



Understanding Policy Failure in Pakistan: Structural Mechanisms and Cross-National Learning

Rizwan Ashraf,¹ & Amna Mahmood²

Abstract:

Political failure in Pakistan has remained a persistent and multidimensional phenomenon, manifesting in governance breakdowns, institutional fragility, policy inconsistency, and recurrent democratic reversals. This paper seeks to develop a structured understanding of political failure in Pakistan by examining its underlying structural mechanisms and drawing comparative insights from cross-national experiences. The paper conceptualizes political failure as a systemic outcome produced through the interaction of institutional design, civil-military imbalance, weak accountability frameworks, elite capture, and constrained state capacity. The paper also discovers reform paths pursued by comparatively successful countries that overcame similar circumstances of instability and moved towards institutionalization. By combining structural explanations with cross-national learning, the paper helps to inform theoretical discourses on state failure and democratic weakness while providing policy-relevant knowledge for institutional transformation in Pakistan. This research uses a qualitative, comparative methodology with case study of Pakistan, rooted in the traditions of political economy and institutionalism. The results seek to shift the discussion from normative arguments to more analytically informed explanations that can help to support sustainable political and governance transformations.

Keywords: Pakistan, political failure, institutional weakness, civil-military relations, political economy, governance structures

INTRODUCTION

Pakistan produced many policies that read as technically sound. Outcomes still disappointed. The puzzle was not a lack of plans. The puzzle was the repeat gap between stated intent and delivered change. The governance literature explained why this gap persisted in many low- and middle-income settings. The World Bank argued that policy outcomes depended on bargaining among actors with unequal power, plus the credibility of commitments, plus the ability to enforce decisions. Fukuyama defined governance as the ability of a state to make and enforce rules, then deliver services. That definition shifted attention from policy text to power and compliance. This article applied that shift to Pakistan. Pakistan also fit a broader debate on hybrid political orders.

¹ PhD Scholar, Department of Politics & International Relations, International Islamic University, Islamabad, Pakistan. Email: rizwan.phdps27@iiu.edu.pk

² Professor, Faculty of Social Sciences, International Islamic University, Islamabad. Email: amna.mahmood@iiu.edu.pk

Competitive authoritarianism research described systems with elections and formal checks, paired with uneven constraints on power. The transition literature warned against linear assumptions about reform in such settings. These arguments mattered for policy failure. Ambiguous accountability weakened discipline over political elites and state agencies. The result was a high tolerance for partial delivery, selective enforcement, and symbolic reform (Shah, 2012).

The public policy field showed why delivery collapsed even when leaders announced clear goals. Pressman and Wildavsky showed how each added decision point multiplied the risk of breakdown. Lipsky showed why front-line staff rewrote policy in practice when caseloads were heavy, resources were scarce, and goals conflicted. These ideas fit Pakistan's routine conditions in service sectors. Development research then added a sharper claim. Pritchett, Woolcock, and Andrews described "capability traps," where governments adopted global best practice templates, yet core functions stayed weak (Pressman & Wildavsky, 1973). Andrews later explained why rule reform produced compliance in form while daily routines stayed unchanged. This article treated that pattern as a mechanism, not a label. Pakistan-specific work supported the expectation of repeat gaps between reform and outcomes. Husain's political economy account described how elite control over the state shaped economic governance choices. Hasnain (2010) linked devolution design to local accountability limits and service delivery distortions. World Bank work on decentralization argued that service delivery improved only when accountability relationships changed, not when authority shifted on paper (Ahmad et al., 2007). Work on the Eighteenth Amendment raised coordination and fiscal stress issues that could re-create fragmentation across tiers of government (Shah, 2012). Taken together, this literature suggested a structural mechanism story: power bargains shaped rules, rules shaped incentives, incentives shaped agency behavior, and agency behavior shaped what citizens received.

Worth-mentioning, this article contributed in three ways. First, it replaced single-cause accounts with a mechanism map. It treated corruption, capacity limits, and "political will" as outcomes of deeper structures, not as standalone explanations. Migdal's state-society lens supported this move. North, Wallis, and Weingast also supported it by linking elite bargains and violence control to the way rules operated (Migdal, 1998). Second, it made the argument testable. It specified pathways that could be traced in documents, budgets, appointments, monitoring records, and front-line practice. It used process tracing guidance to connect evidence to causal claims. Pierson's path dependence works informed expectations about why early choices locked in costly routines (Pierson, 2000). Third, it treated cross-national learning as bounded. Rose described lesson drawing as selective borrowing. Dolowitz and Marsh described policy transfer as a political process with coercion, emulation, and failure risks. James warned that transfer claims could overstate portability when context shaped incentives and enforcement. This article used that caution to separate "portable" tools from "context-bound" bargains (James & Lodge, 2003). The next sections reviewed the literature on policy failure in weak and hybrid states, set out a structural mechanism framework, explained the case selection and evidence strategy, presented mechanism narratives from Pakistan cases, then derived cross-national lessons with clear limits.

LITERATURE REVIEW

The study of policy failure considered weak and hybrid states as environments in which there were formal rules, but the state did not use these rules to create predictable public action. The World

Bank's Governance and the Law report put power, bargaining, and enforcement at the forefront of policy outcomes (World Bank Group, 217). State-society studies also arrived at a similar finding through a different approach. Migdal (1988) argued how patrons, brokers, and organized groups contested the state for compliance and control, which constrained the actions of the state. North, Wallis, and Weingast (2009) framed many developing polities as "limited access orders," where elites stabilized rule by restricting entry and sharing rents, which constrained credible enforcement. This scholarship shifted attention away from policy texts and toward the governing order that shaped policy choice, funding, and enforcement. Political settlements research treated policy failure as a symptom of elite bargains that decided who gained, who paid, and who controlled coercion (Khan, 2011).

In this perspective, reform agendas stalled when they challenged dominant bargains, or when they called on instruments of enforcement that no coalitions were willing to provide. This argument contradicted "capacity-first" explanations of failure, which had posited failure as a lack of skills or resources. It also contested linear "transition" stories, which had predicted a gradual path to liberal democracy, as many countries found themselves in stable mixed categories after adopting electoral rule (Carothers, 2002). Hybrid regime work strengthened the political logic. Diamond (2002) argued that rulers adopted elections and formal checks while maintaining informal control over media, courts, and oversight. Schedler (2006) described electoral authoritarianism as "unfree competition," where rulers used legal and administrative tools to tilt the playing field while preserving contestation. The competitive authoritarianism through uneven contestation and the leverage of state resources, which shaped both policy promises and the credibility of sanctions for noncompliance. These arguments implied a recurring pattern for public policy. Governments announced reforms to signal performance to voters, donors, and courts, yet rulers restricted monitoring and punishment to protect political control. Policy failure then reflected political survival strategies as much as administrative weakness (Hasnain, 2010).

Comparative political economy added mechanisms that linked political structure to policy change. Veto player theory predicted that fragmented coalitions raised the number of actors that could block policy change, raise bargaining costs, and produce diluted reforms (Tsebelis, 2009). In weak or hybrid states, fragmentation did not sit only in legislatures. It appeared inside cabinets, between civilian and security agencies, and across subnational patron networks. This pattern converted policy delivery into repeated bargaining over budgets, appointments, and enforcement priorities. It also created multiple "exit" options for losers, from court challenges to street pressure to bureaucratic noncompliance. The literature therefore warned against treating the state as a unitary actor.

State capacity research refined the "weak state" label by separating distinct capacities. Hanson and Sigman developed measures that distinguished extractive, coercive, and administrative capacity, and they showed that these dimensions moved differently across time and cases (Hanson & Sigman, n.d.). Besley and Persson (2011) linked state capacity to development clusters and stressed fiscal capacity as a foundation for public action, since revenue shaped what governments could credibly fund and enforce. Fukuyama (2013) added a conceptual warning. Analysts used "governance" as a catch-all, yet they measured different objects, which weakened causal inference and muddied policy advice. Together, these works suggested a practical diagnostic. Policies failed when they demanded

capacities the state did not hold. A revenue policy required credible collection. A regulation required inspection and adjudication. A social program required accurate targeting and sustained frontline effort.

These capacity debates also raised a legitimacy question. WDR 2017 treated compliance as a political outcome, since citizens obeyed rules when they expected fair treatment and credible enforcement (World Bank Group, 217). Migdal's account implied a related pattern. Where local brokers-controlled access to jobs, land, and protection, citizens routed claims through patrons rather than through public offices (Migdal, 1998). Policy failure then appeared as a feedback loop. Weak enforcement reduced trust, low trust reduced voluntary compliance, then agencies relied on discretion and extraction. Research on bureaucracy supplied micro foundations for these claims. Evans and Rauch (1999) linked merit recruitment and predictable careers in core economic agencies to higher growth in a cross-national sample, which pointed to the value of professional norms and organizational memory. This strand treated bureaucracy as a political object. Leaders shaped staffing rules to reward supporters, which traded long-term performance for short-term coalition management. That trade-off mattered in hybrid orders, where rulers valued control over impartial rule enforcement. It also mattered in decentralized settings, where patron networks influenced postings and transfers (Migdal, 1998).

Theories of delegation extended similar logic. McCubbins et. al. (1987) showed that elected principals crafted rules to mitigate bureaucratic drift. In weak and hybrid regimes, principals did not always prefer control in the same way. Some principals preferred ambiguity because ambiguous rules left space for selective enforcement and side payments. This preference undermined credible commitment. It also increased transaction costs for citizens and businesses who depended on stable rules. Research on corruption questioned whether principal-agent theories explained policy failure when corruption became institutionalized. Persson et. al. (2013) reframed "systemic corruption" as a collective action problem. Even honest actors expected others to demand bribes, so they complied with the corrupt equilibrium (Persson et al., 2012). This argument implied that technical anti-corruption agencies failed when broader expectations and incentives remained unchanged. It also explained why monitoring reforms produced weak effects when political protection networks shielded violators. This mattered for policy failure research because corruption distorted each link in the delivery chain, from procurement and staffing to inspections (Andrews et al., 2017) and benefit allocation.

Development governance research added a caution about template-driven reform. Grindle (2004) criticized the expansion of "good governance" agendas that demanded a long list of reforms in low-capacity settings, and she argued for sequencing and prioritization under binding constraints. Andrews (2008) criticized indicator-led governance reform that promoted a one-best-way model of public sector form, which encouraged reform plans that chased external benchmarks rather than local problem solving (Andrews et al., 2017). The World Bank (2017) treated this as a political economy problem. External actors could change incentives through finance and conditionality, yet they also pushed reforms that rewarded visible form over hard-to-observe enforcement. Capability trap accounts developed these critiques into mechanism claims. Pritchett et. al. argued that governments and donors adopted reforms that resembled high-capacity states, yet these reforms did not build underlying capability to deliver policies (Pritchett et al., 2010). They described

isomorphic mimicry and premature load-bearing as pathways to persistent failure, where laws, agencies, and plans multiplied while frontline practice stayed unchanged. This research treated failure as a political and organizational equilibrium, not as a one-time design error.

Street-level research clarified why frontline practice diverged from formal intent. Lipsky (1980) demonstrated how frontline practitioners rationed services, simplified rules, and exercised discretion when resources and time were short. In weak states, discretion combined with local power. Brokers and local elites determined who got benefits and who got sanctions. This combined capacity constraint with political control. It also moved analysis of policy failure toward mechanisms that linked center-level compromises to street-level selections. This literature converged on linked explanations. Political bargains shaped what policies rulers announced and what enforcement they allowed (Khan, 2011). Capacity gaps and bureaucratic incentives shaped whether agencies translated policy into routine action (Hanson & Sigman, n.d.). Systemic corruption and frontline discretion distorted delivery chains even when formal designs looked coherent.

A gap remained. Many studies treated these explanations as parallel. They rarely traced how structural features of hybrid orders produced specific failure mechanisms inside budgets, appointments, procurement, and service encounters. The capability trap literature moved closer to mechanism accounts, yet it aggregated diverse political settings under similar label. Hybrid regime scholarship explained incentives for selective enforcement, yet it rarely tested how those incentives travelled through delivery systems across sectors. This article built directly on these strands. It treated policy failure as a patterned outcome of linked mechanisms, which supported cross-national learning based on mechanism similarity rather than surface similarity in laws or agencies across key sectors too.

THEORETICAL FRAMEWORK

This theoretical framework treated policy failure in Pakistan as a causal chain that moved from structure to mechanism to outcome. Structural conditions set the constraints and rewards that policy actors faced. These conditions then triggered mechanisms, meaning repeatable processes that converted incentives into observable results. Outcomes appeared as weak policy design, unrealistic budgeting, uneven service delivery, selective rule enforcement, and repeated reform reversal. This approach followed the mechanism logic used in state capability research, where formal reforms could expand while real performance stayed weak (Woolcock et al., 2017). It also matched classic work on how “great expectations” failed once policies met really administrative and political conditions (Hasnan, 2022).

Pakistan-relevant structural conditions in this model covered seven categories. Fragmented authority multiplied veto points and blurred responsibility, which lowered the chance that any actor owned delivery results (Woolcock et al., 2017). Elite bargaining constraints pushed policy choices toward coalition survival and distributive bargains, which shifted design toward politically safe compromises (James & Lodge, 2003). Fiscal and extractive limits constrained recurrent spending and enforcement capacity, which widened the gap between announced targets and funded delivery chains. Bureaucratic incentives rewarded risk avoidance when posting security stayed weak and when sanctions looked uncertain or politicized. Centre-province tensions after devolution created contested mandates and uneven capability across tiers, which raised

coordination costs and delayed action (Hussain, 1999). The legal-judicial enforcement environment shaped credibility because slow adjudication and strategic litigation could reduce sanction certainty, then encourage procedural defensiveness. Information asymmetries widened agency loss because principals lacked reliable data on field behavior and service outputs.

Six mechanisms linked these structures to outcomes. Capture and rent reallocation treated policy as distribution, so actors redirected rules, exemptions, and contracts toward private gain (Shah, 2012). Credible commitment failure arose when leaders reversed rules after backlash, so firms and citizens expected reversal and reduced compliance investment. Principal-agent slippage emerged when monitoring stayed weak and incentives misaligned, so agents reshaped delivery and resources leaked at handoff points (Pressman & Wildavsky, 1973). Coordination and veto appeared through inter-agency conflict, file circulation, and forum shopping across overlapping mandates. Street-level discretion and coping described how frontline staff rationed services and reinterpreted directives under workload pressure and ambiguity (Woolcock et al., 2017). Symbolic compliance explained paper reforms that signaled progress while routines stayed unchanged, a pattern linked to isomorphic mimicry and capability traps.

The framework generated testable expectations. If capture operated, evidence should have shown repeated procurement anomalies, expanding exemptions, and selective enforcement in politically connected cases. If credible commitment failure operated, evidence should have shown policy reversals aftershocks, arrears build-up, and stalled long-horizon plans. If principal-agent slippage operated, audits should have repeated similar objections and reports should have looked strong while field outcomes stayed weak (Shah, 2012). If coordination and veto operated, agencies should have issued contradictory directives and shifted blame across forums (Woolcock et al., 2017). If street-level coping operated, access rules should have varied across localities and staff should have met targets through reclassification (Migdal, 1998). If symbolic compliance were operational, governments would have introduced new units and indicators without changing their budgets, without stability in tenure, and without using sanctions (Hedstrom & Swedberg, 1998). Cross-national learning was appropriate as lesson drawing that is sensitive to mechanisms. It asked which reforms altered the mechanism chain within the constraints of Pakistan, rather than which templates were modern enough. This was in line with the literature on lesson drawing and policy transfer that emphasized adaptation to domestic politics and administrative realities (Dolowitz & Marsh, 2022).

RESEARCH METHODOLOGY

This research uses a qualitative, comparative methodology rooted in the traditions of political economy and institutionalism. The research uses a case study methodology, with Pakistan as the key case, and uses cross-national comparisons based on a set of developing democracies that have followed similar paths of political instability but have reached different ends on reform outcomes. The data sources include secondary sources such as peer-reviewed articles in academic journals, policy studies, constitutional texts, governance datasets, and historical accounts. To add rigor to the analysis, the research uses process tracing to isolate causal mechanisms between institutional arrangements and political failure in Pakistan. The research uses most-similar systems design to derive comparative findings that can isolate institutional variables to explain different outcomes across cases. The research also uses thematic coding of governance datasets on accountability, civil-

military relations, electoral integrity, and state capacity. This research design approach is geared towards theory-building, rather than hypothesis-testing, and is focused on explanatory depth rather than breadth. The research design approach is particularly well-suited to studying the complex, historically embedded phenomenon of political failure, and allows for cross-national learning for reform.

POLICY FAILURE IN PAKISTAN: AN EMPIRICAL ANALYSIS

This section deals with policy failure as a patterned outcome of structural constraints and recurring mechanisms, not as a sequence of isolated mistakes. Comparative public policy research showed that failure clustered at predictable points in the policy cycle: agenda-setting, design, delivery, enforcement, and feedback. These points mattered because each one created a separate “decision gate” where political incentives, fiscal limits, and administrative behavior could distort the policy chain. Pakistan’s case fit this logic. The evidence base described repeated cycles where policy statements travelled faster than delivery capacity, and where political bargaining reshaped both policy goals and delivery rules. The analysis separated four linked problem sites. First, agenda-setting and problem framing shaped which issues entered government attention and how leaders defined causation and responsibility. Second, design choices translated political constraints into targets, instruments, and delivery chains (Husain, 1999).

Third, policy delivery depended on staffing, procurement, monitoring, and credible enforcement. Fourth, accountability and learning determined whether the state corrected errors or repeated them. Research on “capability traps” and “isomorphic mimicry” helped explain why formal reforms could look credible on paper while producing weak results in practice (Pritchett et al., 2010). Pakistan’s fiscal stress and repeated adjustment cycles also mattered for this section, because fiscal constraint shaped design and delivery in direct ways. Where budgets tightened, governments tended to protect politically salient spending, compress maintenance, delay payments, and shift burdens downwards. That pattern raised the risk of unfunded mandates and “paper compliance” systems that met reporting needs but failed to change outcomes (International Monetary Fund, 2024).

Failure in Policy Formulation

Agenda-Setting and Problem Framing Distortion

Agenda-setting in Pakistan tended to follow crisis attention and elite threat perception more than long-horizon problem diagnosis. Public policy research predicted this pattern when leaders faced high political turnover, weak party program discipline, and strong pressure to signal “action” within short cycles. Pakistan’s reform history also documented stop-start patterns where policy packages entered the agenda during fiscal or security stress, then lost momentum once immediate pressure eased (Husain, 2009). Problem framing influenced the set of solutions that were considered “reasonable.” Evidence from Pakistan’s governance discourse revealed that elite framing was capable of attributing responsibility to administrative failure or “lack of capacity” even when incentive conflicts and distributional bargaining were the drivers of outcomes. This framing constrained the space of admissible instruments and reforms that threatened rent streams or patronage connections. Framing also promoted “solution-first” packages in which leaders identified a preferred instrument and then looked for evidence to support the decision after the choice had

been made. Policy research terms this policy-based evidence, which is a known problem in conditions of weak accountability (Cyan & Pasha, 2017).

Design Trade-offs Shaped by Coalition Management

Bargaining in the coalition altered the design of policies by dilution, exemption, and ambiguity. Comparative research on veto politics revealed that multi-actor bargaining resulted in compromise instruments that maintained balance in distribution, rather than optimizing problem-solving capacity. The political economy literature on Pakistan also argued the same, where the government traded off policy consistency for coalition stability, and then used public institutions as bargaining venues through postings, contracts, and discretionaries (Husain, 2009). This mattered because it changed the "delivery chain" in the design phase. Unclear rules increased less political conflict in the short run, but they also increased discretion to frontline officials and mid-level managers later on. This increased the likelihood of discriminatory enforcement and unequal access, which then appeared as technical failure. Studies of transfers in Pakistan's civil service provided evidence for this channel by demonstrating how political and administrative incentives interacted to produce frequent transfers that reduced ownership of program outcomes (Cyan & Pasha, 2017).

Over-Ambitious Targets, under-Specified Delivery Chains, Unfunded Mandates

A common weakness in formulating targets was when the targets surpassed fiscal capacity and administrative ability. The study of public finance considered this a classic case of design fault, where the government set targets for outcomes without any serious costing, sequencing, and staffing, and left it to the agencies downstream to "find a way." In Pakistan, the tightening of fiscal policies and debt problems widened the gap between policy intentions and budgetary credibility. IMF program documents and World Bank updates explained how there was constant pressure to safeguard core spending while lowering deficits, which often squeezed discretionary operational budgets. This made it less likely to set ambitious service targets (International Monetary Fund, 2024). The shift towards performance budgeting in Pakistan was an effort to bridge this gap, but the legal requirement did not necessarily ensure a change in behavior. The Public Finance Management Act of 2019 introduced the requirement for performance budget reporting, which could help in better alignment of goals and resources. However, international experience indicated that performance management systems would fail if the indicators were treated as compliance documents by the agencies, or if the political leadership used the targets for signaling without funding the chain of delivery.

Donor or Imported Templates Versus Domestic Political Constraints

The external templates entered the Pakistani reform scene via loan conditionalities, technical assistance, and global 'best practice' scripts. The question was not about the role of external actors. The question was when external templates were more aligned with donor reporting requirements than with local political incentives and administrative facts on the ground. The capability trap literature explained the result: countries institutionalized structures that looked like high-capacity systems, but practices did not change. This dynamic preserved legitimacy in the eyes of external audiences while leaving service outcomes weak (Pritchett et al., 2010). Pakistan's fiscal and energy reform cycles illustrated the risk. Reform packages stressed tariff rationalization, loss reduction, and better governance of distribution. These goals could improve sector performance, yet domestic

politics constrained the pace of price adjustment and enforcement against theft or non-payment. Where leaders lacked credible commitment to sustained price reforms, design drift occurred through freezes, exemptions, and arrears accumulation. World Bank (2024) updates on the power sector reform agenda documented these pressures and their macro linkages.

Failure in Implementation

Administrative Capacity Limits Versus Incentive Problems

Capacity limits described what the state could not do with its existing skills, data systems, staffing levels, and managerial routines. Incentive problems described what the state chose not to do because actors gained from delay, discretion, or selective enforcement. The distinction mattered. Evidence from reform research showed that "capacity" language could hide political choice. This misdiagnosis then produced training workshops and new units, while core incentive conflicts remained intact. Pakistan's delivery challenges fit this pattern. Data and personnel shortages limited planning, procurement quality, and field supervision. On the other hand, incentive conflicts influenced postings, procurement, and enforcement selectivity. Studies on postings in Punjab showed that a combined political-bureaucratic logic influenced postings, which affected continuity and accountability for outcomes (Cyan & Pasha, 2017). Findings on compensation packages indicated that non-wage benefits and career rewards influenced behavior in ways that compensation scales could not fully capture, which made it difficult to explain "low pay causes poor performance" narratives (Andrews & Pritchett, 2017).

Intergovernmental Friction and Delivery Gaps

Devolution brought key service delivery responsibilities to the provinces after the 18th Amendment, although federal responsibilities were still present through finance, standards, and national programs. Research on comparative federalism indicated the stress of coordination when responsibilities were not clear or when the forums for intergovernmental interaction did not have binding dispute resolution mechanisms. Pakistan-focused studies indicated that devolution was expected to bring benefits, but it also posed risks during the transition period regarding coordination, capacity, and fiscal balance. There were delivery gaps due to mismatched mandates and funds. The autonomy of the provinces was increased, but they were highly dependent on the transfer of shared revenues, which made the service budgets vulnerable to macro adjustments. The World Bank's reporting on fiscal management indicated the need for improved intergovernmental coordination in the post-18th Amendment scenario, which indicated that devolution was not necessarily an automatic solution to coordination problems (World Bank Group, 2023).

Procurement, Staffing, Monitoring Systems, Leakage Points

Public procurement acted as a core vulnerability because it linked budgets to contracts and created opportunities for discretion. Pakistan's procurement rules and portals showed formal procedures for tendering and evaluation, yet external watchdog reporting and audit materials pointed to recurrent violations and weak compliance incentives (Public Procurement Regulatory Authority, 2025). Audit reporting reinforced the same diagnosis: weak internal controls, irregular payments, and non-compliance with public finance rules. Auditor General documentation across sectors consistently treated these as repeated control failures rather than rare exceptions. This pattern supported a mechanism claim. When procurement oversight lacked credible sanctions, actors

treated rules as flexible, then shifted risk to auditors after the fact (Gondal, 2024). Monitoring systems also faced data credibility problems. Global evidence showed that field monitoring reduced leakage when it changed expected punishment and raised detection probability. It failed when monitoring became performative reporting or when field staff faced pressure to falsify compliance. Pakistan-specific studies of aid and development projects noted weak monitoring and evaluation practice and limited use of findings in decision-making, a pattern consistent with principal-agent problems and goal conflict (Amin et al., 2024).

Enforcement Credibility: Inspection, Regulation, Sanctions, Compliance

Enforcement credibility depended on whether regulated actors believed that the state would detect violations and apply penalties with some consistency. Where enforcement stayed selective, actors recalculated incentives and treated compliance as negotiable. Regulatory political economy research described this as “credible enforcement” failure, a major driver of weak outcomes in tax, power, and service regulation. Pakistan’s macro and sector documents showed recurring stress points where the state struggled to sustain enforcement while also managing distributional politics. Energy sector reforms showed this tension in tariff setting, subsidy management, and arrears accumulation, which fed “circular debt” dynamics. Where enforcement linked to fiscal adjustment, credibility also depended on political willingness to absorb social backlash. IMF program materials discussed reform conditions and performance, which implied that sustaining enforcement required continued political commitment, not only technical fixes (World Bank Group, 2024).

Accountability and Feedback Loops

Policy systems learned when leaders demanded honest reporting, protected evaluators, and used findings to change incentives and budgets. Pakistan’s public sector showed partial movement toward performance reporting through the PFM Act framework and performance monitoring reports. Still, formal reporting did not guarantee learning. Cross-national research on performance budgeting showed that agencies could game indicators, report outputs that lacked outcome meaning, and treat targets as compliance exercises, especially when political leaders used reports for signaling rather than course correction (Alam, 2022). Empirical work on monitoring and evaluation in Pakistan-linked development settings also showed weak uptake of findings, shaped by information asymmetry and goal conflict. Agency theory research described how monitoring failed when principals lacked incentive to act on bad news or when agents-controlled information flows (Awan et al., 2024). World Bank materials on evaluation capacity in Pakistan treated demand for evaluation as rising, yet they still framed system strength as uneven and capacity-constrained, which implied limited routine use of evidence in policy correction.

Judicialization added another feedback channel. Courts sometimes constrained executive discretion and increased attention to legality, yet research on judicial activism in Pakistan argued that strong court intervention could also create uncertainty for administrators and raise risk aversion, especially when officials expected ex post scrutiny without protection for good-faith decisions (Munir & Khalid, 2018). In such settings, bureaucrats shifted toward procedural defensiveness, slow approvals, and file-passing as self-protection strategies. This behavior led to lower adaptive learning because the officials cared more about their personal security than about improving the program. The short political time horizon further increased the rate of reform change. The experience of Pakistan’s reform process was explained as a series of cycles in which new

administrations replaced the previous ones' programs, renamed the entities, and adjusted the targets. This reduced learning because the agencies devoted their efforts to rebadging instead of addressing the delivery bottlenecks (Hasnan, 2022).

Case Vignettes: Two Mechanism Chains (Structure → Mechanism → Outcome)

Vignette 1: Power Sector “Circular Debt” and Reform Drift

Fiscal distress, politically sensitive tariffs, fragmented sector governance, and lack of payment discipline created a situation where leaders were at high risk of backlash from price adjustment. IMF and World Bank reporting indicated reform pressure driven by macro stability and sector arrears. Leaders bargained tariff and subsidy commitments in exchange for temporary political relief. Agencies subsequently delayed payments, built up arrears, and turned to financial solutions rather than long-term loss reduction. Studies on Pakistan's power sector debt indicated that circular debt was associated with increased costs and macro distress, locking in the political economy trap (Malik & Mustafa, 2024). Arrears persisted, reform packages repeated, and the state carried continuing fiscal exposure. The vignette demonstrated a credible commitment problem driven by distributional politics, not a narrow technical gap.

Vignette 2: Post-18th Amendment Service Delivery and Coordination Stress

Devolution increased the power of the provinces, but financial reliance and a lack of clarity on role boundaries created pressures for intergovernmental coordination at the federal, provincial, and local levels. The World Bank and UNDP identified the lack of intergovernmental coordination in the post-18th arrangement as a problem (Shah, 2012). The provinces were characterized by unequal capacity and political incentives, and federal institutions maintained residual powers without effective coordination instruments. Budget constraints and political changes then led to program resets in the short term and unequal monitoring. This mechanism was consistent with expectations in comparative federalism, where common rule bodies had no enforcement powers. Policy outcomes differed among provinces, and intergovernmental coordination failures persisted, with feedback loops remaining weak. The vignette illustrated how structural decentralization could introduce new veto points and information asymmetries unless incentives and coordination rules were altered (Seidle, & Khan, 2013).

Cross-National Learning: What Travels, What Does Not

Comparator logic

This section used lesson drawing as comparison. Lesson drawing asked whether a program that worked elsewhere could travel, under what conditions, and with what redesign. Policy transfer research warned that imported models could arrive through donor leverage or expert networks, so design had to match domestic politics and administrative realities (Dolowitz & Marsh, 2022). Three cases were selected. Bangladesh served as “most similar,” given South Asian state formation pressures, donor density, and comparable service delivery stress. Indonesia served as “most useful variation,” because it combined rapid decentralization with national-level tools for procurement reform and public oversight in a large, diverse polity (Ahmad et al., 2007). Georgia served as “least similar but high-traction,” because it delivered visible anti-corruption change in policing and procurement through strong political backing and procedural redesign. These cases supported mechanism learning, not ranking. Each case illuminated a distinct route by which states had

reduced capture, strengthened commitment, improved monitoring, reduced veto friction, and improved frontline delivery.

Mechanism-Sensitive Lessons

Capture weakened as reforms decreased hidden discretion and increased contestability. Evidence from Indonesia found that the use of e-procurement was associated with fewer cases of procurement corruption in provinces with high levels of procurement expenditure. Studies of e-procurement in the decentralized state of Indonesia still reported workarounds, so control remained partial. In Georgia, open e-procurement minimized non-collaborative bargaining by posting tenders and contracts on the internet. Analysis of work remained alert to dangers such as single-source contracting, which indicated that open data required audit capacity and complaint mechanisms. The transferable lesson remained narrow. E-procurement narrowed some rent channels when governments ensured data integrity, rule enforcement, and genuine dispute resolution. Credible commitment improved when leaders linked reputational capital to rule stability and supported rule enforcement in a visible sector. Reforms in Georgia's policing and public services demonstrated a credibility reset in specific sectors when leaders combined personnel change with credible punishment.

In Pakistan, the lesson from Georgia might be to make enforcement visible against rule-breakers. The strategy hinged on secure posting control and credible punishment. Coalition bargaining and high personnel turnover undermined both in the Pakistani context. The lesson was transported as logic of design, not as tempo. The reform chain in Pakistan needed shorter rules, clear accountability, and safeguarding of crucial tenures.

The monitoring system was enhanced when service channels generated useful data and when managers responded to it. Bangladesh's a2i initiative expanded digital service projects and dashboards, and UNDP evaluation evidence mentioned broader use of digital technology for service delivery and decision-making. Peer-reviewed evaluation still mentioned limitations in terms of skills, funds, and political support, so data did not necessarily lead to correction. The evidence supported a bounded lesson. Digital monitoring was helpful when managers linked data to ground truth and budget allocations. Otherwise, reporting was mere window dressing. Decentralization eased veto power only when division of roles and budget rules remained transparent. Indonesia's experience demonstrated that ambiguous role transfers increased service delivery gaps and provided room for local control. For Pakistan, the lesson supported sectoral role definitions and rules for transfer before setting targets downward to provinces and districts.

Adaptation Constraints in Pakistan

Copy-paste failed when form travelled without function. Capability-trap research described "isomorphic mimicry" and "premature load bearing," where governments adopted advanced forms that exceeded local capability and incentive alignment. Pakistan faced tight fiscal space for recurrent costs and high posting churn tied to coalition bargaining. Centre-province coordination disputes also slowed delivery. Donor-backed models struggled when external financing slowed or ended, because recurrent costs still required domestic budget support. These constraints explained why Pakistan could not treat cross-national learning as template transfer. Pakistan needed mechanism-fit reforms that reduced discretion, strengthened monitoring, and protected rule

stability under fiscal pressure. Map the mechanism chain, test fit, then select an action. Fit used three tests. Did the reform reduce discretion at key leakage points? Did it generate usable information for supervisors? Did it create credible sanctions at a politically realistic cost (Dolowitz & Marsh, 2022). “Adopt” suited low-cost mechanisms with clear rule change. “Adapt” suited promising mechanisms that required sequencing, piloting, or political shielding. “Reject” suited mechanisms that relied on near-term conditions Pakistan lacked, such as stable enforcement and protected tenure.

Discussion: Implications for IR and State Governance

This article treated Pakistan’s policy failure as a source of external power loss, not only as a domestic management problem. In IR, bargaining power depended on material assets and on the state’s ability to make promises that others believed. The mechanism chain identified in this study suggested that repeated distortions in agenda-setting, design compromise, weak delivery control, and weak feedback reduced that belief. The result was a credibility gap. External actors priced that gap through tougher conditionality, narrower trust, and higher monitoring demands. The results helped to better understand what Pakistan as an IR actor looked like. It acted as a two-level state, where international bargains had to live up to the domestic coalition constraints and administrative style. The theory of two-level games foresaw that leaders could make promises abroad while facing limited “win-sets” at home. The case of Pakistan revealed a more extreme version of this problem. The domestic policy failure limited the state’s capacity to translate international commitments into consistent action, which in turn affected international expectations.

Moreover, the research on reputation confirmed this relationship. The leaders’ domestic decisions influenced foreign counterparts’ perceptions of resolve and reliability, not only of their past foreign policy behavior. The research on domestic politics in IO institutions also considered internal bargaining and domestic constraints as key drivers of external behavior. The internal governance failure then undermined external policy credibility in tangible domains. The IMF’s Governance and Corruption Diagnostic for Pakistan identified weak rule of law enforcement, dispersed oversight, and accountability gaps as fundamental constraints on reform and economic performance. These results were important in the context of IR because the IMF program functioned as a series of international bargains that had to be supported by consistent domestic follow-through. When follow-through stayed uneven, Pakistan faced tighter monitoring and higher reputational costs in financial diplomacy (Benson & Jordan, 2011). A similar logic applied to compliance regimes.

Pakistan’s exit from the FATF grey list in October 2022 showed that strong external pressure plus clear benchmarks could push domestic change. The wider lesson remained cautious. Episodic compliance did not remove the deeper incentive and delivery problems described in the mechanism model. The discussion also refined how domestic capacity translated into bargaining strength. Interview-based evidence from international economic negotiations showed that higher bureaucratic capacity improved preference attainment in bilateral deals. For Pakistan, the mechanism model implied that capacity was not only staff numbers or training. It was policy coherence across agencies, stable roles, credible enforcement, and usable monitoring. Where domestic delivery chains stayed weak, negotiators lost leverage because counterparts expected slippage after signing.

Trust research in IR treated credibility as relational and path dependent, shaped by perceived reliability and follow-through. Evidence from climate cooperation also linked credibility of national pledges to domestic political quality, which reinforced the claim that governance conditions shaped external belief (Berge & Stiansen, 2023). This mechanism model added to IR-informed policy scholarship in three ways. First, it moved past broad “weak state” labels by specifying where failure was produced inside the policy process, then tracing how that production reduced external credibility. Second, it linked domestic political economy to global governance pressures through a state transformation lens, where external standards reshaped domestic rules but also triggered local contestation and selective adoption. Third, it generated testable expectations: reforms that reduced discretion in procurement, protected monitoring integrity, and increased sanction certainty should raise international trust and widen negotiating room, even without large shifts in material power (Hameiri & Jones, 2016).

CONCLUSION

This study explained why policies failed in Pakistan by tracing a repeatable chain from structure to mechanism to outcome. It showed that fragmented authority, elite bargaining constraints, fiscal limits, unstable postings, center-province tension, uneven legal enforcement, and weak information systems shaped what policymakers could credibly promise and what administrators could deliver. These conditions activated six mechanisms. Capture redirected rules and resources toward narrow interests. Commitment failure raised reversal risk, so actors delayed compliance investment. Principal-agent slippage widened leakage where monitoring stayed weak and sanctions stayed uncertain. Coordination conflict created veto points, forum shopping, and blame shifting. Street-level coping converted vague directives into informal rationing rules. Symbolic compliance produced paper reforms that satisfied reporting needs while routine practice stayed unchanged. This chain explained why policy ambition and policy results separated, even when governments announced serious reforms.

The article advanced IR-informed policy analysis by specifying testable mechanisms that linked domestic political economy constraints to state credibility, not just to service outcomes. The analysis implied that reform needed to target discretion points, monitoring integrity, posting stability, and sanction certainty, because technical redesign without incentive changes repeated failure. Three limitations shaped interpretation. First, the study relied mainly on secondary evidence and official documents, so it captured patterns well yet it could not observe all informal bargains that shaped design and delivery. Second, mechanism inference faced equifinality, since different mechanisms could produce similar outcomes such as delay or under-spending. Third, sector variation remained under-specified, since the same mechanism could operate differently across energy, education, taxation, or policing. Future research should test the model through paired sector studies and province-level comparisons that traced the same policy instrument across different political and fiscal settings. It should measure posting churn, procurement contestability, audit objections, sanction timing, and compliance rates as observable markers of the mechanisms. It should also test whether targeted changes in monitoring and enforcement raised credibility in specific policy areas, then examine whether credibility gains improved Pakistan’s bargaining position in repeated external commitments, such as program conditionality and compliance regimes.

References:

Ahmad, J., Devarajan, S., Khemani, S., & Shah, S. (2007, May). *Decentralization and service delivery* (WB Policy Research Working Paper 3603). World Bank.

Alam, M. A. (2022, Jan. 19). Evaluating Pakistan's performance based budget. <https://blog-pfm.imf.org/en/pfmblog/2022/01/evaluating-pakistans-performance-based-budget?>

Amin, H., Malik, M., & Scheepers, H. (2024, Nov.). An agency theory unpacking of how monitoring and evaluation affect international development project impact. *International Journal of Project Management*, 42(8), 1-20.

Andrews, M., Pritchett, L., & Woolcock, M. (2017). *Building state capability: Evidence, analysis, action*. Oxford Academic.

Benson, D., & Jordan, A. (2011). What have we learned from policy transfer research? Dolowitz and Marsh Revisited. *Political Studies Review*, 9(3), 366-78.

Berge, T. G., & Stiansen, Ø. (2023, Sep. 28). Bureaucratic capacity and preference attainment in international economic negotiations. *The Review of International Organizations*, 18, 467-98.

Carothers, T. (2002). The End of the Transition Paradigm. *Journal of Democracy*, 13(1), 5-21.

Cyan, M. R., & Pasha, O. (2017). A symbiosis of civil service and politics in transfers: The case of Pakistan's management cadres. *Public Administration*, 95(4), 1077-91.

Dolowitz, D. D., & Marsh, D. (2022). Learning from abroad: The role of policy transfer in contemporary policy-making. *Governance*, 13(1), 5-23.

George, A. L., & Benette, A. (2005). *Case studies and theory development in the social sciences*. MIT Press.

Gondal, M. A. (2024). *Annual report: 2023-2024*. Department of the Auditor-General of Pakistan.

Hameiri, S., & Jones, L. (2016, Dec. 1). Global governance as state transformation. *Political Studies*, 64(4), 1-18.

Hanson, J. K., & Sigman, R. (n. d.). Leviathan's latent dimensions: Measuring state capacity for comparative political research. *The Journal of Politics*, 83(4), 22-34.

Hasnain, Z. (2010). Devolution, accountability, and service delivery: Some insights from Pakistan. *The Pakistan Development Review*, 49(2), 129-152.

Hasnan, Z. (2022). The politics of service delivery in Pakistan: Political parties and the incentives for patronage, 1988-1999. *The Pakistan Development Review*, 23(3), 129-51.

Hedstrom, P., & Swedberg, R. (1998). Social mechanisms: An analytical approach to social theory. 34(2), 234-45.

Husain, I. (1999). *The political economy of reforms: A case study of Pakistan*. Pakistan Institute of Development Economics.

Husain, I. (2009, Winter/Fall). The role of politics in Pakistan's economy. *Journal of International Affairs*, 63(1), 1-18.

International Monetary Fund. (2024, Jan. 19). *First review under the stand-by arrangement, requests for waivers of applicability of performance criteria, modification of performance criteria, and for rephrasing of access—press release; staff report; and statement by the executive director for Pakistan (IMF Country Report No. 24/17)*. Author.

James, O., & Lodge, M. (2003, April 2). The limitations of 'policy transfer' and 'lesson drawing' for public policy research. *Political Studies Review*, 1, 45-55.

Khan, M. H. (2011). *Political settlements and the governance of growth-enhancing institutions*. (Working Paper). <https://doi.org/https://soas-repository.worktribe.com/output/389203/>

Malik, A., & Mustafa, G. (2024). *Power sector debt and Pakistan's economy* (PIDE Working Paper). Pakistan Institute of Development Economics.

Migdal, J. S. (1998). *Strong societies and weak states: State-society relations and state capabilities in the Third World*. Princeton University Press.

Munir, K., & Khalid, I. (2018, Jul.-Dec.). Judicial activism in Pakistan: A case study of Supreme Court Judgments 2008-13. *Journal of South Asian Studies*, 33(2), 321-34.

Persson, A., Rothstein, B., & Teorell, J. (2012, Sep. 7). Why anticorruption reforms fail—Systemic corruption as a collective action problem. *Governance*, 26(3), 449-71.

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, 94(2), 251-67.

Pressman, J. L., & Wildavsky, A. B. (1973). *Implementation: how great expectations in Washington are dashed in Oakland*. University of California Press.

Pritchett, L., Woolcock, M., & Andrews, M. (2010, Dec.). *Capability traps? The mechanisms of persistent implementation failure* (Working Paper 234). Center for Global Development.

Public Procurement Regulatory Authority. (2025). Ensuring transparency in public procurement. <https://ppra.gov.pk/#/>

Seidle, L., & Khan, Z. (2013). *Federalism - 18th amendment report*. United Nations Development Program.

Shah, A. (2012, Sep.). The 18th constitutional amendment: Glue or solvent for nation building and citizenship in Pakistan? *The Lahore Journal of Economics*, 17(SE), 387-24.

Tsebelis, G. (2009). Decision making in political systems: Veto players in presidentialism, parliamentarism, multicameralism and multipartyism. *British Journal of Political Science*, 25(3), 289-325.

Woolcock, M., Pritchett, L., & Andrews, M. (2017). *Building state capability: Evidence, analysis, action*. Oxford Academic.

World Bank Group. (2023, Oct.). Pakistan development update: *Restoring fiscal sustainability*. Author.

World Bank Group. (2024, Apr.). *Pakistan development update: Fiscal impact of federal state-owned enterprises*. Author.

World Bank Group. (2017). *World development report 2017: Governance and the law*. Author.

Date of Publication	December 24, 2025
---------------------	-------------------