Check for updates

Article

Constrained but Not Transformed: Civilian-Led Certification Reform and Officer Misconduct

Political Research Quarterly
2026, Vol. 79(1) 110–125
© The Author(s) 2025

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10659129251382162
journals.sagepub.com/home/prq
Sage

Amanda Lanigan[1] and Ilker Kalin[2,3]

Abstract

In the wake of national protests against police violence, Massachusetts established a civilian-led certification regime to institutionalize external oversight of law enforcement. This study evaluates the short-term behavioral effects of that reform on officers with documented histories of misconduct. Leveraging the alphabetically phased rollout of re-certification, we use a natural experiment and difference-in-differences design to estimate the causal impact of POST participation. While our quantitative analysis does not reveal sweeping changes in the frequency of officer misconduct, it uncovers suggestive evidence of behavioral recalibration: officers appear less likely to engage in misconduct at all following exposure to external review. To contextualize these findings, we draw on semi-structured interviews with officers, supervisors, and oversight officials. These interviews reveal how officers often reject the legitimacy of POST while simultaneously adjusting behavior in response to reputational risk and the threat of sanction. Taken together, the results suggest that externally adjudicated, procedurally binding reforms can induce modest, short-term behavioral change through instrumental compliance.

Keywords

police misconduct, certification reform, natural experiment, civilian oversight

Introduction

In 2025, the Trump administration took the National Law Enforcement Accountability Database offline, dismantling one of the few national tools for tracking misconduct among federal law enforcement officers (Kaste 2025). As efforts to erode federal accountability mechanisms continue, the recent expansion of state-level officer certification represents a primary means of surveilling patterns of police violence in state and local jurisdictions. Yet despite their expansion, we know little about how these interventions shape officer behavior in practice.

This question has demanded renewed attention since the 2020 murder of George Floyd, which catalyzed the largest protest movement in U.S. history and revived long-standing demands for police reform dating back to the killing of Henry Truman, an unarmed Black man murdered by police in 1870 (Blow 2023). Despite this momentum, patterns of coercive state violence persist. More people were killed by police in 2024 than in any other year over the past decade (Mapping Police Violence)[1] with people of color bearing the brunt of violence at the hands of police (Ward et al., 2024). The persistence of police violence raises fundamental questions about institutional capacity to constrain abuse, especially in contexts where internal accountability is weak and organizational cultures resist reform.

As of 2021, thirty-seven states have adopted or strengthened centralized certification regimes, reflecting a broader institutional shift towards the use of professional licensing mechanisms to impose state-administered oversight on decentralized frontline actors (NCSL; Subramanian and Arzy 2021). But whether such mechanisms can meaningfully recalibrate behavior within coercive bureaucracies remains unresolved.

[1]Department of Peace and Conflict Research, Uppsala University, Uppsala, Sweden
[2]Institute for Turkish Studies, Stockholm University, Stockholm, Sweden
[3]Institute for Democracy, Matej Bel University, Banská Bystrica, Slovakia

Corresponding author:
Amanda Lanigan, Department of Peace and Conflict Research, Uppsala University, Box 514751 20, Uppsala, Sweden.
Email: amanda.lanigan@pcr.uu.se

This study turns to a recent recertification reform in Massachusetts to examine how state-level certification affects agent-level behavior in coercive bureaucracies. While research has examined the broader package of public safety reforms implemented in Massachusetts and their effects on police behavior (Cassino and Demir 2024), we investigate whether civilian-led licensing authorities matter for officer compliance with reform mandates, and how formal oversight mechanisms influence the underlying logics of discretion and risk that govern officer decision-making.

Massachusetts offers a unique opportunity to study this question. The 2020 creation of the Peace Officer Standards and Training (POST) Commission marked the first independent, statewide oversight body in Massachusetts tasked with evaluating officer fitness through mandatory, legally binding, periodic recertification. The reform aimed not only to standardize accountability but also to institutionalize civilian oversight by linking continued employment to documented professional conduct. It provides a rare opportunity to assess whether officers perceive certification by an external, civilian-dominated body with sanctioning power as a credible constraint, and whether that perception is sufficient to prompt behavioral change.

To evaluate the effects of POST reform, we begin by leveraging a publicly available police misconduct dataset and exploiting the statewide, alphabetically phased rollout of the recertification process as a natural experiment. By comparing officers recertified in the first wave to those who have not yet participated in recertification, we estimate the short-term effects of participation in recertification on observed misconduct. From there, we conducted semi-structured interviews with members of law enforcement, as well as state and city oversight officials in Spring of 2025. These interviews provide insight into how the certification process is understood and navigated in practice. They shed light on how officers perceive its legitimacy and risks, how administrators implement it, and how both frontline and supervisory actors interpret external oversight in relation to established norms of discretion and accountability.

Together, these methods allow us to examine the behavioral consequences of certification at the officer level. Our findings suggest that external and centralized state-level certification regimes constrain officer misconduct to a limited degree, but do not transform policing norms. They prompt short-term shifts in the degree to which officers modify behavior in response to increases in scrutiny and sanctions.

Quantitatively, we identify a subtle but important distinction: while recertification does not significantly reduce the number of monthly misconduct incidents among offending officers, it does lower the likelihood that officers engage in any misconduct at all. This reduction is concentrated at the extensive rather than intensive margin. These effects are modest but robust across model specifications and appear slightly more pronounced among officers with more extensive prior misconduct records. Interviews help explain these patterns. While many officers perceive POST as a politicized or redundant reform, they nonetheless adjust their behavior in response to the perceived threat of decertification and reputational exposure. Officers who underwent recertification reported greater caution and risk aversion, especially in ambiguous situations. At the same time, interviews with officers across ranks and roles suggest these behavioral shifts are driven by fear of sanction rather than normative acceptance of external oversight and its aims.

Our study contributes to broader debates on the governance of coercive bureaucracies, offering new empirical evidence on whether formal oversight mechanisms can change behavior in insular institutions, and more broadly, on the possibilities and limits of democratic constraint over state violence.

Literature Review

Centralized accountability mechanisms are increasingly used to constrain abusive policing, yet their effects on officer behavior remain uneven and contested. Previous research offers varying explanations for what makes bureaucratic accountability effective (Eriksen 2021). Some scholars emphasize the role of formal authority and the visibility of oversight mechanisms (Bovens 2007, 2010; Grimmelikhuijsen et al., 2017), while others highlight officers' instrumental risk calculations and their normative perceptions of legitimacy (Engel and Worden 2003; Flom 2024; González 2023; Mummolo 2018; Sheffer and Loewen 2019). Still other work focuses on how reforms are mediated by organizational dynamics—through supervisory enforcement, institutional resistance, or symbolic compliance (Brehm and Gates 1994; Brehm and Scott 1997; Clark 2024; González 2019; Lipsky 1980). Yet due to the insular nature of policing and the lack of data critical questions persist about how oversight is internalized, by whom, and with what behavioral effects.

Scholars of public administration have long recognized that street-level bureaucrats pose enduring challenges to public accountability and constituent constraint (Flom 2020, 2024; Lipsky 1980). In policing, this challenge is especially acute. Officers operate with high autonomy and low visibility, their behavior shaped as much by political dynamics, informal norms, individual biases, and peer expectations as by formal rules (Clark 2024; Flom 2020, 2024; Mastrofski 2004; Prottas 1978; Terrill and Mastrofski 2002). As principal-agent theorists argue, agents do not automatically comply with institutional mandates but respond strategically to the credibility of sanctions and the alignment between oversight and

professional identity (Brehm and Gates 1994; Brehm and Scott 1997; Carpenter 1996). In high-impunity institutions like the police, where agency goals diverge from democratic ones, compliance cannot be assumed.

Attempts to constrain officer discretion range from internal affairs units and civilian review boards, to civil litigation, to accreditation programs and certification regimes (Bird and Shannon 2024; Ferdik, Rojek, & Alpert 2013; Filstad and Gottschalk 2011; Johnson 2015; Perez 2000). Yet many of these mechanisms have proven limited in scope or effect (Lawson 2019). Voluntary accreditation often reflects self-selection into reform rather than meaningful change (Johnson 2015; Teodoro and Hughes 2012). Civilian oversight bodies, widely adopted in U.S. cities, tend to lack formal power and are frequently co-opted by police departments themselves (Ali and Pirog 2019; Alpert et al., 2015; Clark 2024; Ferdik, Rojek, & Alpert 2013; Lewis and Prenzler 2017; Prenzler 2000; Rocha Beardall 2022), but can improve the efficiency of arrests and positively affect citizen associations with procedural justice dependent upon their degree of scope and authority (Ali and Nicholson-Crotty 2020; Holliday and Wagstaff 2022; Idrobo, Kronick, & Slough 2024). At the same time, research finds that even when reforms increase scrutiny and sanctioning power more broadly, their behavioral impact is often filtered through the informal institutional structures and collective bargaining protections that govern everyday policing (Biesta 2010; Rad, Kirk, & Jones 2023; Thomas and Tufts 2020), and speaks to the hurdles civilians and elected officials face in constraining the public sector (Fung 2015).

Recent studies highlight the importance of institutional design in shaping the behavioral impact of accountability mechanisms. Interventions that are external, binding, and recurring are more likely to influence officer conduct than those that are internal, advisory, or reactive (Ali and Nicholson-Crotty 2020; Rozema and Schanzenbach 2020). Yet as González (2019, 2023) shows, external oversight reforms are politically contested because they shift disciplinary power away from the profession. Officers and supervisors often resist externally imposed mandates that conflict with occupational norms or appear politically motivated." The result is often symbolic compliance or selective enforcement, mediated by supervisors who interpret and implement reforms on their own terms (Brehm and Gates 1994; Engel and Worden 2003).

Empirical evidence on behavioral outcomes remains mixed. Some studies find that credible oversight reduces misconduct and improves compliance when mechanisms alter risk calculations (Mummolo 2018; Rivera and Ba 2019). Others highlight uneven and racially stratified effects. Bird and Shannon (2024), for instance, show that federal lawsuits against police agencies in North Carolina lead to significant reductions in discretionary traffic stops, but only for white drivers. Their findings suggest that legal accountability alters officer behavior in selective and often blunt ways, and that discretion is not evenly constrained; it is recalibrated across discriminatory lines, revealing the durability of racialized enforcement even under external pressure (Epp, Maynard-Moody, & Haider-Markel 2014).

These findings raise broader questions about the behavioral reach of accountability. While some officers may adjust conduct in response to credible oversight, others remain shielded by institutional norms, weak enforcement, or organizational complicity. Misconduct is known to be concentrated among a subset of repeat offenders, often sustained by peer protection and lack of formal sanction (Ingram, Terrill, & Paoline 2018; Ouellet et al., 2019) Individual-level predictors—such as tenure, self-control, or social dominance orientation—further shape responsiveness to reform (Davis, Baluran, & Hassan 2024; Donner 2019; Donner and Jennings 2014). The persistence of abuse, despite recurring reform efforts, reflects not just gaps in policy but institutional insulation from and resistance to democratic constraint (González 2019, 2023).

Taken together, the literature underscores both the promise and the limitations of current accountability interventions. While research has advanced important theoretical insights into how officers interpret and respond to oversight, few studies systematically examine how centralized, recurring, and externally imposed mechanisms, such as state-mandated certification, affect officer behavior across different organizational and individual contexts. The behavioral consequences of these reforms remain poorly understood, particularly in coercive bureaucracies where professional norms and institutional protections blunt the force of formal sanction. Moreover, little is known about how officers internalize such reforms, or whether these mechanisms are capable of shifting the informal logics of risk, discretion, and peer solidarity that govern frontline decision-making. Motivated by these gaps, this study leverages the unique opportunity offered by the implementation of POST to gain traction on how participation in a recertification process affects officer behavior and perception. Understanding these dynamics sheds light on why democratic governments continue to struggle with institutionalizing control over police behavior—and why misconduct persists despite repeated waves of reform.

Background of the POST Commission and Massachusetts Policing Reform Bill

Although often seen as exempt from national debates on police violence, Massachusetts reflects many of the same institutional challenges. As of 2023, over 10 percent of Massachusetts officers had sustained allegations of

serious misconduct, including excessive force, discrimination, and incidents related to serious injury or death (Becker and Jarmanning 2023).

In 2020, responding to national and local calls for accountability, Massachusetts passed *An Act Relative to Justice, Equity and Accountability in Law Enforcement in the Commonwealth*, creating the Peace Officer Standards and Training (POST) Commission—the state's first independent body with authority to certify, suspend, and decertify officers. POST marks a significant departure from earlier oversight mechanisms, which largely relied on internal disciplinary procedures or external civilian boards limited to nonbinding recommendations. POST holds investigative and sanctioning power, including subpoena authority and the ability to impose binding disciplinary outcomes. Its civilian-majority composition includes members appointed by the governor and attorney general with expertise in criminal law, civil rights, social work, and crisis response (Post Commission Leadership). The Commission is also legally obligated to maintain a public database of officer certification status and disciplinary records, aiding national efforts to prevent "wandering officers" from moving between agencies without accountability (Du 2025; Subramanian and Arzy 2021).

This reform introduces a certification regime to enforce professional standards, expand civilian authority, and institutionalize external oversight. POST certification is now a condition of continued employment and shifts disciplinary power away from internal command structures. This transfer of authority was met with strong resistance from police unions and police leadership across the state, who held public protests in uniform and raised legal challenges to changes to certification, as well as the broader package of public safety reforms (Young 2020). All active-duty officers entered a phased recertification schedule in 2021, based on the alphabetical order of their last names, and officers must renew their certification every 3 years. The process comprises two key phases: an internal departmental review and an external review by the POST Commission. POST can revoke, suspend, or condition certification for a range of violations—including misconduct, training deficiencies, and criminal convictions, and must do so under certain conditions (MA POST Commission).[2] Officers facing decertification receive a public hearing, and final decisions are sent to agency heads and published in a public database listing all officers' certification status and disciplinary records (MA POST Commission, Officer Disciplinary Records).

Decertified officers are terminated, barred from law enforcement statewide (General Law, n.d.) and listed on the National Decertification Index (Kaste 2025). Terminations related to POST decertification are not subject to collective bargaining protections or arbitration, which previously reversed terminations, even cases of egregious misconduct (Katcher 2024).

This institutional architecture introduces a legally binding, recurring review process that compels officers to defend their professional fitness under conditions of formal scrutiny and public visibility. As such, it offers an ideal case for empirically assessing whether centralized certification can have a short-term effect on officer behavior within a coercive bureaucracy historically resistant to external control. This study focuses specifically on how POST recertification shapes officer behavior after participation, offering insight into the behavioral consequences of institutional design and the role of transparency and public accountability in shaping bureaucratic conduct.

Anticipated Behavioral Effects of POST Recertification

We argue that procedurally intensive, externally enforced accountability mechanisms, such as POST recertification, induce short-term behavioral recalibration among frontline officers. Rather than fostering normative change, we theorize that these interventions prompt instrumental shifts as officers reassess the reputational risks of misconduct. When participation is mandatory, outcomes are public, and oversight is external to the organization, participation in mechanisms of scrutiny constrains discretionary behavior, at least temporarily.

We conceptualize participation in POST recertification as an externally imposed accountability mechanism that influences officer behavior by enhancing the credibility of scrutiny and sanctions. As a binding, individualized oversight process administered by a civilian-majority commission, it places officers in a formal forum where they must justify past conduct under conditions of transparency. Recertification marks a rare instance in which the state and the public gain meaningful leverage over officers. The departmental opacity, union protections, and informal networks that have historically diluted reform are weakened, as each officer is now subject to external review, by both a civilian body with subpoena and enforcement authority and by journalists, civil rights organizations, and community members. Because POST decisions are not subject to managerial override or collective bargaining, certification becomes contingent on adherence to external standards. We anticipate that, taken together, these features reshape how officers perceive the reputational risks associated with misconduct, prompting short-term behavioral recalibration.

Although POST evaluates past conduct, our focus is on how its effects unfold prospectively. Research in

behavioral public administration underscores that retrospective review can reshape forward-looking perceptions of accountability, especially when the process is rule-bound and public. As Bovens (2007: 453) argues, "norms are (re)produced, internalised and, where necessary, adjusted through accountability." Participation in POST recertification activates this logic by placing officers in a formal forum where their conduct is evaluated, recorded, and made visible to external actors. We anticipate that this moment of exposure heightens perceived scrutiny and generates short-term behavioral restraint.

However, we do not expect these effects to be durable. Officers adjust behavior in response to a credible threat of sanction, but once that threat recedes, so too does the incentive to maintain restraint. While POST's subpoena and enforcement powers may initially loom large in officers' risk assessments, if formal consequences such as decertification remain rare, the process may come to be seen as symbolic.

We therefore conceptualize POST as a bounded accountability mechanism, capable of producing short-term reductions in misconduct, but unlikely to induce durable behavioral change in the absence of continuous enforcement and sustained reputational threat. This leads to our directional expectation:

> **H1:** Participation in POST recertification will lead to a short-term reduction in documented officer misconduct.

The scope of our argument is delimited in three respects. First, it applies to coercive bureaucracies in consolidated democracies, such as police departments, where frontline agents possess wide discretion and are often shielded from external scrutiny. Second, it pertains to accountability regimes that are binding, individualized, and externally adjudicated, rather than advisory or internal. Third, it assumes that certification outcomes are publicly accessible, thereby introducing reputational consequences that extend beyond formal sanction.

Research Methodology

This study employs a mixed-methods design that captures both behavioral outcomes and perceptual shifts. By combining a natural experiment with semi-structured interviews, the research design enables both the identification of systematic patterns in officer misconduct and a deeper understanding of how individual actors perceive and respond to new accountability mechanisms.

Our quantitative analysis draws exclusively on data from officers with sustained allegations of misconduct, as identified by the Massachusetts POST Commission. This focus allows us to test our hypothesis—that participation in POST recertification reduces officer misconduct—on a particularly challenging and policy-relevant subset of the officer population. Officers with documented misconduct histories are widely seen as least responsive to accountability measures, yet they are precisely the individuals whom such reforms are designed to reach, and who pose the greatest risk to public safety. In this way, the natural experiment offers a conservative test of the intervention's effectiveness by centering those officers most in need of reform and least likely to adjust their behavior in response to novel accountability regimes.

To contextualize these findings, the qualitative component draws on interviews with a hard-to-reach sample of officers across a broader spectrum of disciplinary histories. Our interview sample ranges from officers with no sustained complaints to officers at risk of decertification. This variation allows us to explore how officers with different exposure to the possibility of formal sanction interpret and respond to the POST process. While the quantitative data capture aggregate patterns of behavior change, the interviews offer insight into how recertification participation may influence day-to-day decision-making, perceptions of risk, and interpretations of legitimacy. By spanning both ends of the disciplinary spectrum, the qualitative data helps to contextualize behavioral trends and assess whether observed reductions in misconduct reflect broader shifts in how officers navigate discretionary authority under conditions of external oversight.

Statewide Police Recertification: A Natural Experiment

Data. We construct our panel dataset using publicly available officer misconduct data reported monthly by the POST Commission. The public dataset reported by POST is continuously updated and encompasses the period from January 1989 through the present and comprises 6,340 substantiated incidents of police misconduct. The dataset does not include officers with no disciplinary record, unfounded or unsubstantiated complaints, but only those against whom there are *sustained* allegations of misconduct.[3] As a result, we work with a universe of officers who have at least one sustained allegation in their record. For each sustained allegation, the dataset codes the date, type and description of the misconduct, the discipline imposed, the officer's agency, and certification status (Officer Disciplinary Records—MA POST Commission).

While the POST Commission establishes uniform certification standards and definitions of misconduct, implementation varies across agencies in rigor and transparency. As agencies adjust to new reporting norms, data completeness may improve over time, particularly

after recertification. This likely increase in reporting makes any observed treatment effects more compelling, as it would bias against finding reductions in misconduct. Moreover, POST data include only sustained allegations of serious misconduct, excluding minor infractions and unsustained complaints. While this limits the behavioral scope of our analysis, it also strengthens its relevance: our sample consists entirely of officers with a demonstrated record of serious misconduct, and thus most directly targeted by—and potentially resistant to—oversight reforms.

Using this dataset, we construct an officer-level monthly panel dataset spanning January 1, 2019 to June 30, 2024. To create a balanced panel, we convert the incident-level data into an officer-month format: each officer who committed any misconduct after January 1, 2019 appears in every month of the study period. For each officer-month, we record the number of misconduct incidents and code as zero if no misconduct occurred. This structure allows us to rigorously track changes in misconduct behavior before and after the reform's implementation.

The final dataset comprises 58,608 officer-month observations, each representing an officer's record in a given month, including variables such as misconduct count, certification status, and department. We also construct a control variable—*baseline misconduct*—which captures the number of sustained misconduct incidents each officer committed prior to 2019.

Experimental Design. We leverage the plausibly exogenous, alphabetically phased rollout of officer recertification by the Massachusetts POST Commission to leverage a natural experiment and employ a difference-in-differences (DiD) framework to evaluate the impact of the certification review on officer misconduct. Officers across the state were assigned to recertification cohorts based solely on the first letter of their last name—an administrative decision unrelated to individual or departmental characteristics. Specifically, officers with last names beginning A–H were recertified between July 1, 2021, and July 1, 2022; those with last names I–P between July 1, 2022, and July 1, 2023; and those with last names Q–Z between July 1, 2023, and July 1, 2024.

Alphabetical assignment is a bureaucratic sorting mechanism previously used in quasi-experimental designs, particularly when the ordering variable is uncorrelated with the outcome of interest (Ho and Imai 2008; Meredith and Salant 2013). In this context, last name functions as a non-manipulable variable orthogonal to officer performance, tenure, or disciplinary profile. Consequently, assignment by last name can be viewed as a plausibly exogenous shock. This allows us to form credible counterfactual comparisons and isolate the short-term behavioral effects of recertification from other institutional or temporal confounders.

Although the full rollout involved three distinct cohorts assigned 1 year apart between 2021 and 2024, we adopt a two-group, two-period DiD design. We designate the first cohort (A–H) as the treatment group and the last cohort (Q–Z) as the control group. This decision is motivated by two methodological considerations. First, we exclude the intermediate I–P cohort to ensure a clean separation between treated and untreated officers, as our data specify only the annual review period and not the exact recertification dates. Second, we truncate the dataset before the Q–Z cohort begins its own recertification (ending the sample on June 30, 2024), ensuring that the control group remains untreated and thus suitable for comparison. Still, interactions among officers, particularly within tight-knit departments, may diffuse elements of the treatment through informal conversations and peer influence, potentially contaminating the control group. Nonetheless, the analysis seeks to isolate the direct effects of participation in recertification and focuses only on officers who participated in the recertification process. Additionally, a small number of officers (n = 48) graduated from the academy in 2022 and were certified with their cohort rather than alphabetically. Their inclusion could introduce bias, but the available data does not allow us to identify them.

We conceptualize our treatment as participation in the first round of the recertification process. Importantly, this treatment involves participation in both internal review and external oversight by POST, which serves as the final arbiter of an officer's fitness to serve.[4]

We expect participation in POST recertification to function as a credible signal of the heightened scrutiny and threat of professional sanction imposed by the state.

A key limitation, however, is that we do not know the exact recertification date for individual officers. For instance, although officers with last names A–H were reviewed between July 1, 2021 and July 1, 2022, we cannot pinpoint precisely when any given officer completed their recertification. We therefore use July 1, 2022 as our treatment date as we know that by that date, every officer in the A–H cohort had been exposed to the recertification process and its attendant scrutiny.[5]

We employ two samples for both theoretical and methodological reasons. In the main specification, we retain all officers, including those who were eventually decertified. Theoretically, decertification may itself be a mechanism through which the policy reduces misconduct. However, because officers who exit the force subsequently register zero misconduct incidents, their continued inclusion could artificially deflate post-treatment averages.[6] To address this concern, we also estimate models in which we drop decertified or separated officers

after their exit, while retaining their pre-exit observations. This approach preserves baseline misconduct data while restricting post-treatment analysis to officers who remain at risk of offending.

Results. Our DiD analysis compares trends in misconduct before and after July 1, 2022, between the early (treatment) (A–H) and late (control) (Q–Z) recertification cohorts. The identification strategy relies on the standard parallel trends assumption—that, in the absence of treatment, misconduct would have evolved similarly across both groups. To test this, we ran an event study specification using a "RelativeMonth" variable, which measures each observation's distance (in months) from July 1, 2022. We then interacted these RelativeMonth dummies with the treatment indicator to estimate separate time trends for treated and control officers during the pre-treatment window. The interaction coefficients for pre-treatment months were not statistically distinguishable from zero, supporting the validity of the parallel trends assumption. Full event study results are reported in the appendix.

The treatment group (last names A–H) comprises 595 officers observed over 66 months (39,270 officer-month observations), while the control group (Q–Z) comprises 293 officers over the same period (19,338 observations). This imbalance reflects the natural distribution of surnames in the population; however, it does not threaten the validity of our difference-in-differences (DiD) estimates so long as treatment assignment is exogenous and pre-treatment trends are parallel. To address potential heteroscedasticity resulting from group size differences, we cluster standard errors at the officer level in our analyses.

To test our hypothesis, our baseline specification first uses OLS to regress the monthly misconduct count on the interaction between cohort (treated vs. control) and the post-treatment period, while controlling for each officer's baseline misconduct, and including cohort and year-month fixed effects. This DiD setup compares how misconduct trends evolve for treated versus control officers before and after recertification, effectively differencing out both time-invariant cohort differences and common temporal shocks to isolate the reform's average causal effect (see, Kropko and Kubinec 2020).

Our main specification is:

$$Y_{it} = \alpha + \delta \text{Treated}_i + \sum_m \lambda_m D_{t=m}^{\text{YearMonth}} + \beta(\text{Treated}_i \times \text{Post}_t) + \gamma \text{BaselineMisconduct}_i + \varepsilon_{it}$$

Here, Y_{it} is the number of misconduct incidents for officer i in month t. The term δ Treated$_i$ absorbs any time-invariant difference between the treatment and control recertification groups, while the λ_m year-month dummies control for shocks common to all officers in each calendar month. The coefficient β on the interaction (Treated$_i$ × Post$_t$) is our difference-in-differences estimate that captures how much the change in misconduct among treated officers after certification differs from the simultaneous change among controls. Finally, γ BaselineMisconduct$_i$ adjusts for each officer's prior misconduct history, and standard errors are clustered by officer to account for within-officer correlation.

Treated x Post is our DiD interaction capturing the average impact of the recertification review. Table 1 reports the OLS difference-in-differences estimates for monthly misconduct counts. In Model 1.1, the coefficient on the DiD term is −0.008 (SE = 0.007, $p = 0.268$), indicating that, on average, the recertification review does not have a statistically significant effect on the number of misconduct incidents per month. The estimate

Table 1. Difference-In-Differences (DiD) Results, OLS Regression.

	Dependent Variable: *Count of Misconduct*	
	(1)	(2)
DiD	−0.008 (0.007)	−0.008 (0.007)
Baseline misconduct	0.002* (0.001)	0.002* (0.001)
Group (fixed effect)	✓	✓
Year-month (fixed effect)	✓	✓
Clustered standard errors (Officer)	✓	✓
Log-likelihood	0.001	0.001
BIC	0.352	0.343
Observation	58,608	57,024

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

is close to zero, suggesting no meaningful difference between treated and control officers after accounting for group and year-month fixed effects. With an adjusted R^2 near zero, the model explains virtually none of the variation in misconduct counts, reinforcing that we do not detect a clear treatment effect in this specification. Model 1.2 reports the results after excluding post-treatment observations for treated officers who are no longer in force (i.e., decertified, on leave, or suspended). This second specification yields a similarly small and insignificant DiD coefficient of −0.008 (SE = 0.007, p = 0.256).

These results suggest that the recertification review does not have a detectable effect on monthly misconduct counts. Although the coefficient remains negative in both models, it is not statistically distinguishable from zero, implying that we cannot reject the possibility that the treatment has no effect on misconduct incidents. Baseline misconduct, however, is positively associated with future misconduct, though the result is marginally significant (p = 0.094 in Model 1.1). This suggests that each additional incident of past misconduct is associated with a small increase in the number of future misconduct incidents.

While the OLS results do not support a clear effect on the intensive margin (the number of incidents), recertification could still affect the likelihood that an officer commits any misconduct at all.

To examine this extensive margin (any misconduct versus none), we next estimate a logit model using a binary indicator for whether an officer had at least one incident in a given month. Building on the same two-way fixed-effects specification above, the logistic model uses a logit link and interprets β as the difference-in-differences effect on the log-odds of misconduct due to recertification.

In Table 2, the logit coefficients on the DiD term are negative and statistically significant, indicating that treated officers have lower odds of committing misconduct than their untreated counterparts. Specifically, Model 2.1 suggests a 25.4 percent (exp (−0.293) ≈ 0.746) reduction in the odds (p = 0.046), while Model 2.2, which excludes officers who are no longer active due to decertification, indicates a 24.6 percent (exp (−0.282) ≈ 0.754) reduction (p = 0.057). Figure 1 plots the predicted probability of misconduct occurrence by the experiment groups (based on Model 2.1), suggesting that the treatment group on average is less likely to commit any misconduct compared to the control group in the post-treatment period. We also find that each additional incident of baseline misconduct slightly increases the odds of future misconduct by around 3.5 percent across each cohort.

These findings indicate that recertification review primarily affects the extensive margin (whether an officer engages in any misconduct) rather than the intensive margin (the number of incidents committed). In our DiD analysis of the count of monthly misconduct incidents (OLS), the treatment coefficient is negative but not statistically significant, suggesting no detectable change in incident counts for officers who misbehave. In contrast, when we model a binary outcome ("any misconduct" vs. none) using a fixed-effects logit, we observe a significant reduction in the odds of committing at least one incident after recertification. In other words, while participation in recertification does not appear to lower how many incidents of misconduct an offending officer commits, it does make officers less likely to offend at all.[7]

To increase confidence in our findings, we conduct several robustness checks (see Appendix). First, we re-estimate our baseline DiD using individual-level fixed effects (instead of cohort fixed effects) to account for any unobserved, time-invariant officer heterogeneity (Tables A3 and A4). Second, we shift the treatment cutoff to July 1, 2021 (the start of A–H reviews) and confirm that no "anticipation" effect appears before actual recertification (Tables A5 and A6). Third, we drop the top 1% of officers

Table 2. Difference-In-Differences (DiD) Results, Logistic Regression.

	Dependent Variable: *Misconduct Occurrence*	
	(1)	(2)
DiD	−0.293** (0.147)	−0.282* (0.148)
Baseline misconduct	0.036*** (0.011)	0.034*** (0.012)
Group (fixed effect)	✓	✓
Year-month (fixed effect)	✓	✓
Clustered standard errors (Officer)	✓	✓
Log-likelihood	−5,248.1	−5,130.9
BIC	11,253.8	11,017.5
Observation	58,608	57,024

*p < 0.10, **p < 0.05, ***p < 0.01.

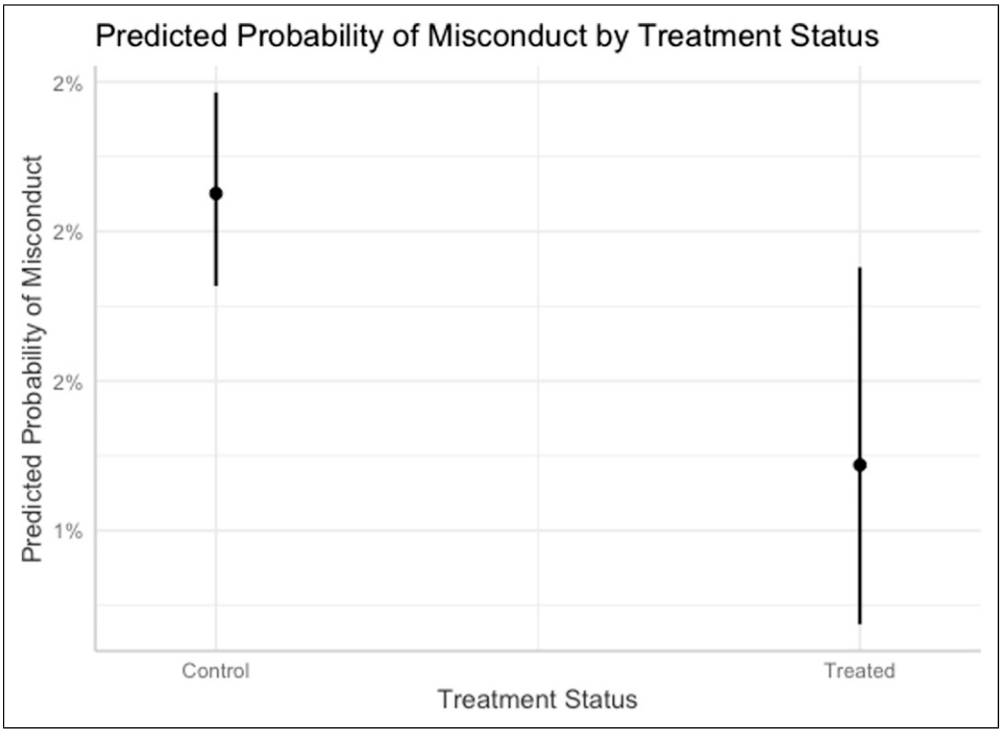

Figure 1. Predicted probability of misconduct occurrence by treatment status.

by total misconduct (outliers) and verify that our main coefficients remain essentially unchanged (Tables A7 and A8). Across these checks, our main results remain robust.

Additionally, we examine the middle (I–P) cohort, which was excluded from the main DiD model to avoid overlapping treatment, by running a simple pre/post comparison around their July 1, 2023 treatment date. These preliminary results (without controlling for time-varying factors) suggest that I–P officers experience a post-treatment decline in both the likelihood of any misconduct and in the average count of incidents.

We also explore whether recertification has a different impact for officers with prior misconduct (Table A9 in the appendix). In these interaction models, the negative coefficient on DiD × BaselineMisconduct (approximately 0.006; $p < 0.01$) implies that each additional pre-2019 incident is associated with an extra 0.006 drop in monthly misconduct counts after recertification. Although this effect is small, it suggests that officers with more documented complaints see a slightly greater reduction in misconduct frequency. By contrast, when we re-estimate a fixed-effects logit on the binary "any misconduct" outcome, the interaction term is not significant, indicating that baseline misconduct does not meaningfully change the odds of committing at least one incident post-treatment.

One limitation is that our analysis treats all substantiated misconduct incidents as functionally equivalent, whether they involve "unprofessional conduct" or more serious offenses like sexual assault or excessive force. In reality, officers may respond differently to the prospect of sanctions for minor infractions versus major violations. Future research should disaggregate misconduct by severity or category for a more nuanced understanding (see, e.g., Cassino and Demir 2024).

In what follows, we present findings from interviews conducted with officers and oversight officials. These qualitative insights help explain our mixed quantitative results by revealing how officers perceive the credibility of POST's scrutiny and sanctions, as well as how those perceptions, in turn, shape their on-the-ground behavior.

Qualitative Analysis

To better understand the dynamics of police misconduct and recertification, we conducted a series of semi-structured interviews in a major Massachusetts city in Spring of 2025.[8] This site was selected for its strategic relevance along several dimensions. Professional and academic connections facilitated access to municipal stakeholders within the police department and city government, which allowed for a referral-driven sampling strategy.[9] The city is also one of ten majority-minority municipalities in Massachusetts (U.S. Census Bureau 2020), offering an important demographic context for studying how police reform plays out in diverse urban environments. Finally, this jurisdiction had prior

instances of officer decertification by the POST Commission, providing a valuable setting for examining how officers respond to the credible threat of sanction.

Participants included patrol officers, supervisors, internal affairs officials, union representatives, a POST official, and a city solicitor. Most interviews were in person, lasted about 60 minutes, and all officer participants had completed the first POST recertification round. Unlike the quantitative sample, which focuses on officers with sustained misconduct, the interviews included both those with and without POST records, allowing comparison across disciplinary profiles. This design enables us to capture a range of perspectives on the recertification process, with specific attention to how officers with different disciplinary profiles interpret the POST regime.

Interview findings should be interpreted as illustrative rather than representative. These qualitative insights are drawn from a single urban police department and reflect a specific institutional context. Furthermore, the interviewer's embeddedness in the local context and positionality as a white, middle-class woman from a neighboring community likely influenced disclosure. Additionally, there were no female officers in the sample and many officers declined interview requests. As a result, the perspectives of a substantial subset remain absent from the interview data.

Interview protocols focused on officers' experiences with POST, perceived legitimacy of the process, behavioral responses, and broader attitudes toward oversight and reform.[10] To understand the ways in which individuals across institutional roles and positions perceive and respond to recertification, we first turn to interview findings that reflect officers' views on the POST Commission. From there we unpack how officers, oversight officials, and city officials tasked with litigating police misconduct perceive the scrutiny and sanctions levied by POST. Finally we turn to the problem of enduring misconduct in the wake of exposure to increased scrutiny and sanctions.

Perceptions of POST: Distrust, Redundancy, and Political Framing. Although POST has reported over 6,000 sustained misconduct cases, officers regarded its introduction as symbolic and unnecessary in Massachusetts, attributing it to controversies elsewhere rather than local conditions Respondents often described recertification as redundant, claiming departments already had robust internal accountability (AK140325TPD, PE032125TPD).

Although the POST Commission was established with input from law enforcement and police unions (Office of the Attorney General 2021), officers across ranks described it as politically motivated and lacking practical understanding of policing. One detective characterized it as "detached from the realities of law enforcement" and implemented "without sufficient consultation with officers on the ground" (AG140325TPD, NE210325TPD, DM140325TPD). Civilian oversight drew particular criticism. Officers viewed civilian-majority boards as disconnected from the daily realities of the job (NE210325TPD, DM140325TPD, AK140325TPD), questioning civilians' qualifications to assess complex or ambiguous encounters. "Police work has never been pretty," one officer said, asserting that civilian reviewers lack the experience to judge officer conduct (NE210325TPD). These views persisted despite the fact that one-third of POST members come from law enforcement backgrounds (POST Commission Leadership), indicating that the label "civilian" carries symbolic weight that overshadows the Commission's actual composition.

The disjuncture between the POST Commission's formal composition and how it is perceived by officers underscores a deeper cynicism toward external oversight institutions. Officers described police as uniquely targeted by external accountability regimes and unfairly judged by the public (NE210325TPD, DM140325TPD, AK140325TPD, PC240325NPD),[11] adopting a posture of occupational victimhood that reframes accountability reforms not as legitimate mechanisms of oversight, but as assaults on professional identity that stigmatize officers and frame rights protections for the public as reducing their own.

Yet some divergence was evident. One officer hired after POST's implementation expressed comfort with recertification and external oversight, describing it simply as a "standard bureaucratic obligation" and a welcome way to hold police accountable (PE032125TPD).

Scrutiny and sanctions. While most officers rejected POST as illegitimate and detached, they still saw its authority as a credible threat to professional autonomy and status. One described decertification as a "death sentence" (NE210325TPD); another stated that once decertified, officers are "done in policing" (PE210325NPD).

Confrontation with the possibility of decertification imposed a clear behavioral constraint on many of the officers interviewed. Officers reported heightened caution, especially around low-level enforcement, with supervisors discouraging proactive stops to reduce liability. One officer recalled new recruits being told not to randomly "stop guys and search guys" due to legal risk (PC240325NPD).[12] Officers described themselves and colleagues as more cautious, less aggressive, and "gunshy" (AG140325NPD), with some stating they prioritized legal safety over personal safety (AG140325NPD; NE210325TPD; PC240325NPD).

The visibility of disciplinary records amplified POST's power. One officer noted a case overturned in arbitration still appeared on his POST record: "If it's

overturned, it shouldn't appear" (PC240325NPD). His supervisor's hesitation to sign the attestation led him to police more cautiously. Several expressed fear of "being placed on a POST list," which they said "absolutely" influenced their decisions (PC240325NPD; NE210325NPD).

Officers also recalibrated behavior after observing colleagues sanctioned under POST. One recounted an officer boarding a school bus to physically confront a student—previously a matter for internal discipline, now grounds for potential decertification (NE210325TPD). This shift was compounded by a sense that institutional protections were eroding. "The real morale killer," one patrolman remarked, "is that the department and city don't back you" (NE210325TPD). Officers noted reduced ability to evade sanctions through internal channels, intensifying perceptions of risk (NE210325NPD; PC240325NPD).

Despite rejecting POST's legitimacy, many officers internalized its disciplinary authority. The credible threat of decertification drove tangible behavioral change, not due to normative alignment with reform, but out of self-preservation. Notably, officers viewed internal review as a routine "nonprocess," comparable to internal affairs oversight (DM140325NPD; AK140325NPD; PE210325NPD). It was the external phase consisting of POST's independent review that heightened the perceived risk of real consequences, reinforcing officers' greater trust in external over internal oversight.

Institutional Perspectives on POST: Insights from a City Solicitor. An interview with a city solicitor highlighted how POST reasserts institutional control over discipline through binding, state-imposed oversight (FK032425CN). Unlike internal accountability mechanisms, POST decertification overrides arbitration: "If the state decertifies an officer, that officer cannot perform the duties of a police officer, and therefore the city cannot keep them employed." This removes the obligation to retain officers with serious misconduct, closing loopholes that let them evade sanctions. "I don't think any arbitrator will reinstate a decertified officer," the solicitor emphasized, underscoring POST's authority as the final arbiter of employability (FK032425CN). Furthermore, POST's authority is structurally independent of collective bargaining agreements. "Last chance" settlements are structured so that POST action is "separate and apart and distinct" from internal resolutions and union protections, meaning that even reinstated officers can be lawfully terminated once they are decertified by POST (FK032425CN).

Decertification also reduces financial liability for municipalities. Severance benefits under union contracts are not triggered by state action, so cities need not provide payouts to officers terminated as a result of decertification. On claims that POST is a career-ending threat, the solicitor was clear: "As it should be." From their perspective, POST empowers municipalities to lawfully dismiss officers deemed unfit for service—officers who had previously been shielded by internal politics and union protections. While acknowledging operational tensions, the solicitor framed POST as a necessary state intervention that recenters accountability on certifiability in the most egregious cases. Yet POST sanctions remain rare. As of April 2025, just 46 officers had been decertified, despite over 6,300 sustained allegations involving 888 officers. Grounds for decertification included felony convictions, excessive use of force, and falsifying police reports, among other violations (MA POST Commission, Officer Disciplinary Records). Next we draw on additional officer interviews to explore why misconduct persists in the new regime.

Enduring Misconduct and the Limits of Reform. Despite some officers reporting that the process constrained their day-to-day decision-making, our large-N analysis reveals that misconduct persists, suggesting limited responsiveness to externally imposed evaluation and discipline. In reference to these patterns, several officers expressed doubts about the rehabilitative potential of POST. Several interviewees described reoffending officers as either inherently unresponsive to reform or as strategic actors adept at navigating oversight. One detective remarked, "I think some guys are just... you just know like, oh boy, here he is," referring to officers whose presence reliably escalated tension. He concluded that this behavior was innate: "It's a personal kind of attribute or like personality" (AG140325NPD). A patrol officer offered a similar view, comparing repeat offenders to "a bully in school who continues to misbehave despite consequences" (PE210325NPD).

A sergeant cited disposition as key to persistent misconduct: "If you're not a rule follower before, you're not going to start following the rules now because there's a fancy POST commission that's going to be involved" (DM140325NPD). He linked noncompliance to greater risk tolerance: "They are more willing to take on the risks of... not doing things the exact way they're supposed to be done." In contrast, he characterized himself as risk-averse: "It's like, you're that type of person or you're not." For him, repeat misconduct reflects a fixed trait: "It's just kind of what they're going to do until they get removed from the job" (DM140325NPD).

Echoing these observations, some respondents expressed willingness to circumvent POST constraints, despite the threat of career-ending sanctions. One officer from a specialized unit remarked, "There is going to be a time where you have to be nasty because you're going to be dealing with animals," and, "You got to kind of check

yourself back in and go the heck with the reform stuff or what society wants from us. You know the bad actors. Go approach and see what you can see" (AG140325NPD). He explained that officers can bypass new protocols that render chokeholds automatic grounds for decertification by switching to other forms of force at their discretion, stating :

> "I cannot choke you to save my own life, I can beat you with this table to death. I can hit you with a brick. I can run you over with my car. I can beat you with this radio. Anything to save my own life, because it's a deadly force situation. However, I cannot choke you.[13]

A city official echoed this skepticism about the reform's capacity to constrain the most problematic officers and described officers engaging in misconduct so egregious that they questioned, "What makes a person think that they can get away with something like this?" They saw persistent misconduct as rooted in personality traits such as a "lack of integrity," a "personality deficiency," or "an ego that's just too big," that are fundamentally incompatible with the responsibilities of police work (FK032425CN).

This tension between perceived reformability and intractability highlights a critical limitation of POST. While the system is designed to promote accountability, many officers and administrators express doubt that those who repeatedly violate norms can ever be rehabilitated, unsurprising given the rarity of actual decertification. Despite the visibility of POST and the formalization of misconduct oversight, officer removal remains an exceptional event. As a result, recalcitrant officers interpret the system as largely symbolic, with meaningful consequences applied only in extreme or highly public cases. One patrolman captured this perception, describing how repeat offenders seem to "know how the system works, almost like the criminals," suggesting an insider's understanding of disciplinary loopholes that enables them to evade sanctions without altering their behavior (NE210325NPD).

Ultimately, interviews reveal how recertification may have limited capacity to shift behavior among officers already embedded in patterns of misconduct. At the same time, the institutional emphasis on certification over contextual judgment and internal accountability has generated uncertainty even among officers who view themselves as compliant. In both cases, the reforms appear to recalibrate professional identity, not through enforcement, but through contested definitions of who qualifies as a "good cop" under a regime of external oversight.

Discussion: Constraint Without Transformation

While certification requirements have long existed in U.S. policing, their impact on officer behavior is difficult to ascertain. The introduction of POST recertification in Massachusetts represents a structural shift in how certification is organized and enforced, offering a distinct opportunity to examine how externally structured accountability affects officer conduct.

Quantitatively, we find no significant reduction in the frequency of misconduct among offending officers, but we do observe a meaningful decrease in the likelihood of misconduct occurring at all following recertification. This may suggest that recertification deters marginal or first-time violations, even though it does not significantly change behavior among officers who persistently offend. Qualitative findings indicate two mechanisms shaping behavior: risk recalibration under external monitoring and reputational sanctioning via public transparency.

First, POST restructures the principal-agent relationship between officers, agencies, and the state. As an external, legally autonomous body, POST bypasses internal protections and overrides local discretion. Officers described delaying enforcement and modifying behavior to reduce liability, rather than follow rights-enhancing protocols. These actions reflect strategic adaptation, with bureaucrats using institutional knowledge to shape decisions in their favor rather than to signal normative change, aligning with research on the persistence of bureaucratic agency under external oversight and broader scholarship on how political actors navigate constraints to protect their interests (Herd et al., 2023; Potter 2019).

Second, POST's public-facing database transforms internal discipline into public signals. Officers stated that publication of certification status and disciplinary histories affected their reputation and community standing. This introduces a second axis of constraint where public transparency and legibility of coercive state power amplifies the symbolic, social, and material costs of misconduct, consistent with theories of governance by visibility in which accountability mechanisms gain force through their legibility to external audiences (Bovens 2007; D. Carpenter 2001). Together, these mechanisms suggest the saliency of accountability mechanisms that have the power to usurp internal protections and engage a public forum to facilitate reputational exposure. How these dynamics evolve as POST recertification becomes routine remains to be seen.

Our findings reinforce a central lesson from the accountability literature: institutional design alone is insufficient, and the legitimacy of procedural mechanisms must be actively sustained (Bovens 2007). In this case, the rarity of decertification renders sanctions theoretically

credible but practically constrained, allowing problematic officers to remain in service and reoffend.

Conclusion

While POST recertification represents a novel attempt to constrain a historically insulated institutional field, our findings highlight the limits of externally imposed accountability. Ongoing data from POST and longer-term studies are essential to assess whether these patterns persist and interagency comparative research can help identify under what conditions centralized oversight produces durable accountability. Further, our qualitative findings highlight the role of racial hierarchies in policing. Future research might examine these dynamics, through analyses of disciplinary outcomes by race or cross-agency comparison of how racial hierarchies shape responses to oversight.

From a public administration perspective, POST exemplifies constrained state intervention via recalibration of discretion through professional sanction, rather than structural change. Its impact is shaped by legal finality and visibility, not deep institutional integration. Absent material consequences, shifts in behavior are unlikely to endure. Decertification remains rare. Officers with documented misconduct remain on the force, and their peers characterized them as beyond reform. Recertification may influence rule-abiding officers but has limited impact on those who view it as a bureaucratic formality. As one sergeant noted, "It's just kind of what they're going to do until they get removed from the job" (DM140325NPD). Recertification thus appears more procedural than transformative, a constraint that shifts risk margins without disrupting entrenched patterns of impunity.

This raises concerns about oversight regimes that tolerate repeated misconduct. Framing this as a "bad apples" problem obscures the systemic failure to remove them. The issue is not just individual misconduct but a "rotten barrel" dynamic that undermines reform from within (Bains 2018; Chalfin and Kaplan 2021). For oversight to function, scrutiny must be paired with consistently enforced sanctions. Furthermore, as Cassino and Demir (2024) suggest, system-wide changes are more effective than standalone accountability measures. Certification is symbolically and procedurally significant but cannot succeed in isolation. It requires integration into an infrastructure of reform. Even then, serious misconduct persists (MA POST Commission, Officer Disciplinary Records).

The Massachusetts case illustrates a deeper accountability challenge: the gap between public legitimacy and internal acceptance. POST is a public-facing effort to restore community trust in law enforcement, but its operational effectiveness is contingent on officer investment in the systemic changes it aims to trigger. As long as this remains partial, its effects will be limited. This analysis highlights the consequences that arise when external oversight bodies refrain from fully exercising their authority. Yet, accountability mechanisms that appear limited or symbolic in one institutional context may prove more effective under different organizational or political conditions (Sheffer and Loewen, 2019). Future research should explore how analogous certification regimes operate across other democratic security forces, or within distinct bureaucratic domains, to better understand the conditions under which they induce behavioral change.

Acknowledgements

The authors gratefully acknowledge Kristine Eck, Annekatrin DeGlow, Sophia Hatz, Kathleen Klaus, Thorsten Rogall, Mihai Croicu, Faruk Aksoy, and Hüsne Akgöl for their insightful comments and constructive feedback. We also thank the audiences at the 2025 International Studies Association Annual Convention, the 2025 Annual Conference of the Midwestern Political Science Association, and the Research Paper Seminar at the Department of Peace and Conflict Research, Uppsala University, for their valuable suggestions and engagement. Special thanks to Yanilda González for her generous support and guidance during the fieldwork phase of this project.

Ethical Approval

This project underwent rigorous ethical review, received ethical approval, and complies with all ethical guidelines from the Central Ethical Review Board [Centrala etikprövningsnämnden], the National Ethical Review Authority responsible for ethical review and approval of research on human participants. The diary numbers for the review at the Swedish Ethics Review Authority are: 2022-04133-01 and 2023-06948-02.

Consent to Participate

Before each interview, participants received a detailed consent form and provided both verbal and written informed consent for the interview, audio recording, and potential follow-up contact. The form included contact information for the interviewer, the supervising researchers, the Ethical Review Board, and the data management center enabling participants to raise concerns or withdraw their participation at any time. Participants were informed—both verbally and in writing—of their right to end the interview at any point, as well as to retract or revise their statements at a later stage. To protect their privacy, participants were offered the option to remain anonymous. In this study, all identifying information has been removed, and the name of the municipality where interviews took place has been withheld to further ensure participant confidentiality.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No 101000385).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data and replication materials that support the findings of this study are available in the Harvard Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/AYQZAW.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. According to Mapping Police Violence, police killed 1376 people in 2024.
2. See Appendix for a more detailed review of the re-certification process, and conditions for decertification.
3. As of August 2025, officers with sustained allegations represent roughly 13 percent of all active officers in Massachusetts, based on figures reported on the POST website.
4. Please see appendix for detailed information about the POST recertification process.
5. We acknowledge this cutoff is imperfect. Since we lack individual certification dates, any chosen date is somewhat arbitrary. However, selecting July 1, 2022 guarantees that all A–H officers have "received" the treatment according to our theoretical expectation. As a robustness check, we also re-run our analyses using July 1, 2021 (the start of A–H reviews) as the cutoff; those results appear in Tables A5 and A6. In line with expectation, the DiD coefficients become insignificant when using the earlier cutoff, indicating no anticipation effect prior to actual recertification.
6. The number of officers who exit or are suspended due to decertification or non-certification are quite limited. Please see Table A2 in the appendix for statistics on this.
7. It would be reasonable to expect that agency heads might reassign officers to less public-facing roles in the wake of recertification, thus removing them from the sample of officers "available" to commit misconduct. For a note on how these administrative gymnastics might impact our results, please the appendix.
8. In accordance with informed consent and anonymization protocols established during the interview process, and to protect participant privacy, we do not name the law enforcement agency under study. Instead, interviewees are identified using codes that preserve anonymity with general references indicating their general institutional role.
9. The interviews were conducted by one of the authors.
10. Interview guides can be found in the appendix.
11. Teachers and social workers in Massachusetts are also subject to state-level certification.
12. It should be noted that in Massachusetts police must have a reasonable suspicion that someone actually violated the law in order to conduct a valid stop.
13. While this officer (AG140325NPD) may be seen as an outlier, he is one of the primary officers in charge of use-of-force and firearms training for the agency under study.

References

Ali, Mir Usman, and Sean Nicholson-Crotty. 2020. "Examining the Accountability-Performance Link: The Case of Citizen Oversight of Police." *Public Performance and Management Review* 44 (3): 523–59.

Ali, M. U., and M. Pirog. 2019. "Social Accountability and Institutional Change: The Case of Citizen Oversight of Police." *Public Administration Review* 79 (3): 411–26.

Alpert, G. P., T. Cawthray, J. Rojek, and F. Ferdik. 2015. "Citizen Oversight in the United States and Canada: Applying Outcome Measures and Evidence-Based Concepts." In *Civilian Oversight of Police: Advancing Accountability in Law Enforcement*, 179–204.

Bains, Chiraag. 2018. "A Few Bad Apples': How the Narrative of Isolated Misconduct Distorts Civil Rights Doctrine." *Indiana Law Journal* 29 (1). https://www.repository.law.indiana.edu/ilj/vol93/iss1/3

Becker, Deborah, and Ally Jarmanning. 2023. "Mass. Police Misconduct Database Is Released, but Excludes Many Complaints." *WBUR*; June 27, 2025. https://www.wbur.org/news/2023/08/22/mass-police-misconduct-database-post-complaints

Biesta, Gert. 2010. "Pragmatism and the Philosophical Foundations of Mixed Methods Research1." In *Sage Handbook of Mixed Methods in Social & Behavioral Research*, 95–118. Thousand Oaks, CA, USA: Sage Publications, Inc.

Bird, Christine C., and Brooke N. Shannon. 2024. "The Heat Is On: Does Civil Litigation Affect Policing Practices?" *Political Research Quarterly* 77 (4): 1212–29.

Blow, Charles M. 2023. "Opinion | Dead Men at the State of the Union." *The New York Times*; May 2, 2025. https://www.nytimes.com/2023/02/08/opinion/state-of-the-union-police-killings.html

Bovens, Mark. 2007. "Analysing and Assessing Accountability: A Conceptual Framework1." *European Law Journal* 13 (4): 447–68.

Bovens, Mark. 2010. "Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism." *West European Politics* 33 (5): 946–67.

Brehm, John, and Scott Gates. 1994. "When Supervision Fails to Induce Compliance." *Journal of Theoretical Politics* 6 (3): 323–43.

Brehm, John O., and Gates. Scott. 1997. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor, USA: University of Michigan Press.

Carpenter, Daniel P. 1996. "Adaptive Signal Processing, Hierarchy, and Budgetary Control in Federal Regulation." *American Political Science Review* 90 (2): 283–302.

Carpenter, Daniel. 2001. "The Political Foundations of Bureaucratic Autonomy: A Response to Kernell." *Studies in American Political Development* 15 (1): 113–22.

Cassino, Peter P., and Mustafa Demir. 2024. "The Effect of Police Reform on Overall Police Misconduct and Misconduct That Involves the Use of Force." *Journal of Criminal Justice* 94: 102263.

Chalfin, Aaron, and Jacob Kaplan. 2021. "How Many Complaints against Police Officers Can Be Abated by Incapacitating a Few 'Bad Apples?'" *Criminology & Public Policy* 20 (2): 351–70.

Clark, Tom S. 2024. "A Political-Economic Model of Police Administration." *The Journal of Politics* 86 (4): 1462–78.

Post Commission, Leadership. "Commission Leadership – MA POST Commission." April 23, 2025. https://mapostcommission.gov/about-post/meet-the-post-commissioners/

Davis, Jill A, Darwin A Baluran, and Shahidul Hassan. 2024. "Support for Democratic Policing Among Frontline Police Officers: The Role of Social Dominance Orientation." *British Journal of Criminology* 64 (2): 434–51.

Donner, Christopher M. 2019. "The Best Predictor of Future Behavior Is…': Examining the Impact of Past Police Misconduct on the Likelihood of Future Misconduct." *Journal of Crime and Justice* 42 (3): 300–15.

Donner, Christopher M, and Wesley G Jennings. 2014. "Low Self-Control and Police Deviance: Applying Gottfredson and Hirschi's General Theory to Officer Misconduct." *Police Quarterly* 17 (3): 203–25.

Du, Susan. 2025. "New Tool Will Allow Anyone to Look up Minnesota Police Officers' Employment History." June 3, 2025. https://www.startribune.com/new-tool-will-allow-anyone-to-look-up-minnesota-police-officers-employment-history/601206648

Engel, Robin Shepard, and Robert E Worden. 2003. "Police Officers' Attitudes, Behavior, and Supervisory Influences: An Analysis of Problem Solving." *Criminology* 41 (1): 131–66.

Epp, Charles R., Steven Maynard-Moody, and Donald P. Haider-Markel. 2014. *Pulled over: How Police Stops Define Race and Citizenship*. University of Chicago Press.

Eriksen, Andreas. 2021. "Accountability and the Multidimensional Mandate." *Political Research Quarterly* 74 (2): 364–76.

Ferdik, Frank V., Jeff Rojek, and Geoffrey P. Alpert. 2013. "Citizen Oversight in the United States and Canada: An Overview." *Police Practice and Research* 14 (2): 104–16.

Filstad, Cathrine, and Petter Gottschalk. 2011. "Performance Evaluation of Police Oversight Agencies." *Policing and Society* 21 (1): 96–109.

Flom, Hernán. 2020. "Controlling Bureaucracies in Weak Institutional Contexts: The Politics of Police Autonomy." *Governance* 33 (3): 639–56.

Flom, Hernán. 2024. "Voting with Their Guns: An Integrated Framework of How Police Politically Administer Violence." *Perspectives on Politics* 23: 1–16.

Fung, Archon. 2015. "Putting the Public Back into Governance: The Challenges of Citizen Participation and its Future." *Public Administration Review* 75 (4): 513–22. doi:10.1111/puar.12361.

General Law_Chapter 6E. "General Law - Part I, Title II, Chapter 6E, Section 1." February 12, 2025. https://malegislature.gov/Laws/GeneralLaws/PartI/TitleII/Chapter6E/Section1

González, Yanilda. 2019. "The Social Origins of Institutional Weakness and Change: Preferences, Power, and Police Reform in Latin America." *World Politics* 71 (1): 44–87.

González, Yanilda María. 2023. "Reforming to Avoid Reform: Strategic Policy Substitution and the Reform Gap in Policing." *Perspectives on Politics* 21 (1): 59–77.

Grimmelikhuijsen, Stephan, Sebastian Jilke, Asmus Leth Olsen, and Lars Tummers. 2017. "Behavioral Public Administration: Combining Insights from Public Administration and Psychology." *Public Administration Review* 77 (1): 45–56.

Herd, Pamela, Hilary Hoynes, Jamila Michener, and Donald Moynihan. 2023. "Introduction: Administrative Burden as a Mechanism of Inequality in Policy Implementation." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 9 (4): 1–30.

Ho, Daniel E., and Kosuke Imai. 2008. "Estimating Causal Effects of Ballot Order from a Randomized Natural Experiment." *Public Opinion Quarterly* 72 (2): 216–40.

Holliday, B. S., and J. H. Wagstaff. 2022. "The Relationship between Citizen Oversight and Procedural Justice Measures in Policing: An Exploratory Study." *American Journal of Criminal Justice* 47 (3): 567–88.

Idrobo, Nicolas, Dorothy Kronick, and Tara Slough. 2024. *The Effect of Oversight on the Quantity and Quality of Policing*. CIPR.

Ingram, Jason R., William Terrill, and Eugene A. Paoline. 2018. "Police Culture and Officer Behavior: Application of a Multilevel Framework." *Criminology* 56 (4): 780–811.

Johnson, Richard R. 2015. "Examining the Effects of Agency Accreditation on Police Officer Behavior." *Public Organization Review* 15 (1): 139–55.

Kaste, Martin. 2025. "Trump Took Down Police Misconduct Database, but States Can Still Share Background Check Info." *NPR*; June 3, 2025. https://www.npr.org/2025/02/28/nx-s1-5305281/trump-police-misconduct-database-background-checks

Katcher, William. 2024. "Police Officer Accused of Sexually Harassing Students to Contest His Firing." *MassLive*; May

29, 2024. https://www.masslive.com/news/2024/05/police-officer-accused-of-sexually-harassing-students-to-contest-his-firing.html

Kropko, Jonathan, and Robert Kubinec. 2020. "Interpretation and Identification of Within-Unit and Cross-Sectional Variation in Panel Data Models." *PLoS One* 15 (4): e0231349. doi:10.1371/journal.pone.0231349.

Lawson, Edward. 2019. "TRENDS: Police Militarization and the Use of Lethal Force." *Political Research Quarterly* 72 (1): 177–89.

Lewis, C., and T. Prenzler. 2017. "Police Oversight Agencies: Measuring Effectiveness." In *Promoting Integrity: Evaluating and Improving Public Institutions*, 199–219.

Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.

Mapping Police Violence. nd "Mapping Police Violence." Accessed August 19, 2025.https://mappingpoliceviolence.org/

Mastrofski, Stephen D. 2004. "Controlling Street-Level Police Discretion." *The Annals of the American Academy of Political and Social Science* 593: 100–18.

Meredith, Marc, and Yuval Salant. 2013. "On the Causes and Consequences of Ballot Order Effects." *Political Behavior* 35 (1): 175–97.

Mummolo, Jonathan. 2018. "Modern Police Tactics, Police-Citizen Interactions, and the Prospects for Reform." *The Journal of Politics* 80 (1): 1–15.

NCSL. *Developments in Law Enforcement Officer Certification and Decertification*. NCSL. June 27, 2025. https://www.ncsl.org/civil-and-criminal-justice/developments-in-law-enforcement-officer-certification-and-decertification

Office of the Attorney General. 2021. "Governor and Attorney General Appoint 9 Inaugural Members to the Massachusetts Peace Officer Standards and Training Commission | Mass.Gov." *Mass.gov*; June 3, 2025. https://www.mass.gov/news/governor-and-attorney-general-appoint-9-inaugural-members-to-the-massachusetts-peace-officer-standards-and-training-commission

Officer Disciplinary Records – MA POST Commission. "Officer Disciplinary Records – MA POST Commission." February 21, 2025. https://mapostcommission.gov/discipline-status-records/disciplinary-records/

Ouellet, Marie, Sadaf Hashimi, Jason Gravel, and Andrew V Papachristos. 2019. "Network Exposure and Excessive Use of Force: Investigating the Social Transmission of Police Misconduct." *Criminology & Public Policy* 18 (3): 675–704.

Perez, Thomas E. 2000. "External Governmental Mechanisms of Police Accountability: Three Investigative Structures." *Policing and Society* 10 (1): 47–77.

Potter, Rachel Augustine. 2019. *Bending the Rules: Procedural Politicking in the Bureaucracy*. University of Chicago Press.

Prenzler, T. 2000. "Civilian Oversight of Police." *British Journal of Criminology* 40 (4): 659–74.

Prottas, Jeffrey Manditch. 1978. "The Power of the Street-Level Bureaucrat in Public Service Bureaucracies." *Urban Affairs Quarterly* 13 (3): 285–312.

Rad, Abdul Nasser, David S. Kirk, and William P. Jones. 2023. "Police Unionism, Accountability, and Misconduct." *Annual Review of Criminology* 6 (1): 181–203.

Recertification – MA POST Commission. "Recertification – MA POST Commission." February 12, 2025. https://mapostcommission.gov/certification/recertification/

Rivera, Roman, and Bocar A. Ba. 2019. "The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago." *SSRN Electronic Journal*. doi:10.2139/ssrn.3317952.

Rocha Beardall, Theresa. 2022. "Police Legitimacy Regimes and the Suppression of Citizen Oversight in Response to Police Violence." *Criminology* 60 (4): 740–65.

Rozema, Kyle, and Max Matthew Schanzenbach. 2020. "Does Discipline Decrease Police Misconduct? Evidence from Chicago Civilian Allegations." *SSRN Electronic Journal*.

Sheffer, Lior, and Peter John Loewen. 2019. "Accountability, Framing Effects, and Risk-Seeking by Elected Representatives: An Experimental Study with American Local Politicians." *Political Research Quarterly* 72 (1): 49–62.

Subramanian, Ram, and Leily Arzy. 2021. "State Policing Reforms Since George Floyd's Murder." February 17, 2025. https://www.brennancenter.org/our-work/research-reports/state-policing-reforms-george-floyds-murder

Teodoro, Manuel P, and Adam G Hughes. 2012. "Socializer or Signal? How Agency Accreditation Affects Organizational Culture." *Public Administration Review* 72 (4): 583–91.

Terrill, W., and S.D. Mastrofski. 2002. "Situational, and Officer-Based Determinants of Police Coercion." *Justice Quarterly* 19 (2): 215–48.

Thomas, Mark P, and Steven Tufts. 2020. "Blue Solidarity: Police Unions, Race and Authoritarian Populism in North America." *Work, Employment & Society* 34 (1): 126–44.

U.S. Census Bureau. 2020. "Massachusetts State Profile." In **Explore Census Data*, data.census.Gov*. https://data.census.gov/profile/Massachusetts?g=040XX00US25. Accessed August 19, 2025.

Ward, Julie A., Javier Cepeda, Dylan B. Jackson, Odis Johnson, Daniel W. Webster, and Cassandra K. Crifasi. 2024. "National Burden of Injury and Deaths from Shootings by Police in the United States, 2015–2020." *American Journal of Public Health* 114 (4): 387–97.

Young, Colin. 2020. "Gov. Baker Signs Landmark Policing Reform Law." *WBUR*; December 31, 2020. https://www.wbur.org/news/2020/12/31/gov-baker-signs-landmark-policing-reform-law