The Bureaucracy Is Part of the Treatment

A policy reaches people through approval chains, caseloads, frontline judgment and organizational incentives. When those arrangements change the service, they change what the evaluation is about.

The Bureaucracy Is Part of the Treatment

An evaluation says a new eligibility tool increased benefit enrollment. A second department buys the same software, trains its staff and expects a similar result. Six months later, applications are waiting in a longer queue.

The software may be identical. The authority to act on its recommendations may not be.

The people, places, organizations and figures in this document's examples are fictional.

In the first department, caseworkers could approve eligible applications. In the second, each flagged case required a supervisor's signature and additional documentation. The tool identified more cases, but the organization did not gain capacity to process them.

This example illustrates a problem with treating implementation as background scenery. The administrative arrangement can determine whether people receive the intervention, how quickly they receive it and what it consists of.

© 2026 Prova LLC · All rights reserved · PP-000243-10

01 Locate the decision that changes delivery

A policy team usually knows the visible object being introduced: software, a curriculum, a benefits rule or a procurement incentive. The causal question becomes clearer when it also names the decisions through which that object operates.

For an eligibility tool, those decisions might include whether staff see the recommendation before reviewing a file, whether they can override it, who approves exceptions and what happens when the queue exceeds capacity. Each choice can change exposure, timing or selection into the service.

There is empirical evidence that these arrangements matter. In a field experiment in Punjab, Pakistan, Bandiera, Best, Khan and Prat shifted authority from monitors to procurement officers and introduced financial incentives across 600 officers in 26 districts. Autonomy alone reduced prices by 9 percent without reducing quality; the effect was stronger where monitors tended to delay approvals until the end of the fiscal year.

That finding concerns a particular allocation of authority in a particular setting. It gives a reason to examine where approval power sits. It does not establish that removing oversight will generally improve procurement.

A study by Rasul and Rogger in Nigeria linked management surveys to independently assessed completion of approximately 4,700 public projects. Greater autonomy was positively associated with completion, while some incentive and monitoring practices were negatively associated. These are observational relationships, not randomized effects. Their relevance here is that variation in management belongs in the question being investigated.

© 2026 Prova LLC · All rights reserved · PP-000243-10

02 Separate what changed from the conditions around it

Not every contextual difference should be called part of the treatment. Geography, population characteristics and a pre-existing labor market can influence whether a result travels without being components of the intervention.

The useful distinction is between **what the policy changes**, **how that change is delivered**, and **the conditions under which it is expected to work**. Write those separately before deciding which belongs in the causal claim.

Suppose the policy deliberately delegates approval to caseworkers. Delegation is a component of the intervention. If the new tool increases referrals and thereby creates a queue, the queue may be part of the pathway through which its effects occur. A staffing level that existed before rollout may instead condition how well the intervention works. Those roles imply different evaluation questions.

The Magenta Book, section 1.8, treats process and impact evaluation as complementary. Process work examines delivery and its pathways; impact work asks what difference the intervention made. The distinction is useful because evidence of a bottleneck and evidence of an effect answer different questions.

A process study can show that households waited longer after additional approvals were introduced. A credible impact design is still needed to establish what those approvals caused relative to the relevant alternative. Conversely, an average impact estimate may leave the delivery team unsure which part of the operating model produced it.

© 2026 Prova LLC · All rights reserved · PP-000243-10

03 Collect the records that distinguish the explanations

The U.S. Government Accountability Office's *Designing Evaluations*, chapter 4, describes implementation questions about whether activities occur as intended, reach the intended population and encounter management problems. It also stresses the need for a criterion against which performance can be assessed.

For the eligibility example, "understand implementation" is too loose a data request. A more useful plan would distinguish three possibilities:

Possible explanation	Evidence to look for
The tool fails to identify eligible cases	Check a sample of recommendations and omissions against independently assessed eligibility
Correctly identified cases stall in approval	Trace timestamps, handoffs, unresolved exceptions and authority to clear the queue
Approved households still cannot access support	Trace notification, enrollment and receipt; ask affected households where the process breaks

These are proposed diagnostic checks, not findings from an actual program. Their purpose is to connect each data request to an explanation that would change the next decision.

Interviews and observation can be essential. A workflow chart may say that caseworkers can override a recommendation, while staff explain that doing so routinely triggers an audit. Records can locate delay; people may explain the rule or incentive behind it. The explanation then needs checking against other cases and evidence.

Collect the smallest set of information capable of distinguishing the consequential possibilities. More implementation variables do not automatically produce a better diagnosis.

© 2026 Prova LLC · All rights reserved · PP-000243-10

04 Define the delivery questions before the verdict arrives

A disappointing average result can prompt a search for an office characteristic that rescues a favored policy. With enough subgroups and operational measures, a plausible explanation is easy to find.

Before rollout, identify the few administrative conditions the intervention requires. Specify the delivery measures, the comparisons of interest and which analyses will be exploratory. If new explanations emerge, record them as new explanations and seek a test that could disconfirm them.

For the eligibility tool, a short implementation profile could state:

- who may approve, override, pause and resolve exceptions;
- the queues and handoffs a case must pass through;
- available staff capacity and relevant skills;
- the performance signals that staff respond to; and
- what counts as successful receipt, including timeliness.

The profile should describe the comparison service as well. Knowing how the new arrangement operates is insufficient if the existing arrangement remains a blank box.

© 2026 Prova LLC · All rights reserved · PP-000243-10

05 Make the operating conditions travel with the finding

A useful report allows another department to ask whether it can reproduce the conditions that mattered.

In the example, the headline “The eligibility tool increased enrollment” leaves several questions open. A more informative account would identify that caseworkers had approval authority, how exceptions were handled, the capacity available and which households actually reached the service. It would distinguish conditions supported by the evaluation from plausible explanations still awaiting a test.

That detail guides a practical scale decision: reproduce the relevant arrangement, adapt it deliberately or test a different one. Buying the same product is only one part of that decision.

Before transferring a successful policy, trace one case from eligibility to receipt. At each handoff, ask who can act, who can delay and what happens when demand rises. **The operating conditions that made the result possible belong beside the result.**

RELATED: The Trial Tested the Offer, for distinguishing assignment from the service actually received; What to do with disappointing evidence, for turning an unwelcome result into a bounded next decision.

© 2026 Prova LLC · All rights reserved · PP-000243-10

06 Further reading

1 HM Treasury, *The Magenta Book* (2026), section 1.8, p. 28, and section 2.2.3, pp. 51–53 — complementary process and impact questions.

2 U.S. GAO, *Designing Evaluations*, 2012 revision, chapter 4, printed p. 31 onward — implementation questions, comparison criteria and evaluation design.

3 Bandiera et al., "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats" (2021) — the Punjab procurement experiment

4 Rasul and Rogger, "Management of Bureaucrats and Public Service Delivery" — observational evidence from Nigerian public projects; the linked repository copy is the accepted manuscript.

QUESTION IT

Found an error, or want to question this document? Contact Prova at hello@byprova.com.

© 2026 Prova LLC · All rights reserved

Cite as: Prova (2026). The Bureaucracy Is Part of the Treatment. byprova.com

Examples are fictional unless sourced. Not legal or financial advice.