



Ministry
of Justice

Better by Design: **The Ministry of Justice Evaluation and Prototyping Strategy**

Contents

1. Introduction	5
2. Progress from 2023-2025	7
3. Why we need to go further	14
4. Our Guiding Approach	19
5. Focus for 2026–2029	22
6. Delivering the Strategy	32

Forewords

At the Ministry of Justice, we are responsible for policies and services that affect people at some of the most important moments in their lives. Achieving better outcomes requires a stronger, more deliberate approach to learning, one that ensures decisions are evidence-based, informed by real-world insight, and responsive to what works in practice.

Better by Design builds on our previous strategy and strengthens how a ‘Test and Learn’ approach can be applied consistently throughout the policy cycle. This requires collaboration across policy, operations and analysis and a shared commitment to learning and improvement.

We will place greater emphasis on prototyping to help reduce risk, strengthen design and ensure that policies are feasible and effective in real-world settings. This will be achieved by testing critical assumptions early, working closely with frontline staff and service users, and using evidence from delivery to refine and improve policies before they are scaled.

Evaluation remains central to this approach. High-quality, proportionate evaluation will continue to provide robust evidence on implementation, impact and value for money,

supporting decisions about what to continue, adapt or stop. This means considering evaluation during policy development and implementation and investing in the right data, methods and partnerships to understand what difference our policies make. Through high quality evidence we can guide future decisions with confidence.

By embedding ‘Test and Learn’ as a way of working, we will deliver a justice system that is more adaptive, more effective, and better equipped to meet the needs of the public.

Emma Churchill, Director General Policy

High-quality evidence is central to helping decision makers in the Ministry of Justice make better decisions in a complex and often uncertain environment. As Director of Analysis, I am committed to ensuring that the department’s analytical work not only generates robust evidence but is used effectively to inform choices about policy and delivery. This is not just about retrospective evaluation, but about ensuring evidence is available when it is most needed—supporting timely, confident decision-making.

The Strategy sets out how the department is embedding a Test and Learn approach, to ensure that evidence is generated and used at each stage of design and delivery throughout the policy cycle. Through Better by Design, we recognise that alongside robust evaluation, there is a need to adopt more innovative and agile approaches - in particular, prototyping, to explore uncertainty, test assumptions early, and provide timely insights from real-world settings.

By strengthening our collective capability and embracing a culture of continuous learning, we will ensure that evidence remains at the heart of decision-making—supporting a more adaptive, evidence-driven justice system.

Chris Drane, Director of Analysis

When we published our first MoJ Evaluation and Prototyping Strategy in 2023, we set out how our ambitions are centred around a simple message: better evidence enables better decision making which delivers better outcomes. For a world class justice system that works well for everyone in society, it is essential to develop a robust understanding of which policies and interventions work, which don't and how we can improve them.

With the publication of this second MoJ Evaluation and Prototyping Strategy, we report on progress made, and where we want and need to go

next. We have built strong foundations that have led to great examples of evaluation and prototyping in the department. MoJ analysts have invited collaborative challenge, transparently shared evidence, and are applying a wide variety of innovative methodologies in the justice setting.

Given MoJ's complex policy remit and challenging operational landscape, delivering on our evaluation and prototyping ambitions is a dynamic endeavour. There is still considerable potential to fill critical justice evidence gaps for better decision-making, and there are exciting opportunities to go further. To truly embed a Test and Learn approach, we will be inclusive in how we achieve this and inspire ourselves and others to produce timely evidence that 'shifts the dial' - all with the aim to improve justice system outcomes.

I would like to thank the specialist MoJ Evaluation and Prototyping Hub for driving many of our successes over the last few years. However, decision-focused evaluation and prototyping cannot thrive in siloed environments. A commitment to evaluation and to a Test and Learn approach is everyone's business. We are therefore grateful for - and look forward to continuing - our collaborations with current and new partners across the justice system, the civil service and academia.

Alexy Buck, Chief Social Researcher

Introduction

High-quality evaluation is essential for understanding whether government policies and programmes:

- Achieve the outcomes they intend
- Are delivered as designed
- Provide value for money for the public.

As set out in HM Treasury's Magenta Book¹ and Green Book², evaluation helps government learn what works, for whom, and in what circumstances, and supports better decisions about whether to continue, adapt or expand interventions.

Across government, there is increased focus on evaluation and understanding what works.

The Cabinet Office and HM Treasury established the Evaluation Task Force (ETF) in 2021 to strengthen evaluation practice, improve the use of evidence in spending and policy decisions, and support departments to plan and undertake proportionate, credible evaluation.

The Ministry of Justice (MoJ) is a major UK Government department responsible for the administration of the justice system in England and Wales.

It works to protect and advance the principles of justice, overseeing courts and tribunals, prisons and probation services, and a range of services to help victims, vulnerable people and those seeking access to justice. The department collaborates with numerous agencies and public bodies to ensure that sentences are served, offenders are managed effectively, and people's rights under the law are upheld.

The MoJ established the Evaluation and Prototyping Hub in 2021 to provide specialist evaluation and prototyping support to policy, operational and analytical teams. The Hub helps the department to meet cross-government expectations by using evidence systematically to improve outcomes and filling priority evidence gaps.

In 2023, the MoJ published its first **Evaluation and Prototyping Strategy**, setting out how the department would

¹ [The Magenta Book - GOV.UK](#)

² [The Green Book and accompanying guidance - GOV.UK](#)

strengthen the use of evidence, evaluation and early testing in policy and operational decision-making.

The 2023 Strategy articulated an ambition to embed evaluation earlier in the policy cycle, increase the use of proportionate and robust methods, and use prototyping to test assumptions and feasibility before committing to large-scale delivery. It also signalled a commitment to building capability, improving data and evidence infrastructure, and working more closely with delivery partners and external experts.

This updated Strategy builds on that work and reflects the progress made since 2023 in embedding evaluation and prototyping across the department.

Alongside the continued focus on evaluation, one of the ambitions of our 2026 Strategy is to place prototyping on an equal footing as a core part of how policies and services are designed, tested and refined. While evaluation remains essential for understanding impact, implementation and value for money, prototyping plays a complementary role by enabling early learning and risk reduction before decisions are locked in.

Prototyping involves working with those who use or deliver services to test early versions of an intervention in a safe, low-cost way. As the National Audit Office³ has set out, early testing can reduce delivery risk and strengthen the evidence base for future decisions.

MoJ's use of prototyping aligns with the government's ambition to make the state more agile and productive and supports the wider public sector reform agenda. The department has championed this approach across government and co-authored the recently published [Test and Learn Annex to the Magenta Book](#), setting out how early testing and evaluation can work together to improve policy and delivery.

Progress from 2023-2025

Since the launch of our first Evaluation and Prototyping Strategy in 2023, MoJ has made progress in how we learn and apply evidence. Over the past few years, MoJ has delivered high-quality evaluations and research that have shaped important decisions — from sentencing reform to offender management and access to justice. This has been enabled by delivering key actions outlined in the 2023 Evaluation and Prototyping Strategy:

- **Pillar One** – Establish processes to ensure proportionate evaluation and prototyping
- **Pillar Two** – Build capability to deliver quality evaluation and prototyping
- **Pillar Three** – Timely and accessible evidence to improve decision making

2.1 Pillar One

Establish processes to ensure proportionate evaluation and prototyping

- We have developed a bespoke tool, to provide a department-wide view of evaluation activity. This improves transparency, enables better coordination across policy areas, and supports more strategic prioritisation of analytical resources.
- We have refined the approvals process for evaluation and prototyping, to ensure that evaluations are proportionate and methodologically robust and ethically sound prior to investment.
- We have increased the adoption, and improved the quality of, Theory of Change models to identify risky assumptions and to structure prototyping and evaluation around testing the right questions.
- We have collaborated closely with the Evaluation Task Force to influence, adopt and embed cross-government methodological best practice.

2.2 Pillar Two

Build capability to deliver quality evaluation and prototyping

- We are conducting more impact evaluations using rigorous analytical methods, such as randomised controlled trials, difference in differences and propensity score matching (see [case study A](#)). Several of these evaluations have been conducted in complex and operationally challenging settings, including prisons, demonstrating that robust counterfactual methods are feasible in live justice environments. This experience has paved the way for wider, more routine use of these approaches in future evaluations.
 - MoJ's Data First programme funded by Administrative Data Research UK has continued to set the precedent across government in its transparent use of data and academic collaborations to deliver innovative and impactful new evidence. Nine justice datasets have been shared, with almost 70 academic research projects facilitated to date.
- Projects are leading to robust new insights across MoJ priority areas, such as ethnic disparities in the justice system and the intersection of care and criminal justice proceedings.
- We have invested in the learning and upskilling of MoJ analysts through the Evaluation and Prototyping Hub expanding access to practical resources to meet evaluation and prototyping needs, delivering targeted training, and continuing to act as a critical friend and advisor to priority projects across the department.
 - MoJ's Economics Hub has continued to bring together data and develop best practice for monetising the costs and benefits of delivering justice policies and services.

CASE STUDY A:

Building Causal Evidence in Prisons

Over the last three years, the MoJ has strengthened its ability to test interventions rigorously in operational prison settings. A central example was a randomised controlled trial (RCT) of Incentivised Substance Free Living (ISFL) wings across four Category C prisons, funded through the Evaluation Task Force Accelerator Fund.

What we were trying to learn

ISFL wings aim to support prisoners to live free from drugs and alcohol. An important element of the intervention was that ISFLs created a safer and more stable prison environment compared to non-ISFL wings: to allow people living on the wing the space, time and support to work on their recovery. But in practice, did this happen?

How we tested it

The evaluation focused on a critical assumption in the ISFL Theory of Change: that separating and supporting prisoners committed to substance-free living would improve safety and stability.

Using a waitlist RCT design,⁴ eligible prisoners were randomly allocated either to an ISFL wing or to remain on a non-ISFL wing. Outcomes were measured using administrative data, tracking incidents of assault, self-harm and disorder over time.

Because sample sizes are often small in prison-based research, the team used Bayesian survival modelling, allowing credible estimates of impact despite operational constraints.

What we found

Across all models, there was a high probability (over 90%) that ISFLs reduced incidents of assault, self-harm and disorder.⁵

The most likely effect size suggested around a 30% reduction in incident risk for prisoners on ISFL wings compared to those on non-ISFL wings.

What we learned

- Robust causal evaluation is possible in live prison environments when interventions are designed to be evaluable.
- Innovative analytical methods can overcome common constraints such as small samples and complex operational settings.
- Testing a clear mechanism — rather than the whole policy at once — produces more interpretable and decision-relevant evidence at earlier stages of policy development.

4 A waitlist randomised controlled trial (RCT) is an evaluation design in which eligible participants are randomly split into two groups: one receives the programme straight away, and the other waits for a set period. Outcomes are compared during the waiting period to estimate the programme's impact, after which the waiting group also receives the programme.

5 [A Waitlist Randomised Controlled Trial in Four Prisons: Estimating the Impact of Incentivised Substance Free Living Wings on Prison Stability](#)

2.3 Pillar Three

Timely and accessible evidence to improve decision making

- We have demonstrated how prototyping, while still relatively new to government, can deliver clear value. Teams who have used it describe faster progress, stronger collaboration with operational staff and better designed interventions (see [case study B](#)).
- We recently published the [MoJ Areas of Research Interest \(ARI\) 2025](#) which outlines MoJ's research priorities, aligned with the department's strategic objectives for the justice system and setting out where new research can have most impact for policy and practice.
- We have built new impactful academic partnerships with research networks and funders to embed academic evidence and expertise within the department and directly inform decision-making (see [case study C](#)).
- Many analysts are now embedded in policy teams, making it easier for analysts to be involved early in policy design and improve how evidence is used.
- We ensure that high-quality evaluation reports are routinely and systematically accessible to the public. Our transparency is reinforced through active participation in the cross-government Evaluation Registry.

CASE STUDY B:

Building Choices – A Phased Approach to Developing and Evaluating What Works

What we were trying to learn

Building Choices is part of His Majesty's Prison and Probation Service's Next Generation of Accredited Programmes delivered in prisons and probation settings. It is designed to help prisoners and those on probation to develop skills such as managing emotions, improving relationships and making better decisions — all of which are linked to reducing reoffending.

As part of the development of the programme, the MoJ wanted to understand:

- Was the right group of people being selected?
- Could the programme be delivered consistently in real-world settings?
- Were participants beginning to show early signs of positive change?
- Did delivery challenges differ between prisons and the community?

How we tested it

The team ran a small-scale “design test” across a small number of prisons and probation areas, gathering evidence in three main ways:

- **Listening to people's experiences:** interviewing participants, prison and probation staff, and court staff to understand what was working and what was difficult.
- **Tracking early changes:** Participants completed short measures to show whether they were developing skills such as emotional regulation and healthy thinking.
- **Reviewing delivery data:** We examined attendance, drop-out rates and practical delivery issues to see how the programme operated in different settings.

This test stage was not designed to prove long-term impact. It focused on identifying and refining delivery assumptions in its Theory of Change to strengthen the programme design prior to larger-scale evaluation.

What we found

The design test showed encouraging early signs, alongside clear areas for improvement.

- The assessment process broadly identified appropriate participants but required refinement.
- Some aspects of group size, session timing and guidance materials needed adjustment to improve delivery.
- Participants reported early improvements in key skills linked to behaviour change.
- Delivery conditions mattered: drop-out rates were higher in community settings, where participants faced competing responsibilities.
- Practical constraints — such as time pressures and lengthy guidance — affected consistency of delivery.

With the programme more stable, it has moved into the “Grow” phase. A randomised controlled trial and process evaluation is underway in early adopter sites to assess whether Building Choices improves short-term skills, for whom, how and why. A further quasi experimental evaluation is planned after wider rollout to assess impacts on misconduct and reoffending.

What we learned

- Testing foundational assumptions early reduces delivery risk later.
- Phased evaluation with iterative feedback loops builds a cumulative and credible evidence base.
- Understanding mechanisms and context is essential before focusing on long-term outcomes.
- Test and Learn supports better decisions about when — and how — to scale through collaboration.

Why we need to go further

Despite considerable progress, there is potential to go further and use evaluation and prototyping even more consistently and strategically at the MoJ. This Strategy focuses on addressing six system-wide challenges to further embed evidence and learning at the heart of policy and delivery.

3.1 Embedding practical evidence use throughout the policy and operational delivery cycle

In fast-moving policy and operational contexts, opportunities to test assumptions or design interventions in ways that support robust evaluation can be challenging, particularly where there is a need to move quickly to implementation. To address this, analytical insight needs to shape policy and operational delivery from the outset. Early testing and phased implementation help us understand what works, for whom and under what circumstances before committing to wider rollout. This approach supports learning and improvement throughout delivery, and increases the likelihood that later evaluation is timely, feasible and useful for decision-making.

3.2 Balancing rigour with pace

Policy and operational demands require timely insights that still meet robust analytical standards. By combining rapid evidence cycles with proportionate evaluation, we will ensure decisions are informed by credible evidence at each stage of the policy cycle development.

The increasing availability of digital and AI-enabled tools is changing how analysis, policy design and operational delivery are carried out across government. These developments create new opportunities to support timely insight and learning and reinforce the value of evaluation and prototyping in understanding how evidence is generated and used in practice.

3.3 Strengthening the conditions for prototyping

Innovative approaches benefit from structured opportunities to test ideas in real delivery settings before models are fixed. By working with frontline staff and justice system users to co-design and prototype options early, we can identify the approaches with the greatest potential to improve outcomes and invest in scaling those that show evidence of impact

3.4 Improving data access and infrastructure

Robust evaluation requires accessible, reliable and linked data. We will continue to strengthen the data foundations that underpin good evidence, including outcome measures, contractual data provisions, and data-sharing infrastructure, so that evaluations can be conducted efficiently and proportionately.

3.5 Building capability and confidence

We will support colleagues across policy, operations and analysis to further develop the skills needed to design, commission and interpret prototyping and evaluation. This includes building confidence in rapid-testing methods and strengthening expertise in economic and impact evaluation.

3.6 Fostering a Test and Learn way of working

To drive continuous improvement, we will embed ways of working in which evaluation and prototyping are used to learn and adapt. Approaching change with curiosity will help us focus on what interventions deliver meaningful improvements, and stop or redesign interventions when evidence suggests they are not achieving intended outcomes.

Summary

- **Shape policies and operations to be evaluable from the outset.** This will mean that evidence can be generated quicker and used to improve delivery in real time. Designing interventions with clear outcomes, explicit assumptions and planned learning points makes it easier to test what is working early, identify issues before they become embedded and adapt approaches without costly rework.
- **When feasible, move rapidly from idea to action by involving frontline colleagues early.** This means co-creating and testing prototypes in live environments within the first phase of policy development, so we surface and test critical assumptions and improve feasibility before deciding which option to scale.
- **Foster collaboration across disciplines.** This means bringing together analysts, policymakers, operational staff and delivery partners, such as NHS organisations, education providers and charities, to generate insight from real-world implementation, build prototyping and evaluation capability, and promote a shared understanding of how evidence can inform decisions.
- **Embed learning and evaluation in decision-making.** This means making evaluation findings visible, timely and influential, ensuring that analysis, monitoring and evidence shape policy and operational choices before, during and after delivery. This aligns with the government's ambition for a more agile, productive state that delivers better outcomes for the public.

CASE STUDY C:

Collaborations with academia and research funders

What we were trying to learn

The MoJ has committed to working with academics, research networks and funders to strengthen evidence for decision-making across the justice system.

The department wanted to understand whether a more structured, partnership-led approach to academic engagement could help address long-standing evidence gaps.

How we tested it

Through the Evidence and Partnerships Hub, MoJ has built and coordinated collaborations aligned to its Areas of Research Interest, working with academics and funders to design evaluations that are both methodologically robust and policy relevant.

At a system level, this involved investing in clearer engagement routes for academics, better alignment between policy questions and research design, and the use of shared infrastructure such as linked administrative data. This approach has since been recognised by the Government Office for Science

as representing the gold standard for academic engagement, reflecting its clarity, openness and impact.

One specific example is a partnership between MoJ, the Cabinet Office Evaluation Task Force, the Economic and Social Research Council, and Professor Ian Brunton-Smith (University of Surrey). This collaboration supported a quasi-experimental evaluation to understand the impact of radio frequency electronic monitoring (RF EM) on reoffending. RF EM uses a tag worn by an individual to check that they are present at a specified address during curfew hours.

What we found

This evaluation found that RF EM was associated with lower levels of proven reoffending within 12 months compared with matched control groups. They acted as both a situational barrier to reoffending as well as reinforcing the effectiveness of community sentences.

Three evaluation reports and a technical report were published on GOV.UK in May 2025.

What we learned

- Academic partnerships can unlock methods, expertise and capacity that are sometimes difficult to develop internally.
- Access to high-quality linked data is critical for robust evaluation.
- Aligning research to clear policy questions increases relevance and impact.
- External collaboration strengthens credibility and confidence in findings.
- Well-designed academic partnerships can deliver rigorous evidence in complex justice settings.
- Linking administrative data at scale materially improves the feasibility and quality of evaluation.
- Early alignment between policy needs and research design increases both relevance and impact.

Across this and other collaborations, the Evidence and Partnerships Hub approach has demonstrated that:

The Hub will continue to invest and scale these academic collaborations, including an expansion of our secondment schemes.

Our Guiding Approach

Our approach draws on the new [Test and Learn Annex to the Magenta Book \(2026\)](#). When the problem is complex, when there is uncertainty about the best approach, or conditions are changing, the best policies are built through collaboration, curiosity, and continuous improvement. Test and Learn, as set out in the Magenta Book, is built around four iterative stages — Explore, Co-design, Test, and Grow — which can be applied flexibly to different types of reform and align with the ROAMEF policy development cycle (as included in Figure 1 below).

Test and Learn is a way of working that uses rapid, iterative feedback loops to check whether key assumptions behind a policy or service hold in practice, adapting quickly when they do not. Instead of designing a full solution upfront, we start small, test the uncertain and riskiest elements early, and

refine the approach based on evidence from delivery settings. This helps ensure interventions meet user needs and fit delivery conditions.

Once an approach becomes stable and well understood, we will use counterfactual evaluation, whenever feasible, to establish whether the intervention delivers its intended outcomes and value for money. This provides credible evidence to decide whether to scale, adapt, or stop an intervention.

Taken together, and as set out below, this end-to-end approach reduces the cost of failure, strengthens design, and increases the likelihood that government policies deliver the intended benefits.

Explore: Building a shared understanding

We aim to bring together relevant stakeholders to build a shared understanding of the system we want to influence. Through system mapping and early evidence reviews, we will seek to identify where an intervention is most likely to make a difference.

Co-design: Working with those who deliver and use services

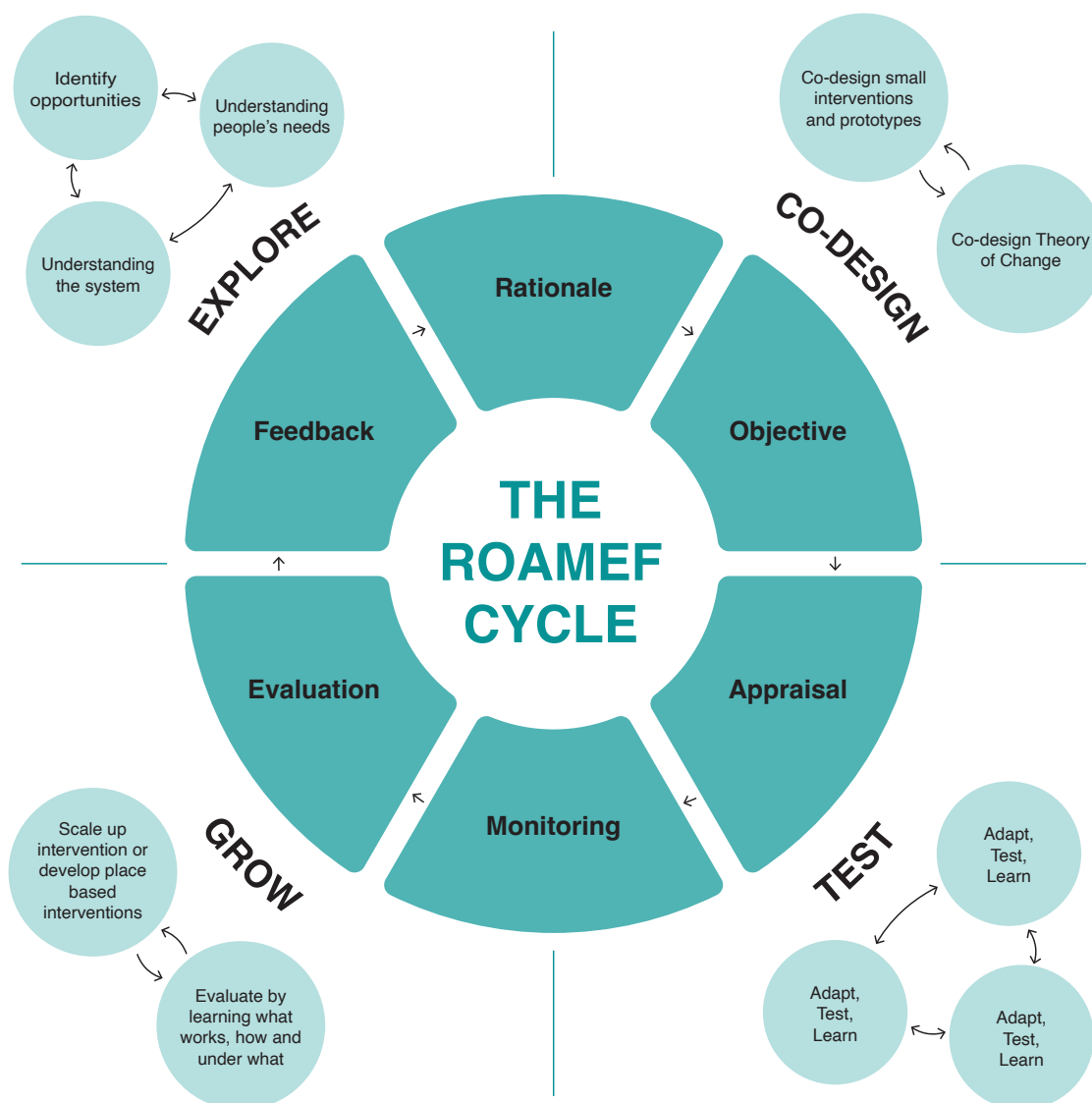
Where appropriate, we will involve frontline staff, delivery partners and service users in shaping ideas and developing multiple prototypes. This approach helps ensure that new interventions are practical, credible and grounded in lived experience, particularly when delivery contexts are complex or uncertain.

Test: Learning what works in practice

We will seek to test the riskiest or least certain parts of a policy on a small scale if feasible, using proportionate and rapid research methods to identify what needs to change before wider rollout. Depending on the decision at hand, this can include qualitative sprints, user testing, mechanism experiments, or nimble randomised and quasi-experimental approaches to understand feasibility, behavioural responses, early outcomes, and unintended effects.

Grow: Evaluating at scale

Once interventions are stable and operating as intended, we will undertake robust evaluations — prioritising counterfactual impact assessments where feasible — to understand effectiveness and the conditions for success at a programme level. Through this, we are renewing our commitments to robust evaluation as set out in our 2023 Evaluation and Prototyping Strategy.



Integrated with ROAMEF, the Test and Learn approach supports earlier, smaller-scale evidence generation to shape policy design. It encourages teams to examine risky assumptions, engage users and frontline staff, and iteratively adapt interventions as new information emerges. This strengthens Appraisal and Monitoring and ensures that, by the time formal Evaluation is undertaken, the intervention is stable, understood, and evaluable by design, consistent with Magenta Book principles.

Figure 1: Test then Invest - How the Test and Learn Framework fits with the ROAMEF Cycle

Focus for 2026–2029

In order to address our ambitions as set out in chapter 3, over the next three years we will focus on three pillars:

5.1 Pillar One

Further improve systems to support Test and Learn and proportionate evaluation

Action One: Focus evaluation efforts where evidence matters most

We will focus evaluation effort on the areas where evidence can have the greatest impact on outcomes, delivery and value for money, supported by clear priorities and a coordinated, proportionate approach to evaluation.

Key areas of MoJ interest over the period of this Strategy include major reforms and operational priorities with significant system-wide implications. These include:

- Monitoring the implementation and impact of the Sentencing Act.
 - Understanding the impact of measures to improve the efficiency of the criminal courts system.
 - Strengthening the evidence base on how best to supervise and support people on probation.
- To support this focused approach, we will take the following steps to ensure evaluation effort is strategic, proportionate and aligned to key decision points:
- Set and regularly review departmental evaluation priorities every six months, ensuring analytical effort is aligned with the most important policy and operational decisions and areas of greatest uncertainty.

- Prioritise proportionate, high-quality impact evaluations, applying counterfactual methods where feasible, alongside mixed-method approaches so that findings are available when key policy, delivery and funding decisions are being taken.
- Improve access to external expertise where needed, to support timely commissioning and delivery of high-quality evaluation.

Action Two: Prototype with the front line, in the front line

We will strengthen our design capability by partnering with those closest to delivery, ensuring frontline expertise informs early prototyping and helps shape policies that are practical and workable on the ground.

- **Explore the problem before designing solutions** – utilise analysts at the outset of the ROAMEF cycle to better understand the problem from the perspective of operations, delivery partners, and service users. Triangulate frontline insight with analysis of administrative data, so experiences are understood in context.

- **Co-create and prototype at the frontline** — multidisciplinary teams work alongside practitioners and service users to design, test and refine multiple ideas in real-world conditions before narrowing down the preferred approach (See [case study D](#)).
- **Nurture small-scale innovation** — continue to provide internal seed funding through innovation grants to support local experimentation, allowing teams to develop early prototypes that address practical challenges.

CASE STUDY D:

Co-Design and Small-Scale Prototyping to Test a New Working Week in Prisons

What we were trying to learn

Prisons face long-standing challenges in providing enough purposeful activity, leaving many prisoners without regular work, training or education.

While the Prisons Strategy White Paper (2021)⁶ set out the potential benefits of a longer working week, there was significant uncertainty about whether a full 35-hour working week could be delivered safely and practically in a real prison environment.

Before designing a national model or investing at scale, the key question was: is this operationally feasible?

How we tested it

Rather than designing a full policy upfront, the MoJ Offender Employment Policy Team worked with the MoJ Evaluation & Prototyping Hub and frontline prison staff to run a small-scale, live prototype.

A one-week prototype was tested in a single workshop, in one prison, using existing facilities and staff.

Simple changes — such as providing packed lunches to reduce prisoner movement — allowed the team to observe how staffing, safety, scheduling and supervision operated under a full working day.

This low-cost, time-limited test focused on learning from real delivery conditions, not on proving impact.

What we found

The prototype surfaced practical issues that would have been difficult to anticipate through desk-based design alone, including:

- Pressure points around staffing and supervision
- Knock-on effects on movement, meals and regime planning
- Operational adjustments needed to make longer working days viable
- The test also demonstrated that a longer working day could be delivered safely in controlled conditions, provided processes were redesigned.

What we learned

It demonstrates how co-design and prototyping support better-grounded, more deliverable justice policy — and why early testing should be a routine part of policy development where feasibility is uncertain.

- Early, live testing reveals operational realities quickly and safely.
- Frontline staff insight is essential for understanding feasibility.
- Small changes to routines can unlock larger system shifts.

Action Three: Leverage partnerships to accelerate evidence generation

Partnership remains central to how we learn and improve. Working with the MoJ Evidence and Partnerships Hub, aligned with their successful academic partnerships programme that builds on the Areas of Research Interest (ARI), we will continue to:

- Work with other government departments, the Cabinet Office Evaluation Task Force, the Cabinet Office Test, Learn, Grow team and HM Treasury to ensure evaluations support investment and spending decisions.
- Build on our successful collaborations with academia and research councils to leverage external expertise and funding (see [case study C](#))
- Engage delivery partners and service users as active participants in testing and learning. This includes understanding effective approaches in engaging justice system users in the co-design, co-production and dissemination of research.

Through these partnerships we will ensure our evidence base reflects the complexity of the justice system and supports innovation at every level.

Action Four: Invest in data and infrastructure

As outlined in the MoJ's Digital Strategy 2025⁷, we will enhance our data and analytics capabilities to inform policy decisions, improve operational efficiency, and deliver better outcomes for users. By creating a comprehensive data ecosystem, the MoJ seeks to foster transparency, accountability, and evidence-based decision-making across the justice system. To do this we will:

- Continue to develop our data management policies (including data quality, data ownership, data standards and data cataloguing standards) across the various datasets and agencies.
- Evaluate new data products ensuring they are fit for purpose and value for money.

5.2 Pillar Two

Invest in People

Action Five: Capability & Skills - building multidisciplinary confidence

We will strive to strengthen skills and confidence across policy, analytical and operational teams to apply Test and Learn methods in practice.

- Further test the model of embedding a prototyping expert within a prison's Senior Leadership Team — to assess how effective and scalable it is in enabling rapid, collaborative problem-solving to improve local performance (see [case study E](#)).
- Run a regular “Test & Learn speaker series” — build on MoJ's successful monthly internal ‘Area of Research Interest Academic Series’ to grow our practical knowledge of early prototyping, and how to set up robust impact evaluations using counterfactual methods, through justice-specific case studies.

Action Six: Continue to invest in the Evaluation and Prototyping Centre of Expertise

We will continue to invest in a strong centre of expertise to build and sustain the department's capability in monitoring, prototyping, evaluation

and evidence-informed decision making. This function will support a more consistent, proportionate and strategic approach to evaluation and prototyping use across the department, aligned with wider cross-government ambitions. The centre of expertise will:

- Provide strategic evaluation and prototyping leadership — supporting alignment between departmental priorities, evidence needs, and analytical activity.
- Set and uphold standards of quality and proportionality – ensure that evaluation and prototyping activity is methodologically robust, ethical and focussed on producing evidence that is useful for decision-making.
- Model and promote good practice — demonstrate how a range of evaluation and prototyping approaches, including prospective impact evaluations, can be applied effectively across different policy and operational contexts.
- Support teams across the department — offer expert advice, guidance and practical examples so that effective evaluation and prototyping approaches can be adopted.

CASE STUDY E:

Embedding Prototyping Expertise to Strengthen Local Problem-Solving in a Prison

What we were trying to learn

Despite sustained effort and investment, some long-standing performance challenges in prisons can be difficult to shift. This case study explored a different question: could embedding specialist prototyping and problem-solving capability directly within a prison's leadership team help staff understand problems differently and generate more effective local improvements?

Rather than introducing a new policy or programme, the uncertainty was about how change happens at the frontline.

How we tested it

A small-scale test was run in one prison, embedding a Service Designer with expertise in prototyping within the Senior Leadership Team for 12 months.

The role was designed to complement — not replace — operational expertise. The embedded specialist worked alongside leaders and staff to:

- Map complex, long-standing issues in the prison
- Surface assumptions about what was causing problems
- Involve staff at different grades in diagnosing issues
- Test small, low-risk changes and adapt them quickly based on feedback

This approach created space to slow down thinking, examine root causes, and learn through doing.

What we found

Staff reported that having dedicated, independent capability embedded in the prison helped them approach problems in new ways. The prototyping approach contributed to tangible operational improvements, including digitalising gate-pass and legal visits processes which reduced errors and saved an estimated 1,800 staff hours per year.

Staff also described:

- Increased confidence in diagnosing and addressing problems
- Stronger collaboration across teams
- A shift towards a more open, learning-focused culture

What we learned

- Proximity to the frontline matters — insight and learning increase when expertise sits inside delivery teams.
- Structured service design and prototyping help teams move beyond symptoms to underlying causes.
- Small, iterative tests can deliver practical improvements while building confidence and capability.
- Learning-focused approaches can support cultural change as well as operational gains.

The early findings suggest this approach is promising, and that further testing across additional sites would help understand its scalability and long-term impact.

5.3 Pillar Three

Developing a collaborative Test and Learn Environment

This strategy sits within a broader cross-government context in which innovation, fast learning and managed risk-taking are seen as essential to improving productivity, resilience and public service delivery. The National Audit office has emphasised that the government must embrace a culture of fast learning and evaluation and articulate a clear risk appetite for innovation; recognising that not every test will succeed but that rapid insight and adaption are key to unlocking productivity and strengthening resilience across public services.⁸

Recent Ministerial speeches have similarly underscored the need to shift from seeking to “get things right first time” towards approaches that allow early testing, rapid refinement and fixing things quickly in response to evidence.⁹ Together, these cross-government perspectives reinforce the central role of evaluation and prototyping in delivering better outcomes for citizens.

Action Seven: Enable people to learn and improve together

Through our prototyping and evaluation work we will promote ways of working that support learning, adaption and continuous improvement

- Promote a “safe-to-learn” approach to improvement, where prototyping and evaluation are used to test assumptions, surface issues early and support rapid learning. This approach recognises that not every test will generate the expected results, but that early insight helps teams adapt quickly, reduces delivery risk and avoids costly rework.
- Share and celebrate case studies — such as the [Working Week Prototype](#) and [Building Choices Programme](#) — to highlight how adaptive, evidence-informed policymaking strengthens design and delivery.

Action Eight: Support development of promising prototypes through early testing

Our ambition is to create space for controlled, early-stage testing to explore how policies and interventions in priority areas work in real-world settings before committing to full rollout.

- When the problem or delivery contexts are uncertain, the Evaluation and Prototyping Hub will work with colleagues to develop a small number of early, small-scale prototypes to test different options. Learning from this work will be shared through case studies to support wider adoption of this approach across the department.

- Where outcomes of interventions have remained difficult to shift, the Evaluation and Prototyping Hub will support small-scale, high-learning tests. These tests will help teams examine underlying assumptions, understand delivery conditions, and identify where different approaches could deliver better results more quickly.

Action Nine: Use evaluation to shape policy and inform decisions

We will strengthen further our approach to evaluation so that evidence informs decisions throughout the policy cycle.

- Embed evaluation and evidence use in programme boards and governance, introducing clear stage-gates that specify what evidence is needed before progressing from testing to wider rollout.
- Adopt a phased approach to learning from prototyping and evaluation, recognising that questions about feasibility, mechanisms, impact and scalability require different methods at different stages. We will distinguish early promise from demonstrated impact, ensuring that expansion of major or higher-risk policies is based on proportionate and appropriate evidence.

Delivering the Strategy

This Evaluation and Prototyping Strategy sets out a clear direction for how the MoJ will continue to strengthen the use of evaluation and prototyping.

We expect to see several important shifts as a result. We aim to achieve a more balanced evaluation portfolio, with greater use of mixed-methods approaches and counterfactual impact evaluation, alongside process and implementation studies. This will strengthen the department's ability to understand not only how policies are delivered, but whether they achieve their intended outcomes and represent value for money.

We also expect to see more consistent use of prototyping in the early stages of policy and operational development, particularly where problems are complex or delivery conditions

are uncertain. Earlier testing of assumptions, feasibility and mechanisms will help reduce risk, improve design, and ensure that larger-scale evaluations are undertaken on interventions that are stable, well understood and evaluable by design.

Taken together, these changes will help to support better decision-making, improving outcomes, and ensuring that public resources are used as effectively as possible. This Strategy therefore marks the next stage in embedding a mature, proportionate and practical Test and Learn approach across the MoJ.