma, Violence, & Abuse*, 25(2):1349–1364.

National Crime Records Bureau (2023). Crime in India 2022: Volume I. Technical report, Ministry of Home Affairs, Government of India.

National Herald India (2022). 200 police outposts in bihar to be upgraded shortly. https://www.nationalheraldindia.com/national/200-police-outposts-in-bihar-to-be-upgraded-shortly. Accessed on April 4, 2024.

NFHS-5 (2021). *National Family Health Survey (NFHS-5), India, 2019–21: Bihar*. MoHFW, Government of India, Mumbai: IIPS.

NFHS-5 (2022). *National Family Health Survey – 5: India Fact Sheet*. MoHFW, Government of India, Mumbai: IIPS.

Page, L. and Pande, R. (2018). Ending global poverty: Why money isn't enough. *Journal of Economic Perspectives*, 32(4):173–200.

Puri, L. (2016). The economic costs of violence against women. UN Women.

Raghavendra, S., Chadha, M., and Duvvury, N. (2018). Cost-benefit analysis of proposed interventions to reduce intimate partner violence in Andhra Pradesh. Andhra pradesh priorities working paper, Copenhagen Consensus Center. An India Consensus Prioritization Project, working draft as of May 16, 2018.

Rasul, I. and Rogger, D. (2018). Management of bureaucrats and public service delivery: Evidence from the nigerian civil service. *The Economic Journal*, 128(608):413–446.

Rehfisch, J. M. (1958). A scale for personality rigidity. *Journal of Consulting Psychology*, 22(1):11.

Sharma, K. (2024). Tackling sexual harassment: Short and long-run experimental evidence from India. Technical report, SSRN. SSRN Working Paper 6460764.

Srivastava, S., Kumar, K., McDougal, L., Kannaujiya, A. K., Sikarwar, A., Raj, A., and Singh, A. (2023). Spatial heterogeneity in intimate partner violence across the 640 districts of India: A secondary analysis of a cross-sectional, population-based survey by use of model-based small-area estimation. *Lancet Global Health*, 11(10):e1587–e1597.

Stewart, R., Wright, B., Smith, L., Roberts, S., and Russell, N. (2021). Gendered stereotypes and norms: A systematic review of interventions designed to shift attitudes and behaviour. *Heliyon*, 7(4):e06660.

Sukhtankar, S., Kruks-Wisner, G., and Mangla, A. (2022). Policing in patriarchy: An experimental evaluation of reforms to improve police responsiveness to women in india. *Science*, 377(6602):191–198.

Sviatschi, M. M. and Trako, I. (2024). Gender violence, enforcement, and human capital: Evidence from women's justice centers in peru. *Journal of Development Economics*, 168:103262.

Tata Trusts (2023). India justice report: Ranking states on police, judiciary, prisons and legal aid (2022). Technical report, Tata Trusts, New Delhi, India. In collaboration with Centre for Social Justice, Common Cause, Commonwealth Human Rights Initiative, DAKSH, TISS-Prayas, and Vidhi Centre for Legal Policy.

Vyas, S., Meinhart, M., Troy, K., Brumbaum, H., Poulton, C., and Stark, L. (2023). The economic cost of violence against women and girls in low- and middle-income countries: A systematic review of the evidence. *Trauma, Violence, & Abuse*, 24(1):44–55.

Appendix

A Tables and Figures

Table A1: Summary—Police Officers in Sample Districts

Officer Rank	Male	Female
	(1)	(2)
Station Head Officer (SHO)	388	6
Inspector	42	0
Sub-Inspector of Police (SI)	1549	230
Assistant Sub-Inspector (ASI)	1627	39
Head Constable	400	84
Writer Constable	349	30
Constable	4250	1596
Total	8605	1985

Notes: This table summarizes police officers across various designations in the sample districts at the time of the baseline survey.

Table A2: Individual Characteristics and Attrition

	(1)	(2)
Expressive Arts	-0.043***	0.138
	(0.015)	(0.165)
Expressive Arts × Social desirability score		0.005
		(0.015)
Expressive Arts × Age		-0.009**
		(0.004)
Expressive Arts × Years of service (force)		0.006*
		(0.003)
Expressive Arts × Years of service (station)		0.024**
		(0.009)
Expressive Arts × Married		0.044
		(0.065)
Expressive Arts × Caste: SC		0.026
		(0.090)
Expressive Arts × Caste: ST		0.138
		(0.122)
Expressive Arts × Caste: OBC		0.079
		(0.086)
Expressive Arts × Caste: General		0.073
		(0.087)
Expressive Arts × Education: 10th		0.001
		(0.073)
Expressive Arts × Education: Plus 2		0.029
		(0.075)
Expressive Arts × Education: Diploma		-0.067
		(0.168)
Expressive Arts × Education: Some college		-0.304
		(0.367)
Expressive Arts × Education: B.A.		0.017
		(0.062)
Expressive Arts × Education: Post Graduate		0.000
		(.)
Expressive Arts × Rank: ASI		-0.043
		(0.062)
Expressive Arts × Rank: SI		-0.017
		(0.058)
Expressive Arts × Rank: PSI		0.016
		(0.089)
Expressive Arts × Rank: Inspector		-0.260
		(0.287)
Expressive Arts × Rank: SHO		0.000
		(.)
Strata FE	Yes	Yes
Individual controls	No	Yes
N	3526	3526
P-value (F-test, interactions)		0.442

Table A3: Treatment Effects on Endline Outcomes: Lee Bounds

	GBV Skills Index	Attitudes Index	Technical Skills Index
	(1)	(2)	(3)
Expressive Arts	0.082**	0.082**	0.085**
	(0.040)	(0.042)	(0.043)
Control mean	-0.000	0.000	0.000
Lee bound (lower)	0.001	-0.002	0.005
Lee bound (upper)	0.189	0.198	0.201
Strata FE	Yes	Yes	Yes
Baseline control	Yes	Yes	Yes
Number of clusters	417	417	417
Number of officers	2571	2571	2571

Notes: Standard errors clustered at the station level (in parentheses). Regressions include only officers who completed both baseline and endline surveys, including those who switched between control and treatment stations between baseline and endline. All regressions include district and officer strength (at the station level) fixed effects and the baseline value of the outcome. Lee bounds computed by trimming 70 officers (5.03%) from the treatment group to account for the higher response rate in treatment (74.70%) relative to control (70.94%). Lower bound drops highest-scoring treated respondents; upper bound drops lowest-scoring treated respondents. All indices created using inverse covariance weighting (swindex) and normalized to the control-group mean. Source: Baseline and endline surveys; police station facility survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A4: Balance of Decoy Sample

	Control Mean	Treatment Mean	Difference (2)−(1)	SE	P-value
	(1)	(2)	(3)	(4)	(5)
Panel A: Demographics					
Age	45.087	46.052	0.923	0.921	0.318
Years of service (force)	19.660	19.654	-0.037	1.103	0.974
Years of service (station)	1.199	1.098	-0.102	0.078	0.191
Married	0.924	0.887	-0.038	0.030	0.197
Panel B: Caste					
Caste: SC	0.163	0.170	0.012	0.041	0.765
Caste: ST	0.017	0.021	0.004	0.016	0.818
Caste: OBC	0.465	0.433	-0.036	0.051	0.483
Caste: General	0.291	0.325	0.029	0.045	0.519
Panel C: Education					
Education: 10th	0.244	0.253	0.013	0.044	0.776
Education: Plus 2	0.128	0.170	0.038	0.038	0.321
Education: Diploma	0.000	0.005	0.005	0.005	0.327
Education: B.A.	0.552	0.510	-0.041	0.053	0.439
Education: Post Graduate	0.076	0.062	-0.014	0.026	0.599
Panel D: Rank					
Rank: ASI	0.395	0.407	0.010	0.048	0.834
Rank: SI	0.314	0.345	0.037	0.049	0.450
Rank: PSI	0.076	0.062	-0.017	0.023	0.454
Number of decoy visits	258	291	549		
Number of officers	172	194	366		

Notes: Standard errors clustered at the station level. Each row reports results from a separate OLS regression of the baseline characteristic on the treatment indicator, absorbing district and strata fixed effects. The difference column reports the coefficient on treatment with stars denoting significance of the p-value. Officer balance rows are restricted to officers who completed the baseline survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A5: Balance by Female Constable Survey Coverage

	No survey Mean	Survey Mean	Difference (2)−(1)	SE	P-value
	(1)	(2)	(3)	(4)	(5)
Panel A: Demographics					
Age	48.286	47.354	-0.439	0.477	0.359
Years of service (force)	23.199	22.300	-0.261	0.630	0.679
Years of service (station)	1.216	1.198	-0.005	0.054	0.929
Married	0.900	0.901	0.016	0.016	0.319
Panel B: Caste					
Caste: SC	0.193	0.184	-0.013	0.024	0.602
Caste: ST	0.028	0.032	0.004	0.010	0.651
Caste: OBC	0.403	0.377	-0.021	0.030	0.476
Caste: General	0.331	0.362	0.028	0.027	0.306
Panel C: Education					
Education: 10th	0.331	0.327	0.023	0.026	0.380
Education: Plus 2	0.197	0.169	-0.016	0.022	0.456
Education: Diploma	0.010	0.010	0.003	0.006	0.647
Education: B.A.	0.393	0.419	-0.019	0.025	0.450
Education: Post Graduate	0.069	0.072	0.006	0.014	0.654
Panel D: Rank					
Rank: ASI	0.469	0.451	0.007	0.024	0.756
Rank: SI	0.331	0.390	0.014	0.026	0.574
Rank: PSI	0.059	0.062	-0.003	0.011	0.783
Number of female officer surveys	0	1810	1810		

Notes: Standard errors clustered at the station level. Each row reports results from a separate OLS regression of the baseline officer characteristic on an indicator equal to one if female officers from the station completed the female constable survey, absorbing district and strata fixed effects. The difference column reports the coefficient on the survey indicator with stars denoting significance of the p-value. Officer balance rows are restricted to officers who completed the baseline survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A6: Transfers are Orthogonal to Treatment Assignment

	Moved across treatment arms		
	(1)	(2)	(3)
Expressive Arts	-0.018	-0.014	-0.016
	(0.014)	(0.014)	(0.014)
Observations	2571	2571	2571
Clusters	417	417	417
Strata FE	Yes	Yes	Yes
Officer controls	No	Yes	Yes
Station controls	No	No	Yes

Notes: Standard errors clustered at the baseline police station level in parentheses. The dependent variable equals one if the officer moved between treatment and control stations between baseline and endline. All specifications include district and officer-strength fixed effects. Column (2) adds officer controls (age, years of service, education, rank, caste, marital status). Column (3) adds station controls. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A7: Baseline Balance: Non-Transferred Versus Any Transfer

	Non-Trans.	Transferred	Diff (2)−(1)
	(1)	(2)	(3)
Panel A: Officer characteristics			
Years of service (Bihar Police)	22.758	21.154	-1.604***
			(0.514)
Married	0.914	0.902	-0.012.
			(0.013)
Scheduled Caste (SC)	0.192	0.181	-0.011.
			(0.018)
Scheduled Tribe (ST)	0.031	0.031	-0.001.
			(0.008)
Other Backward Class (OBC)	0.382	0.394	0.011.
			(0.022)
General Caste	0.355	0.368	0.013.
			(0.022)
10th Pass	0.339	0.315	-0.023.
			(0.021)
12th Pass	0.181	0.185	0.004.
			(0.018)
Diploma	0.010	0.006	-0.004.
			(0.004)
College	0.002	0.002	-0.001.
			(0.002)
BA	0.397	0.413	0.016.
			(0.022)
MA	0.069	0.078	0.009.
			(0.012)
Station Head Officer (SHO)	0.090	0.078	-0.012.
			(0.013)
Panel B: Baseline outcome indices			
GBV Skills Index	0.027	0.035	0.009.
			(0.044)
Attitudes Index	0.019	0.047	0.028.
			(0.045)
Technical Skills Index	-0.009	-0.015	-0.006.
			(0.045)
Observations	1918	653	
Joint p-value (indices)			0.809
Joint p-value (controls)			0.122

Notes: Sample consists of officers who completed both baseline and endline surveys. Column (1) reports mean baseline characteristics for non-transferred officers. Column (2) reports means for all transferred officers regardless of transfer direction. Column (3) reports the difference between columns (1) and (2) with standard errors in parentheses from two-sample t-tests. The joint p-values are from a single regression of the transfer indicator on all listed baseline characteristics, with strata fixed effects absorbed and standard errors clustered at the baseline station level. The first joint p-value tests the baseline outcome indices only; the second tests the officer characteristics only. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A8: Treatment Effects on Endline Outcomes: Controlling for SDB

	GBV Skills Index	Attitudes Index	Technical Skills Index
	(1)	(2)	(3)
Expressive Arts	0.086**	0.084**	0.085**
	(0.040)	(0.041)	(0.042)
Control mean	0.000	0.000	0.000
Strata FE	Yes	Yes	Yes
Baseline control	Yes	Yes	Yes
SDB Baseline	Yes	Yes	Yes
Station controls	Yes	Yes	Yes
Officer controls	Yes	Yes	Yes
Number of clusters	417	417	417
Number of officers	2571	2571	2571

Notes: Standard errors clustered at the station level (in parentheses). Regressions include only officers who completed both baseline and endline surveys, including those who switched between control and treatment stations between baseline and endline. All regressions include district and officer strength (at the station level) fixed effects. All regressions include the baseline value of the outcome, social desirability bias score, station controls, and officer controls. All indices created using inverse covariance weighting (swindex) and normalized to the control-group mean. Source: Baseline and endline surveys; police station facility survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A9: Treatment Effects on Endline Outcomes: Including Enumerator Fixed Effects Only

	GBV Skills Index	Attitudes Index	Technical Skills Index
	(1)	(2)	(3)
Expressive Arts	0.090**	0.086**	0.081*
	(0.040)	(0.042)	(0.042)
Control mean	0.000	0.000	0.000
Enumerator FE	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes
SDB Baseline	Yes	Yes	Yes
Station controls	Yes	Yes	Yes
Officer controls	Yes	Yes	Yes
Number of clusters	417	417	417
Number of officers	2570	2570	2570

Notes: Standard errors clustered at the station level (in parentheses). Regressions include only officers who completed both baseline and endline surveys, including those who switched between control and treatment stations between baseline and endline. Regressions include district and strata fixed effects and enumerator (sv_name_bl) fixed effects; baseline outcome control, SDB control, station controls, or officer controls are included. All indices created using inverse covariance weighting (swindex) and normalized to the control-group mean. Source: Baseline and endline surveys; police station facility survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A10: Treatment Effects: Attitudes Index and Components

	Attitudes Index	Victim Blame Index	Empathy Index	Attitudes GBV Index	Ext. Resp. Index	Truthfulness Index
	(1)	(2)	(3)	(4)	(5)	(6)
Expressive Arts	0.082**	0.009	0.078*	0.011	0.052	0.117***
	(0.042)	(0.042)	(0.047)	(0.043)	(0.041)	(0.043)
Control mean	0.000	0.000	0.000	0.000	0.000	0.000
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline control	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	417	417	417	417	417	417
Officers	2571	2571	2571	2571	2571	2571

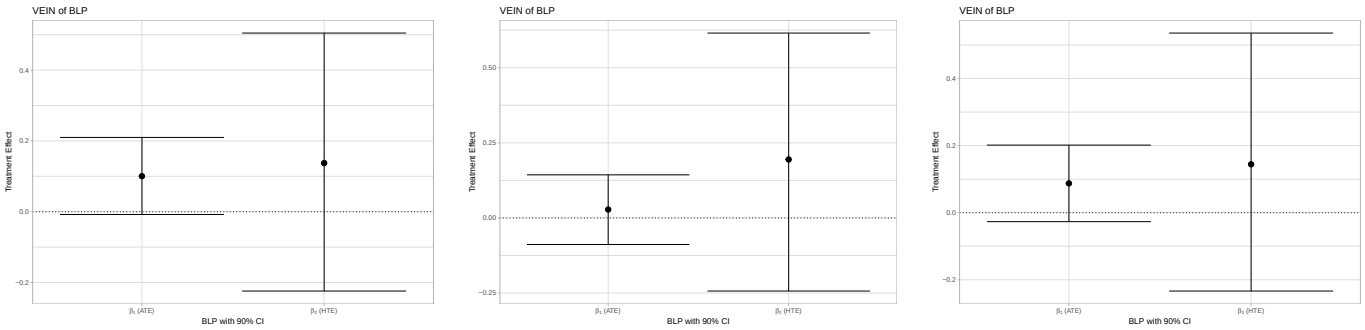
Notes: Standard errors clustered at the baseline police station level in parentheses. Regressions include only officers who completed both baseline and endline surveys, including those who switched between control and treatment stations between baseline and endline. All regressions include district and officer-strength fixed effects and the baseline value of the outcome. Indices are created using inverse covariance weighting (swindex) and normalized to the control-group mean in the index-construction sample. Control means in this table are computed in the regression estimation sample, so they can differ slightly from zero when the estimation sample differs from the normalization sample. Source: Baseline and endline surveys; police station facility survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A11: Treatment Effects: Technical Skills Index and Components

	Technical Skills Index	GBV Identification Index	Legal Knowledge Index
	(1)	(2)	(3)
Expressive Arts	0.085**	0.076*	0.060
	(0.043)	(0.043)	(0.040)
Control mean	0.000	0.000	0.000
Strata FE	Yes	Yes	Yes
Baseline control	Yes	Yes	Yes
Clusters	417	417	417
Officers	2571	2571	2571

Notes: Standard errors clustered at the baseline police station level in parentheses. Regressions include only officers who completed both baseline and endline surveys, including those who switched between control and treatment stations between baseline and endline. All regressions include district and officer-strength fixed effects and the baseline value of the outcome. Indices are created using inverse covariance weighting (swindex) and normalized to the control-group mean in the index-construction sample. Control means in this table are computed in the regression estimation sample, so they can differ slightly from zero when the estimation sample differs from the normalization sample. Source: Baseline and endline surveys; police station facility survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

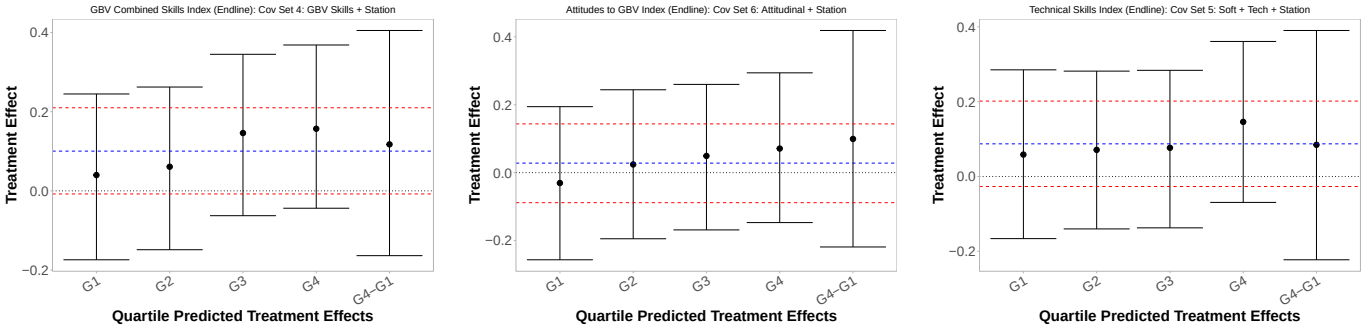
Figure A1: Treatment Effect Heterogeneity (Best Linear Predictor) using Chernozhukov et al. (2018)



(a) GBV Skills Index **(b)** Attitudes Index **(c)** Technical Skills Index

Notes: Each panel plots the estimated relationship between the predicted treatment effect and the corresponding outcome. Predicted treatment effects are generated using the Generic Machine Learning proxy predictor $S(Z)$ (Chernozhukov et al., 2018), estimated from baseline covariates Z via an ensemble of lasso and random forest learners. The analysis tests for treatment-effect heterogeneity using the heterogeneous treatment effect (HTE) coefficient from the Best Linear Predictor specification. Across the three outcomes, the estimated HTE coefficients are statistically insignificant, providing no statistically significant evidence of systematic heterogeneity in treatment effects along the dimension captured by the proxy predictor.

Figure A2: Treatment Effect Heterogeneity (Groupe ATEs) using Chernozhukov et al. (2018)



(a) GBV Skills Index **(b)** Attitudes Index **(c)** Technical Skills Index

Notes: The figure reports group-specific treatment effects from the Classification Analysis of Heterogeneity (CLAN) framework. The sample is divided into four groups ($G1$–$G4$) according to the predicted treatment effects from the heterogeneous treatment-effect model, with $G1$ and $G4$ representing the groups with the lowest and highest predicted treatment effects, respectively. The estimated treatment effects for $G4$ are not statistically significantly different from those for $G1$ for all three outcomes.

Table A12: Effect of Expressive Arts Training on Police Responses to GBV Decoy Victims, Controlling for GBV Caseload During Month of Treatment

	Waste of time	Cases false	Blames victim	Referred to WPS	Explained alternative	Asks unnecessary questions
	(1)	(2)	(3)	(4)	(5)	(6)
Expressive Arts	-0.06***	-0.08***	-0.15***	0.08***	0.08*	-0.14***
	(0.02)	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)
Control mean	0.06	0.15	0.72	0.09	0.21	0.63
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
GBV caseload control	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	183	183	183	183	183	183
Number of decoy visits	549	549	549	549	549	549

Notes: Standard errors clustered at the station level (in parentheses). The unit of observation is a decoy visit. Regressions include district and strata fixed effects, and the number of GBV cases recorded at that police station (FIR admin data) in the same calendar month as the officer's training date. Control means are computed in the regression estimation sample. Source: Decoy survey; FIR admin data. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A13: Effect of Expressive Arts Training on Time Use

	Investigating CAW	Follow-up CAW	Assisting victims
	(1)	(2)	(3)
Expressive Arts	4.099***	2.147**	1.298
	(1.268)	(1.080)	(1.233)
Control mean	5.955	3.733	5.748
Strata FE	Yes	Yes	Yes
Baseline outcome	Yes	Yes	Yes
SDB baseline	Yes	Yes	Yes
Station controls	Yes	Yes	Yes
Officer controls	Yes	Yes	Yes
Clusters	417	417	417
Officers	2571	2571	2571

Notes: Standard errors clustered at the baseline police station level in parentheses. Outcomes are minutes and are restricted to values of 480 or less, following the draft. Regressions include district and officer-strength fixed effects, baseline outcome, baseline social desirability, station controls, and officer controls. Source: Baseline and endline surveys, police station facility survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B The Intervention in Detail

This Appendix provides the full design, curriculum, delivery, and monitoring detail underlying Section 3: the diagnostic process behind the curriculum, the day-by-day session structure, a mapping of curriculum components to the three channels of change, facilitator recruitment and training, the district-level implementation protocol, fidelity monitoring, and qualitative evidence from training sessions.

B.1 Ideation and Diagnosis

The curriculum was developed in three stages. First, the research team consulted former and active police personnel, including senior IPS officials, to identify gaps in officers' response to GBV that mandatory in-service training does not address. Second, the team consulted domain experts and applied-theatre practitioners to identify a training format suited to closing those gaps. Third, the team conducted extended field visits to police stations, observing routine operations and speaking with active-duty officers, to calibrate the intensity and pacing required for the training to produce a tangible effect rather than a pro forma compliance exercise.

This process converged on five recurring institutional gaps, listed in the order in which they emerged from fieldwork:

(i) bias and prejudice against women specifically;

(ii) an inability to respond to complainants with empathy;

(iii) patriarchal norms that manifest as victim-blaming;

(iv) a lack of clarity around the legal and practical meaning of consent; and

(v) a general absence of agency, with officers reporting that they did not see GBV outcomes as within their power to change.

These gaps were not abstractions. During scoping interviews, for instance, officers repeatedly volunteered a specific and revealing narrative: that women who file domestic violence complaints are frequently doing so opportunistically, to gain leverage over household resources and decision-making rather than because they have experienced genuine harm. Narratives of this kind – which recast a survivor's complaint as a manipulative act – were the concrete evidence that the binding constraint was not officers' knowledge of the law but the mental models through which they interpreted a complaint before the law was ever applied. Addressing this required directly unsettling and re-categorizing these narratives rather than layering additional legal instruction on top of them.

Fieldwork indicated that the knowledge gap among officers was comparatively narrow and tractable, while the harder constraint was officers' preconceived, gendered notions, which shaped attitudes and conduct even among officers who could correctly state the relevant law. This diagnosis directed curriculum design toward attitudinal and normative content first, with technical and legal content anchoring the final day rather than constituting the bulk of the training. The resulting session objectives were to (i) sensitize participants to gender and GBV as institutional issues rather than purely individual ones; (ii) build a shared professional premise around service-centric policing; and (iii) improve the quality of officers' direct interactions with survivors.

The curriculum itself was developed in 2019 in collaboration with *Mittika*, a Delhi-based NGO and one of India's leading practitioners of applied theatre and participatory methods for social change on gender, caste, and related forms of inequality; reaching the final version took several months from inception. Several senior IPS officers were directly involved in shaping the curriculum, alongside the research team, which contributed field-based input drawn from qualitative scoping and semi-structured interviews with police officers. The curriculum draws methodologically on Applied Theatre and Theatre of the Oppressed, building on the foundational work of Boal (2014), Freire (2020), and Blatner and Blatner (1988), whose immersive, experiential techniques have been documented as effective across a range of settings. Chaudhuri et al. (2025) apply an analogous applied-theatre method to revise community norms around wife-beating and find reduced domestic violence in villages in West Bengal. The lead trainer on the curriculum, Ms. Jaya Iyer, is an award-winning applied-theatre practitioner with three decades of experience and a recipient of the UNESCO Aschberg Bursary to study Theatre of the Oppressed under Augusto Boal;[21] her direct involvement in curriculum design, alongside the inputs above, culminated in the final training curriculum and an accompanying handbook on handling GBV cases, distributed to officers in both Hindi and English (Section B.7).

B.2 Workshop Design Constraints

The design team set a batch size of 15–30 officers and a duration of three full days (approximately 24 hours), guided by three constraints. First, the workshop must create an environment in which dissent and disagreement are treated as productive rather than discarded, since officers who feel judged rather than engaged are unlikely to update their beliefs. Second, the process must leave enough time and space for peer discussion and for divergent views to be heard, rather than becoming purely didactic. Third, the workshop must be experiential, requiring officers to physically occupy the perspective of a survivor of GBV, rather than discuss that perspective in the abstract. Absent these constraints, the design team judged that the module would default to the same top-down, information-centric format as existing in-service trainings, which officers and administrators alike regarded as ineffective. The three-day arc described below was built as a graded progression addressing denial and resistance before introducing the harder GBV content.

B.3 Curriculum: Day-by-Day Session Structure

Tables A14–A16 report the full session-by-session structure of the three-day curriculum summarized in Section 3.

[21] Ms. Iyer subsequently led the Training of Trainers described in Section B.4.

Table A14: Day 1 Session Structure: Setting the Context

Session	Name	Objective	Rationale
1.1	Introduction	Make officers comfortable; set premises for the workshop; align expectations	Lays the groundwork for trust and open engagement
1.2	Envisioning the Ideal	Articulate current realities as individual officers and as a collective force; enumerate attributes of an "ideal" police station	Motivates participants by anchoring the workshop in a better version of their own role
1.3	Purpose of Policing	Build an understanding of the constituency served; analyze the purpose of policing	Connects policing to community service and democratic accountability
1.4	Of Perceptions and Bias	Examine concepts of bias and stereotype; analyze their impact on police duties	Introduces self-awareness and impartiality without yet naming GBV directly

Notes: No GBV content is introduced on Day 1. The exclusive purpose of Day 1 is to prepare officers, through non-threatening exercises, to absorb the more confrontational material introduced on Day 2.

Table A15: Day 2 Session Structure: Challenging Beliefs

Session	Name	Objective	Rationale
2.1	Recap	Recap of the previous day	Reinforces prior takeaways ahead of new content
2.2	Of Inequality and Discrimination	Articulate linkages between power, inequality, and gender	Encourages reflection on the structural roots of gender-based discrimination
2.3	Gender Stereotypes and Violence	Analyze the impact of gender stereotypes	Helps officers identify how stereotypes normalize or enable violence
2.4	Life Cycle of Violence	Consolidate the different forms of GBV women experience at different life stages	Illustrates GBV as systemic and persistent, not an isolated incident
2.5	Victim Blaming	Explore who is responsible and accountable for acts of GBV; discuss the meaning and consequences of victim-blaming	Directly challenges harmful narratives and builds accountability norms
2.6	Open Space	Decompress and prepare for the final day	Provides emotional recovery time after cognitively demanding content

Notes: Sessions 2.2–2.5 constitute the curriculum's core normative content and correspond most directly to the attitudes and social norm channels described in Section 3.3.

Table A16: Day 3 Session Structure: Skills and Action

Session	Name	Objective	Rationale
3.1	Recap	Recap of the previous day	Reinforces learning and maintains continuity
3.2	Of Rules and Laws	Examine technical/legal knowledge on GBV; walk through the practical handbook	Equips officers with legal clarity and a lasting reference
3.3	Survivor-Centric Policing	Examine and develop skills for interacting with survivors	Promotes an empathetic, rights-based approach to victim interaction
3.4	Mapping Action Points	Examine and improve police interaction; articulate concrete action points	Encourages specific, personal commitments rather than abstract intent
3.5	Change Ahead	Summarize key learnings; build an individual action plan	Solidifies key messages and operationalizes behavior change

Notes: Session 3.2 corresponds to the technical knowledge channel; Sessions 3.3–3.5 are designed to convert attitudinal and normative change into concrete, station-specific practice.

B.3.1 Mapping Curriculum Components to Main Channels

Table A17: Curriculum Components and Channels of Change

Curriculum component	Attitudes channel	Workplace channel
Day 1: perceptions and bias exercises	✓	
Day 2: role reversal / re-enactment	✓	✓
Day 2: collective scene negotiation		✓
Snakes and Ladders game	✓	✓
Web of Relations exercise		✓
Day 3: survivor-centric practice	✓	
Day 3: personal action plan	✓	✓

Notes: A ✓ indicates that the research team, in designing the curriculum, identified the corresponding channel as a primary intended mechanism for that component. Several components are designed to operate through more than one channel simultaneously; this is intentional, since the theory of change (Section 3.3) treats attitude change and norm formation as mutually reinforcing rather than substitutes.

B.4 Facilitator Recruitment and Training

Facilitators were screened on five criteria: (i) clarity and conviction regarding gender equality and justice; (ii) stamina and skill for facilitating experiential processes; (iii) prior experience working with people in positions of authority (desirable, not required); (iv) facilitation acumen, troubleshooting, and improvisational skill; and (v) resilience, teamwork, and self-motivation.

The Training of Trainers (ToT) was held in September 2022 as a six-day intensive residential program, led by Jaya Iyer and Sutapa Sanyal (former Director General of Police, Uttar Pradesh). Of 19 facilitators invited to the ToT, 8 were natives of Bihar and 11 were from other Indian states; the cohort comprised 10 women and 9 men, all with prior experience as applied-theatre practitioners. The ToT included intensive reading of the curriculum, pair-wise dry runs of every session, and structured peer and participant feedback after each dry run, followed by a three-day pilot in Arwal district (September 18–20, 2022; a non-sample district chosen for its proximity to Patna and its small size), attended by 17 officers.

Facilitators were graded continuously throughout the ToT and pilot on their understanding of GBV-related issues, content delivery, energy and conviction, teamwork, and process flow. Of the 19 facilitators trained, 14 (74 percent) were retained for the main intervention; retained facilitators were subsequently paired, with each batch assigned two facilitators to share delivery responsibilities and reduce psychological strain associated with repeatedly enacting distressing scenarios.

B.5 Implementation Protocol

Implementation followed a standard operating procedure developed after the pilot, with six components.

1. **Staggered rollout.** Training was staggered across districts (one or two districts at a time) rather than run simultaneously, both to avoid overstretching the implementing partner's capacity and to preserve the research team's ability to monitor fidelity closely.

2. **Coordination with senior police leadership.** The team held detailed discussions with Director General (DG) and Additional Director General (ADG) level IPS officials to align the curriculum with the operational context of Bihar Police and to sequence data collection and training around law-and-order priorities in each district.

3. **District-level approval.** Cooperation depended on approval from each district's Superintendent of Police (SP), who issued a letter of approval to selected stations and assigned a nodal officer (typically a Deputy SP) to coordinate logistics, confirm that only sampled stations received treatment, and supply updated station-level rosters to track transfers over the study period.

4. **Baseline first.** The baseline survey was completed in every station of a district before training began, both to support rigorous evaluation and to cross-check attendance rosters against the SP's official list.

5. **Field team structure.** Each training session was staffed by enumerators (reflection surveys), Project Supervisors (logistics and attendance), and Project Assistants (overall coordination), all reporting to a Research Associate responsible for monitoring the full intervention. SHOs were contacted ahead of each session to release invited officers from other duties; where attendance could not be secured directly, the SP or nodal officer was asked to intervene.

6. **Batch construction.** No more than 3–4 officers from a single station were invited to any one batch, generating 3–4 batches per district, with the interval between batches kept short to limit knowledge spillovers from earlier to later batches within the same district.

B.6 Monitoring and Fidelity

Fidelity to the curriculum was tracked through five concurrent monitoring instruments: (i) batch-level attendance records maintained by field staff; (ii) detailed field notes recording logistics, facilitator performance, officer response, and qualitative evidence of attitudinal change, maintained by Research Associates for every batch; (iii) a grading/monitoring sheet scoring each session on logistics, facilitator engagement, officer engagement, attendance, and participation; (iv) a structured reflection session with facilitators and field staff at the end of each training day; and (v) a monthly review meeting between the lead facilitator and the full facilitator team to share peer feedback and harmonize delivery across batches. The lead facilitator also conducted bi-monthly field visits to directly observe sessions and assess whether genuine behavioral change, rather than rote compliance, was occurring over the three days.

B.7 Training Handbook and Qualitative Evidence

Officers received a bilingual (Hindi/English) handbook summarizing key GBV-relevant laws, procedures, and referral pathways (Figure A3), designed as a lasting reference that officers could consult after the training concluded.

Selected photographs from training activities are reported in Figures A4–A7.

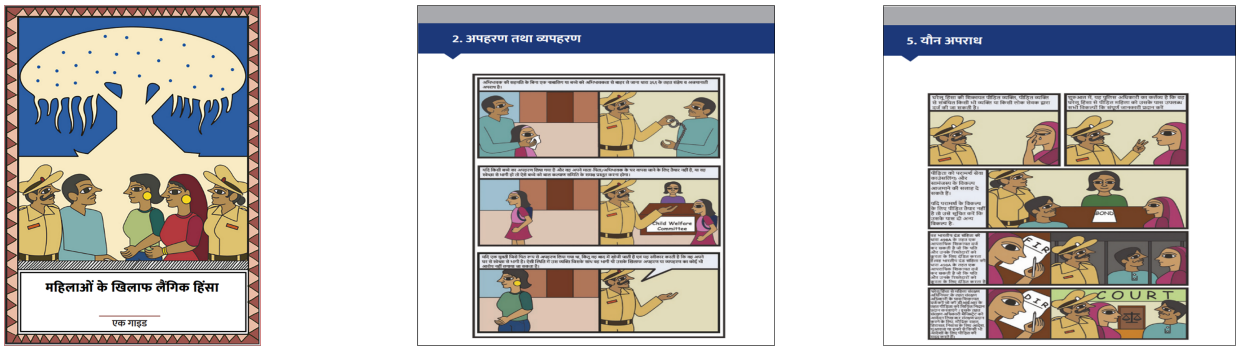


Figure A3: Sample pages from the practical handbook for police officers

Notes: The handbook served as a quick reference guide for officers, summarizing key laws, procedures, and practical guidance for handling gender-based violence cases in a more informed and gender-sensitive manner. Distributed in both Hindi and English.

Perceptions before training **Reflections after training**

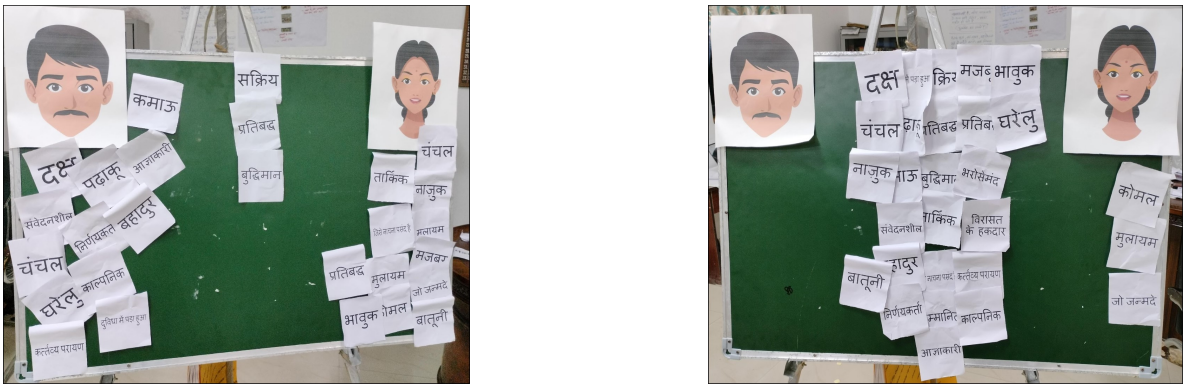


Figure A4: Perceptions of gender roles, before and after training

Notes: This activity, part of the Day 1 session on perceptions and bias (Table A14), asked officers to record their views on gender roles before the training began and to revisit them at its conclusion, surfacing how associations between traits and professional roles are socially constructed rather than innate.

Envisioning the ideal police station **Key outcomes from discussions**

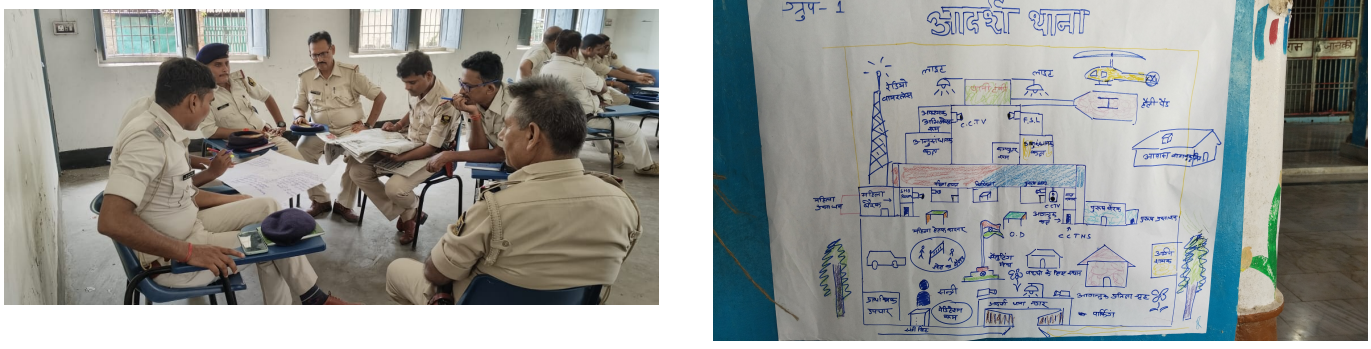


Figure A5: Participatory exercise: envisioning the ideal police station

Notes: Officers discussed the institutional and interpersonal changes needed to make police stations more survivor-centered, accessible, and responsive, and identified areas for individual and institutional improvement.



Figure A6: Interactive activity: Snakes and Ladders (*Saanp-Seedhi*)
Notes: Officers moved through realistic GBV scenarios. Survivor-centered responses moved participants forward, while harmful practices such as victim-blaming moved them backward, reinforcing best practices and common pitfalls in handling complaints.



Figure A7: Interactive activity: Web of Relations
Notes: This activity mapped systemic interactions within a police station, showing how individual actions shape broader institutional culture and policing norms.

Facilitator field notes. Among other observations, field notes recorded moments of visible attitudinal engagement during role-play. In one exercise, an officer enacting the responding-officer role in a domestic violence scenario stopped the scene, appearing visibly affected, to attend to the officer playing the survivor before resuming. In another batch, officers engaged with sufficient enthusiasm in the role-play that they independently produced a poster summarizing the session's key lessons, which they indicated they intended to display at their home station.

Participant reflections. The following reflections, recorded during closing sessions and translated from Hindi, illustrate the range of officer responses to the training:

> *"Hamare niyantran me bhi bohot kuch hai. Hum nyay ki pehli seedhi hai!"*
> Translation: We can also influence a lot of things. We are the first step towards justice.

> *"Aise practical way me hi sikhana chahiye chizon ko. Hamari baaki trainings bohot boring hoti hain, toh kuch seekhne ko nahi milta!"*
> Translation: Things should be taught in this practical way. Our other trainings are often very boring, so we end up learning very little.

> *"Mera nazariya hi badal gaya bilkul."*
> Translation: My perspective toward GBV has completely changed.

These reflections are illustrative rather than representative and are reported to contextualize the quantitative first-stage evidence on empathy and perspective-taking presented in Section 6.1; they are not used as outcome measures in any regression reported in the paper.

C Main Outcomes and Indices Construction

This appendix describes the construction of the outcome indices and lists the survey questions underlying each index.

C.1 Procedure for Index Construction

The indices for our primary and secondary outcomes are constructed by aggregating responses to several individual survey questions, following Anderson (2008) and implemented via the `swindex` command in Stata. For each index, the component variables are first standardized to have mean zero and unit standard deviation, using the control group mean and standard deviation as the reference. The standardized components are then aggregated into a summary index using a weighted average, with weights derived from the inverse of the covariance matrix of the standardized components. This weighting scheme assigns relatively less weight to components that are highly correlated with one another, reducing redundancy and increasing the efficiency of the index. The resulting index is itself re-standardized to have mean zero and unit standard deviation in the control group, so that estimated treatment effects can be interpreted in standard deviation units.

C.2 Primary Outcomes

The primary outcomes combine into a single GBV Skills Index, itself composed of an Attitudes Index and a Technical Skills Index. The Attitudes Index aggregates five component indices: the victim-blaming, empathy, acceptance towards GBV, externalizing responsibilities, and truthfulness indices. The Technical Skills Index aggregates two component indices: the GBV identification and legal knowledge indices. We describe each component index and its underlying survey questions below.

C.2.1 Victim-Blaming Index

Victim-blaming involves devaluing the circumstances of an incident and holding an individual responsible for the circumstances that led to their own victimization. Blaming individuals for the circumstances or behaviors associated with crime victimization can discourage victims from reporting and engaging with a criminal investigation, for instance by lowering self-esteem or confidence, or by reducing a survivor's capacity to recall the circumstances of an incident. Victim-blaming therefore ultimately affects police performance by lowering the quality of questioning, investigation, and the likelihood that an FIR is filed. The endline victim-blaming index aggregates the following seven survey questions:

1. It is justified for a husband to hit his wife if she goes out without telling him.
2. It is justified for a husband to hit his wife if she neglects the children.
3. It is justified for a husband to hit his wife if she burns the food.
4. It is justified for a husband to hit his wife if she argues with the husband.
5. It is justified for a husband to hit his wife if she refuses to have sex with the husband.
6. (Referring to vignette) Out of 10 cases like this, in how many cases the woman faced such a situation because she did not behave in a socially acceptable manner?
7. (Referring to vignette) To what extent do you agree that this incident happened because Alka was traveling alone?

C.2.2 Empathy Index

The empathy index measures empathy among officers. Items 1 to 6 are drawn from the Empathy Quotient (Baron-Cohen and Wheelwright, 2004); items 7 to 9 are vignette-based perspective-taking items. The endline empathy index aggregates the following nine survey questions:

1. I tend to have very strong opinions about morality.
2. I find it easy to put myself in somebody else's shoes.
3. I am good at predicting how someone will feel.
4. I am able to make decisions without being influenced by others' feelings.
5. I can tune into how someone else feels rapidly and intuitively.
6. I can usually appreciate the other person's viewpoint, even if I don't agree with it.
7. (Referring to vignette) I can imagine what it must be like to be in Reena's place.
8. (Referring to vignette) I can imagine what it must be like to be in Alka's place.
9. (Referring to vignette) I can imagine what it must be like to be in Preeti Manjhi's / Varsha Gupta's place.

C.2.3 Acceptance towards GBV Index

This outcome measures officers' views on what forms of abuse constitute a GBV act, what should be tolerated by or not perpetrated by women, and how women should be expected to respond. Less equal norms about what women should tolerate and act upon regarding GBV could lower officers' effort and performance in handling such cases. The endline acceptance towards GBV index aggregates the following five survey questions:

1. (Referring to vignette) Out of 10 cases like this, in how many cases is recommending a compromise without filing an FIR enough to solve the issue?
2. Cases related to women receive too much attention by the police, relative to other crime and law and order issues.
3. (Referring to vignette) Now, thinking about the story, do you think that this consists of abusive behavior by Suresh?
4. The independence enjoyed by women these days is causing such problems.
5. Now, thinking about the story, do you think that Vijay Kumar is at fault in any way?

C.2.4 Externalizing Responsibilities Index

This outcome measures the extent to which officers externalize their responsibilities onto colleagues or victims to remove themselves from the process of justice. The endline externalizing responsibilities index aggregates the following six survey questions:

1. The woman does not report to the police because the police cannot help her and it is an internal matter between the husband and the wife.
2. The woman does not report to the police because such incidents are very common and happen too often.
3. The woman does not report to the police because she fears being beaten up by her husband again if she reports to the police.
4. (Referring to vignette) Do you think this is a problem that the police can help with?
5. Do you think the police can register a case against the men based on the complaint?
6. (Referring to vignette) Do you think this is a problem that the police can help with?

C.2.5 Truthfulness Index

Truthfulness of complaints refers to relying on instinct and intuition, rather than facts and logical rationale, to judge the veracity of a criminal complaint quickly. Such beliefs about the veracity of complaints are often described as gut feelings or hunches, and likely reflect a form of decision-making rooted in experience with past cases. While relying on experience is common to any profession, doing so may lead to an overestimation of false cases when such perceptions are formed on the basis of biased beliefs about victims and case types, in turn leading to the quick dismissal of reports or lower investigative effort. The endline truthfulness index aggregates the following nine survey questions:

1. (Referring to vignette) Out of 10 cases like this, in how many such complaints are false?

2. (Referring to vignette) Out of 10 cases like this, in how many are women pressing false charges related to sexual harassment in order to gain unfair advantage in land dispute cases?
3. (Referring to vignette) Out of 10 cases like this, how many such complaints are false?
4. (Referring to vignette) Out of 10 cases like this, in how many cases is the woman trying to frame the man after a fall-out of the relationship?
5. The complaint is usually more believable if the woman is also accompanied by her relatives rather than when she comes alone.
6. (Referring to vignette) The complaint filed by Reena is true.
7. (Referring to vignette) To what extent do you believe that Alka is speaking the truth?
8. (Referring to vignette) The complaint filed by Preeti Manjhi/Varsha Gupta is true.
9. (Referring to vignette) This is a typical case that I have seen being reported to the police station where a woman tries to frame a man.

C.2.6 Technical Skills Index

The Technical Skills Index measures officers' understanding of the correct procedures to employ in GBV cases. It is composed of two component indices: a GBV identification index, measuring officers' ability to correctly identify whether a described scenario constitutes a chargeable GBV offense and the applicable legal provision, and a legal knowledge index, measuring officers' familiarity with procedural requirements such as third-party filing of complaints on a victim's behalf and appropriate referral pathways. The endline Technical Skills Index aggregates the following eight survey questions:

1. (Referring to vignette) Can the sister file the complaint on the victim's behalf?
2. (Referring to vignette) With whom does the burden of proof lie?
3. If a married woman is evicted from her matrimonial house but has not been subjected to any other physical abuse, does that constitute domestic violence?
4. A female victim says she does not feel safe at her home and wants the police to assist her in finding shelter. Can the police assist with this?
5. A woman accuses a man who is an acquaintance for having verbally abused her in a marketplace. Is this a chargeable offense?
6. If yes, under what section of the IPC will you book him?
7. A police officer you work with has released information about a sexual abuse victim to the media. Is this a chargeable offense?
8. If the answer to the above question is yes, will you book him? If yes, under which section of the IPC shall you book him?

C.2.7 Reflection Index

The Reflection Index measures empathic concern and perspective-taking among police officers immediately following the intervention, the constructs targeted by the expressive arts intervention and the Theatre of the Oppressed methodology. Respondents rate how well each statement describes them on a five-point scale, from "does not describe me well" to "describes me very well." The Reflection Index aggregates the following fourteen survey questions, comprising the

Empathic Concern (EC) and Perspective-Taking (PT) subscales of the Interpersonal Reactivity Index (Davis et al., 1980):

Empathic concern.

1. I often have tender, concerned feelings for people less fortunate than me.
2. Sometimes I don't feel very sorry for other people when they are having problems. (R)
3. When I see someone being taken advantage of, I feel kind of protective towards them.
4. Other people's misfortunes do not usually disturb me a great deal. (R)
5. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (R)
6. I am often quite touched by things that I see happen.
7. I would describe myself as a pretty soft-hearted person.

Perspective-taking.

8. I sometimes find it difficult to see things from the "other guy's" point of view. (R)
9. I try to look at everybody's side of a disagreement before I make a decision.
10. I sometimes try to understand my friends better by imagining how things look from their perspective.
11. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments. (R)
12. I believe that there are two sides to every question and try to look at them both.
13. When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
14. Before criticizing somebody, I try to imagine how I would feel if I were in their place.

Questions marked (R) are reverse-coded. All questions are recoded so that higher values indicate greater empathic concern and perspective-taking. The index is constructed using inverse covariance weighting (swindex) and standardized to the control-group mean.

C.3 Secondary Outcomes

Secondary outcomes capture real-world policing behavior and organizational spillovers that are not observed in the officer survey. They are drawn from three sources: a standardized victim (decoy) audit, administrative First Information Report (FIR) records, and a survey of junior female constables. Unlike the primary indices, these measures are collected outside the officer survey; the audit records an officer's observed conduct in an unannounced interaction, and the constable survey is administered separately by female enumerators.

C.3.1 Standardized Victim Audit Outcomes

After each unannounced decoy visit, the standardized victim records the responding officer's observed conduct along the same dimensions the officer survey measures by self-report. Because these outcomes are based on observed behavior rather than self-report, they are less susceptible to social desirability bias. These measures are grouped by the construct they capture:

Victim-blaming. The following four survey questions record the officer's conduct:

1. Did the officer blame you or say that you were at fault?
2. Did the officer try to explain how you could have done things differently to avoid the situation?
3. Did the officer blame you for not reporting the case earlier?
4. Did the officer ask you unnecessary questions that were not relevant to the case (for example, "What were you wearing?", "Whom were you going with?", or "Why did you need to go outside?")?

Externalizing of responsibilities. The following three survey questions record the officer's conduct:

1. Did the officer tell you to visit a Mahila Thana (women's police station) for this case?

Views and attitudes towards GBV (GBV cases only). The following two survey questions record the officer's conduct:

1. Did the officer appear to be dismissive towards your case?
2. Did the officer make any of the following remarks: that such cases are usually a waste of the officer's time; that most such cases are false; that such cases should only be dealt with in a Mahila Thana; or that women invite these things upon themselves?

C.3.2 Administrative Records

We draw on administrative FIR data for the sample stations from the Bihar State Crime Records Bureau. From these records we construct two station-level outcomes: the number of days between a reported incident and its formal registration, and the total volume of GBV First Information Reports filed at a station over time. Because these records are generated independently of the research team as part of routine police record-keeping, they are not subject to self-report bias.

C.3.3 Workplace Culture Outcomes

To assess whether shifts in senior officers' attitudes translate into a changed workplace environment, we survey junior female constables at each sample station. The survey is administered exclusively by female enumerators. We measure three outcomes: an anxiety score, reported harassment, and perceptions of male officers.

Anxiety (GAD-7). The following seven survey questions record how often, over the past two weeks, the respondent was bothered by:

1. Feeling nervous, anxious, or tensed or worried.
2. Not being able to stop or control worrying.
3. Worrying too much about different things.
4. Trouble relaxing.
5. Being so restless that it is hard to sit still.
6. Becoming easily annoyed or irritable.
7. Feeling afraid as if something awful might happen.

Reported harassment. The following six survey questions record whether a senior officer has ever:

1. Made her feel threatened or unsafe by shouting, scolding, or reprimanding her loudly.
2. Made her feel intimidated or threatened that he or she could physically harm her by pushing or hitting her.
3. Hit, slapped, punched, tripped, or otherwise intentionally caused her physical harm.
4. Made sexual advances such as touching her inappropriately.
5. Requested her to meet alone with an officer in a way that made her feel uncomfortable.
6. Made remarks about her in a sexual manner or shown her inappropriate pictures or videos.

Perceptions of male officers. The following five survey questions record agreement with statements about senior male officers (of rank ASI and above) in the station:

1. Male officers are understanding of my family responsibilities.
2. I had no problems taking maternity leave when I was expecting.
3. Male officers have expressed that they do not like to work with female officers.
4. Male officers make it difficult for me to conduct my job.
5. Male officers are unaware of the challenges female officers face when joining the force.

D Standardized Victim Audit: Sampling, Protocols, and Ethical Safeguards

This Appendix provides the full sampling frame, training, implementation protocol, and ethical safeguards underlying the standardized victim audit described in Section 4.1. The audit supplements the self-reported officer survey with a measure of real-world behavior collected without officers' knowledge that they are being observed, addressing concerns about social desirability bias and Hawthorne effects that self-report alone cannot rule out.

D.1 Sampling

The audit was conducted in the six districts where the district Superintendent of Police showed clear support for the exercise: Bhojpur, Gopalganj, Muzaffarpur, Nalanda, Sitamarhi, and Siwan. Table A18 reports the number of sample stations and decoy visits by district. Each station received three visits over a pre-agreed three-month window for a total of 549 visits across 183 stations, approximately 44 percent of the full 417-station sample. Because randomization was conducted district-by-district in a staggered manner, each audit district contains a roughly equal share of treatment and control stations with balanced observable characteristics across the two arms, so decoy visits are not concentrated in either arm.

D.2 Standardized Victims and Case Assignment

Standardized victims (SVs) were trained to attempt to register a case at a police station in the same manner as an actual complainant, following the approach of Banerjee et al. (2021) and Das et al. (2016). With the support of DAI and professional theatre practitioners, the research team developed intensive training exercises that simulated the range of situations SVs could encounter at a police station. These exercises prepared SVs to respond naturally while adhering to standardized

Table A18: Police Station Visits by District (Standardized Victim Audit)

District	Police stations	Visits per station	Total visits
Bhojpur	35	3	105
Gopalganj	20	3	60
Muzaffarpur	38	3	114
Nalanda	39	3	117
Sitamarhi	24	3	72
Siwan	27	3	81
Total	183		549

scripts and predefined responses to likely follow-up questions, thereby ensuring comparability across visits.

Both the case presented and the timing of the visit within the three-month window were randomized in advance by the research team. Cases were deliberately minor and non-serious, while still meeting the legal threshold for an on-the-spot *kaccha FIR* (pre-FIR) or a full FIR. SVs were dressed and briefed to match the local context and the circumstances of their assigned case; female SVs were accompanied by male SVs when appropriate.

Within one hour of leaving the police station, the SV completed a structured questionnaire recording the officer's response and the quality of the interaction. A subset of interactions was also audio-recorded. The audit followed a double-blind protocol during the interaction: SVs were not informed of the station's treatment assignment, and officers did not know that the complainant was an SV. The SV's identity was revealed only at the conclusion of the interaction, following the protocol described in Appendix D.3.

D.3 Implementation Protocol

Implementation followed a five-part protocol.

1. **Permissions and approvals.**The Superintendent of Police (SP) in each district issued a signed letter to every sample police station indicating that a standardized visit could occur within the following three months. The letter did not disclose the identity of the standardized visit or the timing of the visit. Such notifications were part of routine administrative practice, as police stations frequently received advance notices of official visits. The broad three-month window was intended to preserve uncertainty about the timing and identity of the visit while ensuring the necessary administrative authorization for the visits.

2. **Randomized timing.** The timing of each visit was determined by the research team and staggered across waves to limit cross-station communication once the SV identity became known at a given station.

3. **Case presentation.** At each visit, the SV presented one of a small set of pre-assigned minor GBV or non-GBV cases, with consistent answers prepared for likely follow-up questions.

4. **Recording the interaction.** Immediately after the visit, the SV completed a structured exit survey (within one hour) recording the officer's conduct, the quality of the interaction, and whether the officer appeared to suspect the visit was a decoy.

5. **Decoy reveal.** To prevent an official record of a false case, the SV revealed their identity to the officer at the end of the interaction and presented the SP's official letter.

D.4 Mitigating Risks and Ethical Safeguards

Five safeguards governed the audit, designed to protect both the SVs and the integrity of the exercise.

1. **Safety of the SVs.** The research team maintained close coordination with the district Superintendent of Police, and each visit was supervised by a dedicated field team.
2. **Confidentiality.** The identities of both the SV and the responding officer were anonymized, and all data were stored on encrypted servers with access limited to authorized project staff.
3. **Avoiding fraudulent case registration.** SVs were trained to disclose the nature of the visit to the officer before departing, and to confirm that no official record, online or offline, of the presented case remained on file.
4. **Psychological support for SVs.** Given the emotionally demanding nature of presenting sensitive scenarios, psychological support and counseling services were made available to SVs before, during, and after each interaction.
5. **Debriefing officers.** SVs conducted a short debriefing with the responding officer after each interaction, framing the exercise explicitly as an assessment of the station rather than an evaluation of the individual officer, to address any residual concerns.

E Cost and Benchmarking: Additional Detail

This appendix reports the full cost breakdown underlying the summary figures in Section 7, reports non-monetized cost-effectiveness measures, discusses why we do not construct a monetized cost-benefit ratio for GBV outcomes, and reports the full benchmarking comparison against comparable interventions.

E.1 Program Cost

Table A19 reports the budgeted and actual cost of each component of the intervention (Panel A) and each component's share of the actual total, alongside the related research and evaluation cost categories that are funded separately and excluded from this figure (Panel B). The training of trainers (ToT) prepared the pool of facilitators who later delivered the curriculum in the field. The main intervention cost reflects the three-day training sessions delivered to officers across the 217 treatment stations. The follow-up cost reflects a one-day refresher session delivered after the main training. The monitoring cost reflects the ongoing supervision of facilitators during implementation, including weekly reflection calls, monitoring visits by the lead trainer, and facilitator grading, intended to safeguard the fidelity of curriculum delivery across batches and districts. The Panel B categories were funded separately through the J-PAL Crime and Violence Initiative and the J-PAL Gender Initiative and reflect the cost of the research and evaluation activities surrounding the intervention, not the cost of delivering the intervention itself.

Table A19: Cost of the Expressive Arts Intervention

	Budgeted Cost (EUR)	Actual Cost (EUR)	Share of Actual Total
Panel A: Costs Included in This Figure			
Training of trainers	26,162.39	26,564.53	19%
Main intervention	50,199.01	49,380.48	36%
Follow-up session	35,407.24	35,192.44	26%
Monitoring	24,509.87	25,422.53	19%
Total	136,278.52	136,560.00	100%
Panel B: Related Costs Funded Separately (Not Included Above)			
Personnel at University (PIs, RMs/PMs/RAs)		Funded separately	
Fringe Benefits at University		Funded separately	
Travel (PI, etc.)		Funded separately	
Materials and Supplies at University		Funded separately	
Field Costs (personnel, travel, survey costs, and materials)		Funded separately	
Indirect Costs		Funded separately	

Notes: Source: Fund for Innovation in Development, program financial records. At the exchange rate implicit in the funder's own reporting (approximately EUR 1 = USD 1.11), the actual total corresponds to approximately USD 151,600. Panel B categories were funded separately through the J-PAL Crime and Violence Initiative and the J-PAL Gender Initiative and are not included in the total above. The training of trainers and monitoring components in Panel A (38% of the total combined) are largely fixed costs of establishing and maintaining facilitator quality, while the main intervention and follow-up components (62% combined) scale more directly with the number of officers trained.

Because 38 percent of total program cost reflects training-of-trainers and monitoring activities that contain substantial fixed-cost elements, rather than costs that scale one-for-one with the number of officers trained, the per-officer cost of the intervention would likely fall if implemented at substantially larger scale.

The program trained approximately 1,400 officers across the 217 treatment stations. Using the funder's own reported figures (136,560 EUR total, 1,405 officers), the actual cost of the program was approximately EUR 97.20 (USD 107.89) per officer trained, modestly above the budgeted EUR 97.00 (USD 107.66) per officer.[22] At the level of police stations, the total actual cost of EUR 136,560 implies an average cost of approximately EUR 629 (USD 699) per treated station. Assuming population is distributed evenly across the study's randomized stations, the 217 treated stations serve approximately 21.8 million of the 42 million residents covered by the full study sample.[23] This implies that the total cost of the intervention corresponds to less than one US cent per resident of the area served. From a policy perspective, this population-served figure is particularly informative: although the training reaches officers directly, the population that stands to benefit from improved policing is far larger.

[22]This per-officer figure reflects the funder's own denominator of 1,405 officers trained. If program administrative records indicate a materially different exact count, this figure should be recomputed directly from that count rather than reused from the funder's calculation.

[23]This population figure is derived by scaling the full sample's reported population of 42 million by the share of stations assigned to treatment (217 of 417), and assumes population is distributed evenly across stations. It is not based on station-level population data.

E.2 Cost-Effectiveness

Because we do not construct a monetized value for improvements in GBV outcomes (Section E.3), we cannot report a benefit-cost ratio. We can, however, report the cost of achieving a given magnitude of measured effect, which does not require monetization and is a standard complement to benefit-cost analysis when a defensible monetary value for the outcome is unavailable.

At a program cost of USD 107.89 (approximately USD 108) per officer trained and a measured improvement of 0.082 standard deviations in the GBV attitudes index, the implied cost per standard deviation of attitude change is approximately USD 1,320.[24] Because the combined GBV skills index (attitudes and technical-legal knowledge) improved by a similar 0.082 standard deviations, the same cost per standard deviation applies to that broader index. This figure should be interpreted primarily as a within-study benchmark for comparing across the interventions in Table A21, rather than as an absolute welfare metric or a claim about whether the cost per standard deviation represents good or poor value in any general sense. It is also a survey-attitude metric specifically: because the comparators that report cost do not report the cost of achieving a given change in observed case-handling behavior or in administrative case registration, we cannot extend this cost-per-effect calculation to the behavioral and administrative outcomes documented in Section 6.

This figure is naturally sensitive to how long the measured effects persist. The cost figures above implicitly treat the training as a one-time expenditure whose full USD 108 per-officer cost must be justified by effects observed within our study's follow-up window (approximately one year for the behavioral outcomes measured via decoy audits, and roughly two and a half years for the administrative case-registration outcomes). If instead the attitude and behavioral changes documented in this paper persist over a longer horizon without decay, the annualized cost per officer falls proportionally. Table A20 illustrates this under three illustrative scenarios, each of which assumes that effects persist at their full measured magnitude for the stated number of years and then are not tracked further, rather than assuming any particular pattern of decay within that window. We do not have the longer-run follow-up data needed to determine which scenario applies; we present these as illustrative bounds rather than as an estimate of actual persistence.

Table A20: Illustrative Annualized Cost Under Alternative Persistence Assumptions

Scenario	Assumed duration of effects	Annualized cost per officer
A	1 year	USD 108
B	3 years	USD 36
C	5 years	USD 22

Notes: Annualized cost is calculated as the actual per-officer cost of USD 107.89 divided by the assumed number of years over which effects persist. These calculations assume effects persist at their full measured magnitude without decay for the stated duration; they are illustrative only, we do not have follow-up data beyond approximately two and a half years for the administrative outcomes, and we do not claim that any particular scenario is more likely than another. Cost-effectiveness improves mechanically under this calculation if measured effects persist beyond the study's follow-up window without decaying, and worsens if effects decay faster than one year.

[24] We report this figure to one significant digit given the imprecision of both the underlying cost figure and the estimated effect size. It should not be interpreted with more precision than the underlying data support.

E.3 Why We Do Not Monetize Benefits

GBV has no single agreed-upon benchmark linking attitudinal or behavioral improvements to an economic value, and existing estimates of the aggregate economic cost of violence against women vary too widely to support a defensible conversion factor. Puri (2016) estimates the global cost of violence against women at approximately USD 1.5 trillion, or roughly 2 percent of global GDP. Using a different methodology over a similar period, Hoeffler and Fearon's assessment for the Copenhagen Consensus Center estimates the global cost of intimate partner violence alone at USD 4.4 trillion, or 5.2 percent of global GDP (Hoeffler and Fearon, 2014); the same authors' subsequent estimate of the broader category of interpersonal violence (adding homicide and child abuse to IPV) rose to approximately USD 30 trillion annually (Hoeffler and Fearon, 2026). A systematic review of cost-of-violence studies in low- and middle-income countries similarly concludes that methodological inconsistency substantially limits comparability across studies (Vyas et al., 2023). Closer to our own setting, a cost-benefit analysis prepared for the Andhra Pradesh state government estimates the loss to state GDP from prevailing IPV rates at approximately 0.19 percent of gross state domestic product (Raghavendra et al., 2018), smaller by more than an order of magnitude as a share of output than the global estimates above, illustrating how sensitive these figures are to scope and methodology.[25]

Given this range, constructing a monetized cost-effectiveness ratio for our study would require choosing among substantially divergent benchmarks, a choice that would be largely arbitrary given the current state of the literature. A monetized benefit-cost ratio has been constructed in a related Indian setting, for a college sexual-harassment training, by drawing on employer-side estimates of the productivity cost of workplace harassment (Sharma, 2024); that approach, however, rests on a workplace-productivity cost base that does not transfer to the violence outcomes and case-registration behaviors we study, where the relevant costs fall on survivors and the justice system rather than on employers. We instead report the cost-effectiveness measures in the preceding subsection and the benchmarking comparison below.

That we do not attach a dollar value to these improvements does not make their welfare significance ambiguous. The changes we document, more appropriate treatment of survivors during case handling, faster and more frequent formal registration of GBV cases, and a safer workplace for junior female officers, are the proximate links in a causal chain that experimental evidence connects to women's welfare. In a companion police-patrol experiment in Hyderabad, where uniformed patrols reduced severe public-space sexual harassment, women became roughly 30 percent less likely to move elsewhere in response to it, easing the constraints on physical mobility that are linked to women's schooling and labor force participation (Amaral et al., 2025). That study further identifies police visibility and officers' attitudes toward harassment as central to the reduction, the same attitudinal margin our intervention targets directly. Read alongside our own findings, this places the attitudinal and behavioral changes we measure on a path whose endpoint is women's safety and economic participation, even though we do not attempt to price that endpoint.

[25] Bihar itself has the highest state-level prevalence of physical intimate partner violence in India, at 35.1 percent of ever-married women (Srivastava et al., 2023), reflecting the practical importance of the outcomes we study even though we cannot attach a defensible dollar value to them.

E.4 Benchmarking Against Comparable Interventions

We are not aware of published evidence reporting both costs and impacts for a police-focused gender-based-violence *training* program specifically. The closest police comparator, Sukhtankar et al. (2022), evaluates a gender-targeted police reform in Madhya Pradesh, India (the introduction of Women's Help Desks) and speaks directly to the same class of administrative outcome we study, case registration, but it evaluates a bundled reform rather than a stand-alone training and does not report costs on a comparable basis, so it supports a comparison of effect sizes on the registration outcome but not a joint cost-and-impact comparison. This limits the scope for benchmarking our intervention against the most policy-relevant comparators on a common cost basis, and should be read as a limitation of the existing evidence base rather than of this study. A further complication for benchmarking is that our study measures effects at three distinct levels: (i) officers' attitudes and technical-legal skills, measured by survey; (ii) observed case-handling behavior toward survivors, measured by standardized decoy audits approximately one year after training; and (iii) administrative case registration, measured by state crime records over roughly two and a half years. No single comparator speaks to all three levels. Table A21 therefore benchmarks each of our outcome levels against the most relevant available evidence: the police Women's Help Desk reform (Sukhtankar et al., 2022) for the registration outcome; the Sit-D cognitive training for Chicago police (Dube et al., 2025) for the change in officers' field behavior; a school-based gender-attitudes curriculum in India (Dhar et al., 2022) and a men's sexual-harassment training in Indian colleges (Sharma, 2024) for the attitudes and behavior-change outcomes; and a community mobilization program in Uganda (Abramsky et al., 2014) as broader behavioral-training benchmarks that report both cost and effect-size data.

Table A21: Benchmarking Effect Sizes and Costs Against Comparable Interventions

Study	Setting & population	Duration	Outcome & effect size	Cost
This study	Senior police officers, Bihar, India (417 stations, 217 treated; ∼1,400 officers trained; 42M pop.)	3-day training + 1-day follow-up	*Attitudes/skills*: attitudes +0.082 SD, technical skills +0.085 SD, combined GBV skills +0.082 SD. *Decoy audits* (∼1 yr): less case dismissal, false-case labeling and victim-blaming; more referral and explanation. *Admin FIRs* (∼2.5 yrs): registration latency −41%; long-run rise in GBV reporting	USD 108 / officer
Sukhtankar et al. (2022)	Police stations with Women's Help Desks, Madhya Pradesh, India (180 stations, 23.4M pop.)	Bundled reform (help desk, training, monitoring, outreach), ∼11 months	FIRs in crimes-against-women cases: +14.1% ($p = 0.08$); +26.4% in woman-run desks; FIRs filed by women: +10.4% ($p = 0.1$); DIRs from ≈0 to ∼1.5/month; no shift in male officers' gender attitudes	Not reported

Continued on next page

Table A21 continued

Study	Setting & population	Duration	Outcome & effect size	Cost
Dube et al. (2025)	Police officers, Chicago, USA (2,070 officers)	16 hours (4 four-hour sessions)	Uses of non-lethal force −23%; discretionary arrests −23% (concentrated among Black civilians); arrests of Black civilians −11%; fewer officer injury days; effects diminish over the year	USD 807–864 / officer trained
Sharma (2024)	Male undergraduate students, three colleges, Delhi, India	Single 3–5 hour in-person session	Reported sexual harassment: −0.06 SD short run (extreme forms −0.125 SD, sig. 1%); persists at −0.09 SD after 2–3 years; awareness +0.07 to +0.09 SD; operates via perceived social disapproval more than intrinsic attitudes	∼USD 16–91 / woman avoiding harassment (benefit-cost ratio 12–509)
Dhar et al. (2022)	Adolescents, government schools, Haryana, India	2.5 years, 28 classroom sessions	Gender attitudes index: +0.18 SD (converts 16% of regressive attitudes); persists 2 years post-program	Not reported in a finalized per-participant figure
Abramsky et al. (2014); Michaels-Igbokwe et al. (2016)	Community members (women and men, 18–49), Kampala, Uganda	∼4-year community mobilization (SASA!)	Social acceptance of IPV among women: adjusted RR 0.54 (significant); past-year physical IPV among women: adjusted RR 0.48, i.e., 52% lower (95% CI 0.16–1.39, not significant)	USD 389 / activist / year; USD 21 / person reached over full program

Notes: SD: standard deviations. RR: risk ratio. CI: confidence interval. FIR: First Information Report. DIR: Domestic Incident Report. Effect sizes and costs are drawn directly from the cited sources and use each study's own outcome measures and cost accounting; they are not necessarily constructed on a directly comparable basis. The comparators are grouped by the outcome level of our own study to which each is most relevant: Sukhtankar et al. (2022) for administrative case registration; Dube et al. (2025) for the change in officers' field behavior; Dhar et al. (2022) and Sharma (2024) for attitudes and behavior change. Sukhtankar et al. (2022) evaluates a bundled reform, not a stand-alone training, and reports no cost on a comparable basis. Sharma (2024) reports cost per woman who avoids harassment rather than cost per participant trained, a different denominator from our per-officer figure. None of the non-police comparators targets police officers or a comparable professional population; they represent the closest available published evidence combining cost and effect-size data, not close substitutes for a policy decision about police training specifically.

On input cost, the Bihar program is inexpensive in absolute terms, at approximately USD 108 per officer for a three-day curriculum plus follow-up and fidelity monitoring. The closest police-training comparator that reports cost, the 16-hour Sit-D program for Chicago police, cost USD 807 to 864 per officer trained (Dube et al., 2025), and the Liberia cognitive-behavioral-therapy benchmark cost USD 530 per participant (Blattman et al., 2017); these gaps are dominated by differences in local wage levels, program length, and target population rather than by efficiency, so we do not read them as like-for-like.

On the attitudes outcome, our three-day training produced an improvement in officers' GBV attitudes of approximately 0.082 standard deviations, roughly half the effect size of the 2.5-year, 28-session adolescent curriculum evaluated by Dhar et al. (2022). Given that adult professionals' attitudes are generally understood to be less malleable than adolescents', and that our intervention required a small fraction of the contact time of the Haryana program, we read this as suggestive evidence that a short, immersive training can generate a meaningful share of the attitude change achieved by a substantially longer program targeting a more malleable population. The single-session men's training studied by Sharma (2024), closer to ours in brevity, produced awareness gains of 0.07 to 0.09 standard deviations, comparable in magnitude to our own gains in officers' technical and legal knowledge (0.085 SD). Dhar et al. (2022) do not report a finalized per-participant cost, and Sharma (2024) reports cost per woman avoiding harassment rather than per participant, so we cannot extend either comparison to a joint cost-effectiveness ranking; and our GBV skills index combines attitudinal and technical-legal knowledge components, while the Haryana program's index measures attitudes alone, so the effect-size figure we cite in the attitudes comparison reflects the attitudes component (0.082 SD) specifically.

On the behavioral and administrative outcomes, which are the outcomes of most direct policy interest, the closest comparator is the Women's Help Desk reform of Sukhtankar et al. (2022). That reform raised registration of crimes against women (FIRs +14.1% overall, +26.4% in woman-run desks) and introduced Domestic Incident Reports where they had been effectively absent, but it left male officers' gender attitudes unchanged and worked primarily through the assignment of female officers and dedicated help desks. Our intervention, by contrast, targets officers' attitudes and skills directly and traces the resulting behavioral change through decoy audits (lower case dismissal and victim-blaming, higher referral and explanation) and through administrative records (a 41 percent reduction in the lag between a reported GBV incident and its formal registration, together with a long-run rise in reporting). The two interventions differ in design (a training versus a bundled structural reform), in state, and in the exact outcome definitions, and Sukhtankar et al. (2022) reports no cost, so the comparison establishes that both approaches can move the same administrative margin rather than a ranking of their relative cost-effectiveness.

A closer precedent in form is the Sit-D cognitive training for Chicago police (Dube et al., 2025): a short behavioral training for serving officers that changed field behavior, reducing uses of force and discretionary arrests by about 23 percent, and, unlike the Women's Help Desk reform, reporting a per-officer cost (USD 807 to 864). It targets uses of force rather than GBV, and its cost reflects a high-wage US setting, so it speaks to the plausibility of training-induced behavioral change in policing rather than offering a like-for-like comparison.

In sum, the expressive-arts training was inexpensive to deliver relative to the comparators that report cost, produced attitude change of a magnitude that is meaningful relative to both a substantially longer program targeting a more malleable population and a comparably brief training in a different setting, and produced a rise in the formal registration of women's cases, the same broad administrative margin moved by the closest comparable police-reform study, while additionally reducing the delay between an incident and its registration. Establishing whether these effects persist over a longer horizon would strengthen the case for scaling this approach, as would the emergence of police-specific benchmarking evidence that reports both cost and impact. Even without a monetary valuation, the cost-of-violence estimates reviewed above, however much they diverge in magnitude, all point to an economically substantial problem; an intervention that shifts frontline officers' attitudes at modest cost is therefore a policy lever worth taking seriously at scale.