Test and learn

A playbook for mission-driven government

Authors

Martina Vojtkova, Laurie Smith and Suraj Vadgama

About Nesta

Nesta is the UK's innovation agency for social good. We design, test and scale solutions to society's biggest problems. Nesta's three missions are to give every child a fair start, help people live healthy lives and create a sustainable future where the economy works for both people and the planet.

For over 20 years, Nesta has worked to support, encourage and inspire innovation. We work in three roles: as an innovation partner working with frontline organisations to design and test new solutions, as a venture builder supporting new and early-stage businesses and as a system shaper creating the conditions for innovation. Harnessing the rigour of science and the creativity of design, Nesta works relentlessly to change millions of lives for the better.

Find out more at nesta.org.uk.

About BIT

BIT is a global research and innovation consultancy which combines a deep understanding of human behaviour with evidence-led problem solving to improve people's lives. We work with all levels of government, nonprofits and the private sector, applying behaviour science expertise with robust evaluation and data to help clients achieve their goals.

Find out more at bi.team.

If you'd like this publication in an alternative format such as Braille or large print please contact us at: information@nesta.org.uk

Acknowledgements

Creating this playbook has been a collaborative effort, and we are deeply grateful to everyone who contributed their time, insights, and expertise. Our thanks for invaluable feedback, insightful case studies and constructive challenge to our Nesta colleagues Moria Sloan, Madeleine Gabriel, Sarah Cattan, Hugo Harper and Tom Leach and to BIT colleagues Elisabeth Costa, Deelan Maru, Michael Hallsworth, Rachel Coyle and Giulia Tagliaferri. Special thanks for their support in reviewing early drafts and sharing government perspectives to Miriam Styrnol and Jack Blumenau from the joint Cabinet Office-HM Treasury Evaluation Task Force and to Jessica Hunt from the Department for Education. Your willingness to share your experiences and best practices has been instrumental in making this playbook relevant and actionable. This playbook is a reflection of your collective knowledge and generosity, and we hope it serves as a useful resource for many. Thank you!

"Test it. Fix the problems. Change the design. Test it again. Tweak it again. And so on, and so on, for as long as you provide the service. Suddenly, the most important question isn't, 'How do we get this right the first time?'. It's 'How do we make this better by next Friday?"

Rt Hon Pat McFadden, Chancellor of the Duchy of Lancaster
Public Sector Reform speech, 9th of December 2024

Contents

Executive summary

To achieve its bold new missions the UK government will need a totally different way of working. Instead of placing big bets on assumptions about ideas when they are first proposed, policies need to be reviewed and improved throughout their development via a process of test and learn. This playbook explains what test and learn is, how to use it throughout the policy cycle and provides a methods toolkit that can be deployed at each stage.

What is test and learn?

Test and learn is an iterative approach to public policy and service development. Conventional policymaking typically uses a step-by-step process called "waterfall" where most of the big choices are made upfront. Yet we know least about how to solve a problem at the beginning.

Instead, government should adopt an iterative approach, starting small, testing critical assumptions early and often through rapid learning cycles, and learning from real-world delivery. Robust evidence informs early design, and appropriate evaluation methods assess whether optimised policies and programmes are achieving results. Insights from local experiments and frontline practice guide wider policy and service rollout. As these are scaled, they should be refined through ongoing test-and-learn cycles, ensuring they remain effective, responsive, and grounded in evidence.

Importantly, test and learn should not be understood as only involving a series of small design experiments that can replace rigorous evaluation. Done well, test and learn brings together iterative development and robust evaluation to deliver on long-term outcomes such as the UK Government's mission goals.

What would an ambitious UK government test and learn programme look like?

There are already encouraging examples of test and learn within the UK government. In the mid-2010s the huge Universal Credit scheme, initially developed using traditional waterfall methods, faced serious implementation challenges. A

strategic reset that applied test and learn saved the programme. However, applying this method to policy tends to be the exception rather than the rule.

An ambitious test and learn programme would build an R&D ecosystem for the UK government, using test and learn approaches to de-risk new complex policies, optimise existing services, and design and scale new public programmes. Iterative 'learning loops' would test critical assumptions early and provide feedback to ensure data and evidence regularly inform policy design and implementation.

To achieve this goal, the UK government can draw on the experiences with test and learn from other sectors. At Nesta and BIT, for example, we embed test and lean in how we work towards our missions and with our partners. For example, we used test and learn techniques like iterative prototyping in our Visit a Heat Pump scheme and rapid, affordable nimble trials in the evaluation of the National Tutoring Programme.

What is needed to deliver test and learn successfully?

The UK government already has many of the building blocks required to work in a test and learn way. This includes the joint Cabinet Office-Treasury Evaluation Task Force, £100m funding for place-based test and learn initiatives, and the essential capabilities required to work in a test and learn way such as designers, data scientists and evaluators.

However, embedding test and learn at scale will require the government to change the way it operates. This will require:

- **Outcome-based accountability** – teams tasked with delivering outcomes not policies, activities or outputs. This will require firmly defining goals while remaining flexible in the ways to reach them.
- **A shift to agile ways of programme design and delivery** – away from linear "waterfall" approaches where everything is decided upfront; agile is not new but remains the exception in government.
- **Changes to funding rules** – funding that increases when risk decreases as more is learnt about programmes alongside more locally controlled funding.
- **Flexible procurement mechanisms** – enabling more flexible contracting with more decision points part way through delivery.

- **New units of delivery** such as multi-disciplinary teams combining policy, delivery, operations, service design, innovation and evaluation expertise.
- **Improved data infrastructure and digital capabilities** at central and local government and improved access, linkage and use of data.
- **More local capacity and new capabilities** to innovate, monitor, evaluate and iteratively adapt.
- **Legislative frameworks** that allow the creation of "sandboxes" for local experimentation.
- **Working in the open** – transparently sharing insights, successes, and failures.

Test and learn approaches can be used for almost any policy, service or programme. A good place to start is with those that have:

1. **Available data** – particularly on leading indicators that signal the likelihood of success.
2. **Tractability** – where we have a basic understanding of the problem and mechanisms.
3. **Established provider(s)** – able to deliver relevant services.
4. **Delivery infrastructure to scale** – to double down on opportunities with the potential for growth.
5. **Recognition and appetite to solve the problem** – especially by local partners where most of the delivery capacity sits.

While government may want to start applying test and learn to the most tractable problems, the technique is well placed to tackle knottier challenges like the transition to net zero as test and learn can help de-risk new complex policy and handle the associated uncertainty and evidence gaps.

Why should the Government adopt test and learn?

At a time when public services face mounting pressures and public finances are constrained, test and learn is critical to enabling the government to cost-effectively deliver on its five missions and the Plan for Change. This approach would de-risk

policy and funding decisions, help navigate the uncertainties of addressing complex problems, and accelerate the delivery of impactful solutions.

The UK Government now has a unique opportunity to try a different way of doing things. If the right choices are made about test and learn today then by the end of the next Parliament, UK public services will be transformed for the benefit of the British people.

Background

The UK government has a problem. We are facing some of the toughest policy and public service challenges in a generation – climate change, stagnant growth, and persistent inequalities. These are highly interconnected, requiring coordinated responses across government departments. The rapid pace of technological, social, and economic change makes long-term planning difficult. Public expectations are also rising. People want services that are responsive, personalised, and joined-up– more like what they experience in the private sector. Our legacy policymaking tools weren't built for this kind of complexity.

The Government's Plan for Change reflects the scale of this challenge. It sets ambitious targets across five national missions, from improving school readiness to reforming the NHS, all to be delivered at pace and with limited resources. But traditional policymaking, which relies on upfront planning and waterfall delivery is not well-suited to deliver on these challenging missions. For example, the Opportunity Mission target is a 7.3 percentage point increase in the number of school-ready 5-year-olds in England over 3 years. Yet in the most recent year, we achieved an increase of only 0.5 percentage points.

This doesn't mean conventional approaches have no place – they still work well where the path from policy to implementation is clear. But to tackle complex problems under pressure, we need a new approach – one that embraces uncertainty, continuous adaptation, and a relentless focus on outcomes. We need to test and learn. This isn't a nice-to-have – it's essential to delivering impact and making every pound count.

Test and learn is not a new idea. Its principles draw from a wide range of disciplines and methodologies, from scientific experimentation, design thinking, behavioural science, continuous improvement, the Agile Manifesto and more. Even in the 19th century, Prussian general Helmuth von Moltke the Elder recognised that "no plan survives first contact with the enemy."

In the UK government, test and learn already has a foothold. The Government Digital Service pioneered agile development in the early 2010s. Later, Universal Credit was rescued from failure by shifting from a waterfall approach to a test and

learn model – re-focusing from outputs to outcomes, bringing together previously separate policy, technology and delivery teams, and adapting service design in real-time. The programme now supports around five million households. Defra's Future Farming and Countryside programme offers another example. By designing policy iteratively and incorporating direct feedback from farmers, it replaced the EU's Basic Payments Scheme with a service that drew praise in newspaper industry headlines.

Yet despite these successes, test and learn remains the exception, not the norm. In 2019, only 8% of government spend on major projects (£35 billion of £432 billion total expenditure) had robust evaluation plans in place and 10 out of 16 chief analysts noted that the opportunity to learn was not adequately built into policy design and delivery. Too often government programmes still skip straight from policy to large-scale delivery and evaluation. The failure of the Green Deal – where flawed assumptions weren't tested until three years costing the taxpayers £240 million –shows the cost of this approach.

Test and learn exists on a spectrum. Some government departments pilot initiatives or hold regular check-ins, allowing for minor course correction. Most use rigorous impact evaluation to assess the effectiveness of some of their interventions. But few embed test and learn throughout – aligning teams around outcomes, embracing evidence-informed adaptation, and shifting mindsets and organisational culture. That end-to-end commitment is what is needed to deliver real change at scale.

With ambitious mission targets to achieve, public services under strain and budgets tight, it's more important than ever to innovate quickly and affordably. Test and learn provides a new way to drive change at scale. It helps government improve public services by reducing risks, tackling complex challenges, and speeding up the delivery of effective solutions.

What is test and learn?

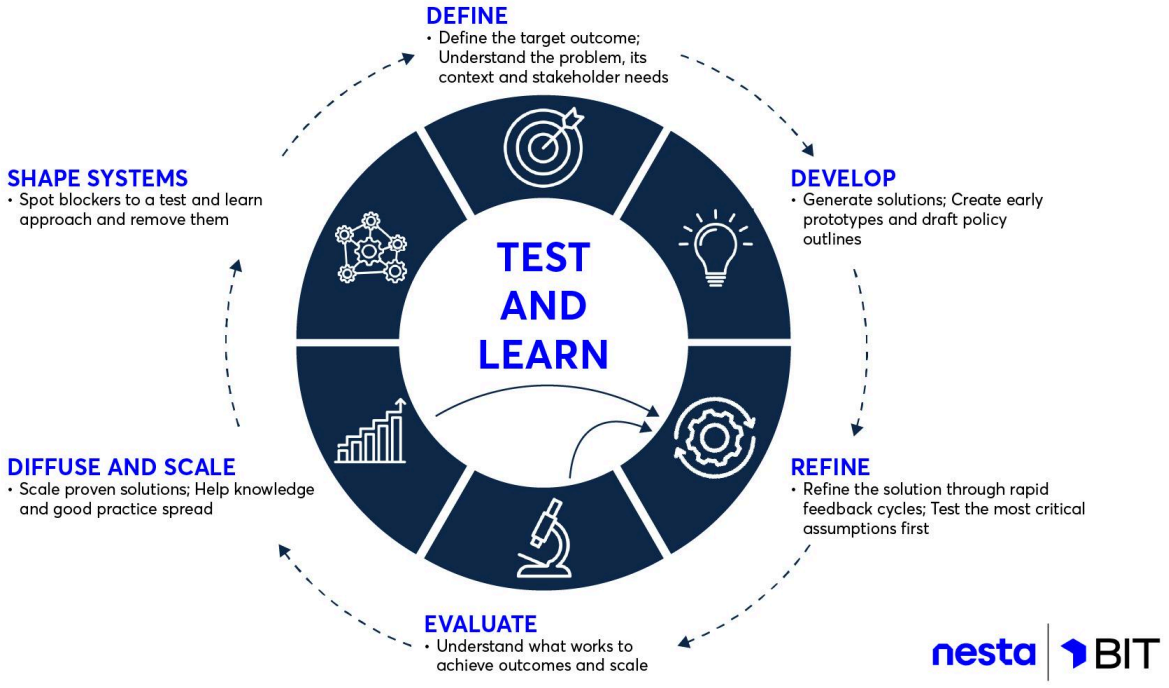
Test and learn is an iterative approach to public policy and service development that starts from the ground up – policy is shaped by practice, not the other way around. It closely inter-twines policy design, delivery and testing in repeated iterative cycles, drawing on the best of design and digital practices, applied research, evaluation and more. This enables policies, programmes and services to adapt, pivot and effectively scale in response to early and ongoing data and evidence.

Importantly, test and learn is not simply a series of small, rapid design experiments that can replace rigorous evaluation. Done well, test and learn combines methods for agile policy design with robust impact evaluation to deliver on long-term outcomes.

An ambitious test and learn programme would build **an R&D ecosystem for the UK government**, applying the approach to different challenges, such as new policy development, existing service optimisation and public programme design and scale-up.

A framework for test and learn is described in the figure below.

Figure 1: Test and Learn framework



Fundamentally, test and learn is a flexible framework – not a prescriptive set of steps. It's grounded in a systematic process and a set of core shifts in how we design, deliver, and adapt policy. Drawing on methods from multiple disciplines, it builds on HM Treasury's commitment to evidence-based policymaking.

Below we outline four of the core shifts:

1. **Put practice before policy** – a new place to start.
2. **Work in learning loops** – a new way to make progress.
3. **Test assumptions first** – a new way to handle risk.
4. **Shape conditions for success** – a new way to manage and lead.

1. Put practice before policy

Typically polices are made in Whitehall and then delivered across the country. However, what seemed like a sensible policy in a London boardroom may fail on the ground, where real-life challenges look very different.

Some of the most impactful national policies were built on a number of small, local experiments. The British welfare state, founded in response to the Beveridge report after World War Two, was in part inspired by earlier smaller-scale experimentation by voluntary social insurance schemes.

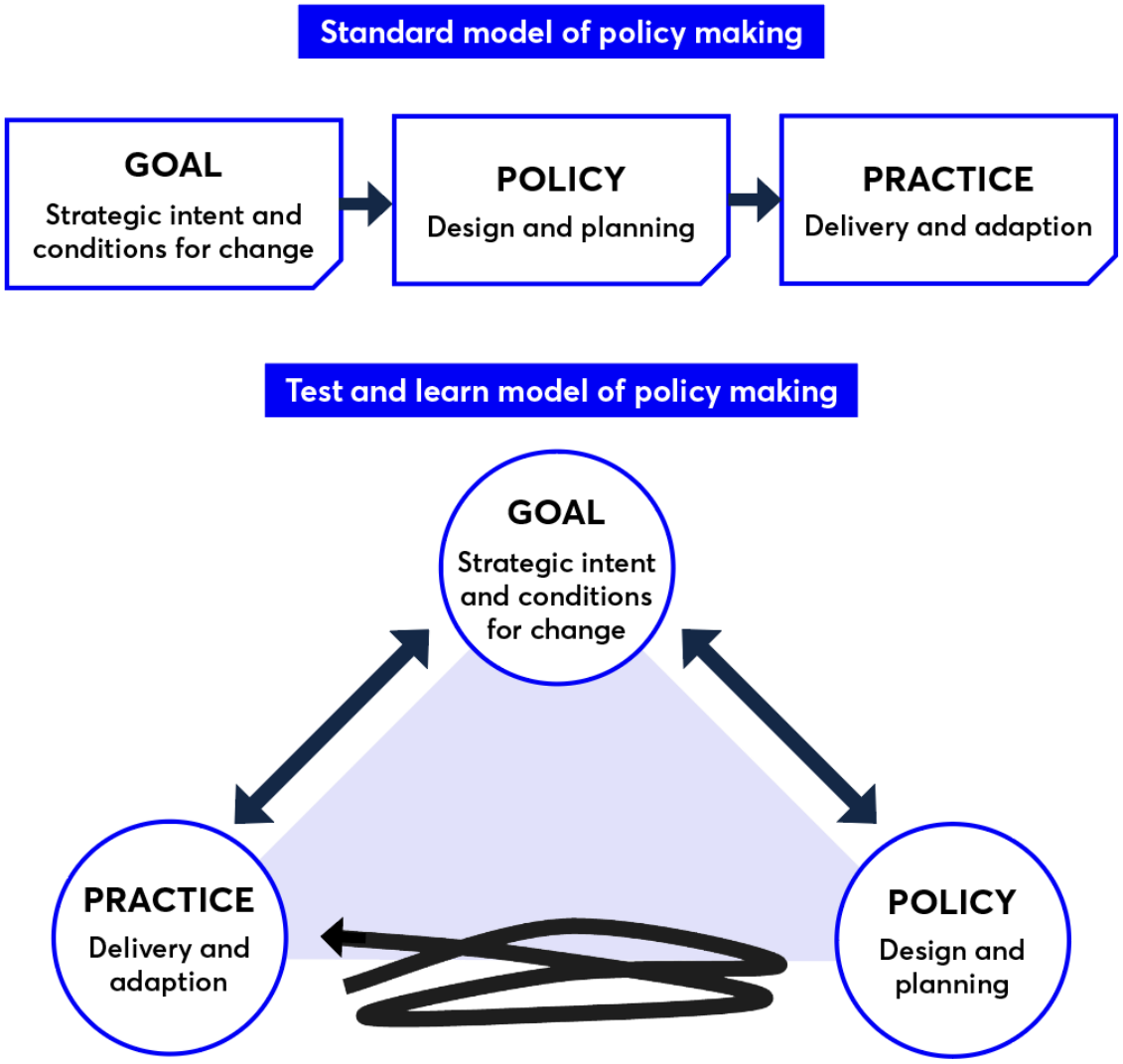
As our Chief Behavioural Scientist Michael Halsworth noted in his work on System Stewardship, policy and delivery are deeply intertwined. Policies evolve as they're implemented and frontline staff often re-make them in practice. This isn't policy going wrong – it's those closest to the work responding creatively and pragmatically to realities on the ground. Test and learn makes this adaptation intentional, not accidental.

This is a fundamental flip. **Policy starts with practice.** First comes existing evidence and examples of what already works, including learning from cases of **positive deviance** – where individuals or organisations achieve exceptional results despite facing the same constraints as others. Then comes grounded action – small-scale experiments, frontline insight, and real-world problem-solving. From there, we build outward. Policy emerges from what works, **informed by prior evidence and direct feedback from those closest to the issues.**

This is not a one-off step, but an ongoing, iterative process. Practice informs policy, policy shapes the next round of practice, and the cycle continues – refining, adjusting, improving.

Importantly, this shift doesn't eliminate strategy or central government's role – it reframes them. Strategic intent shapes the conditions for experimentation: prioritising the problems we want to solve, setting broad goals, and creating the system conditions for change. Instead of assuming control, the centre becomes a steward of learning loops, helping join up what's being tried in different places and making sense of what's emerging. Policy becomes a living system of feedback, rather than a single shot from a starting gun.

Figure 2. Models of policy making



2. Work in learning loops

The test and learn framework doesn't completely re-invent policy development. It builds on standard frameworks such as the ROAMEF (Rationale, Objectives, Appraisal, Monitoring, Evaluation, Feedback) outlined in the HM Treasury Green Book. However, instead of following a **linear** "waterfall" approach where evaluation and feedback happen at the end, policy and services are developed through **shorter, iterative learning loops** that provide rapid feedback. To manage risk we want to 'close' these learning loops as quickly, simply and safely/ethically as possible. At its simplest, closing a learning loop involves: 1) developing something; 2) making contact with reality; 3) learning from the results.

This is not a simple addition of a few more stage gates to an otherwise unchanged policy development paradigm. In contrast, test and learn emphasises **continuous data gathering, experimentation, and user feedback at every stage of the policy lifecycle**, ensuring policies adapt based on real-world evidence. As policies and services evolve from early prototypes to broader rollouts, the nature of the data and evidence informing decisions also changes. In the initial phases, data is used to validate direction and offer early signals of what might work. As initiatives scale, more rigorous methods for assessing impact become essential. This helps government **reduce risk incrementally**, refining policies based on live insights rather than delivering based on untested assumptions.

3. Test assumptions first

Policy design is based on assumptions about how things influence each other such as how a service affects people's behavior or how national and local government interact. If these assumptions are wrong, policies will not work as intended. In test and learn we start with the end outcomes in mind and then work back to identify our key assumptions and uncertainties. We then (through the learning loops discussed above) test the most critical assumptions first, at a small scale. This way, we take small safe-to-fail steps to build our confidence in our solution. In other words, we think big and start small, growing our efforts as confidence and results build.

The major assumptions we make when developing public services and policy often fall into a few areas:

1. **Stakeholder value** – how are people's needs, behaviours and context met?

2. **Impact** – what is the likely size of the benefit?
3. **Viability** – what is economically viable and sustainable?
4. **Feasibility** – what is deliverable and technically possible?
5. **Political acceptability** – what will the public and media accept?
6. **Scalability** – what is the potential for impact at scale?
7. **Unintended consequences** – how might we cause and avoid harm?

Different methods and different forms of evidence are going to be needed to explore each of these types of assumptions. For example, randomised controlled trials might be useful for measuring impact while user interviews might be more suitable for exploring stakeholder value.

4. Shape conditions for success

With the flexibility and adaptability of test and learn come some tradeoffs. In particular, we need to give up direct control. Instead, our role is to shape the conditions for success. This doesn't mean renouncing responsibility – rather, control is **indirect but intentional**. It's a little like being a football coach, designing the training sessions and setting the strategy, but letting the players read the play and make decisions on the field.

Having less direct control can feel scary. It also requires specific skills and experience to manage work in this way – rather than a legible plan we have a messy collection of learning loops all happening at once requiring diverse and different people to collaborate.

At the heart of making this work is **trust** – trust in the process, in the people involved, and in the ability to course-correct as new insights emerge. Leaders and managers need to empower those closest to the work to have the agency to experiment, make decisions, and drive meaningful change, and those delivering need to have the skills and experience to work in an **agile and data-driven way**.

Rather than relying on rigid top-down control, this approach thrives on c**ollaboration and co-creation**, bringing together a range of perspectives to shape and refine solutions in real-time. While test and learn doesn't require a **place-based approach**,

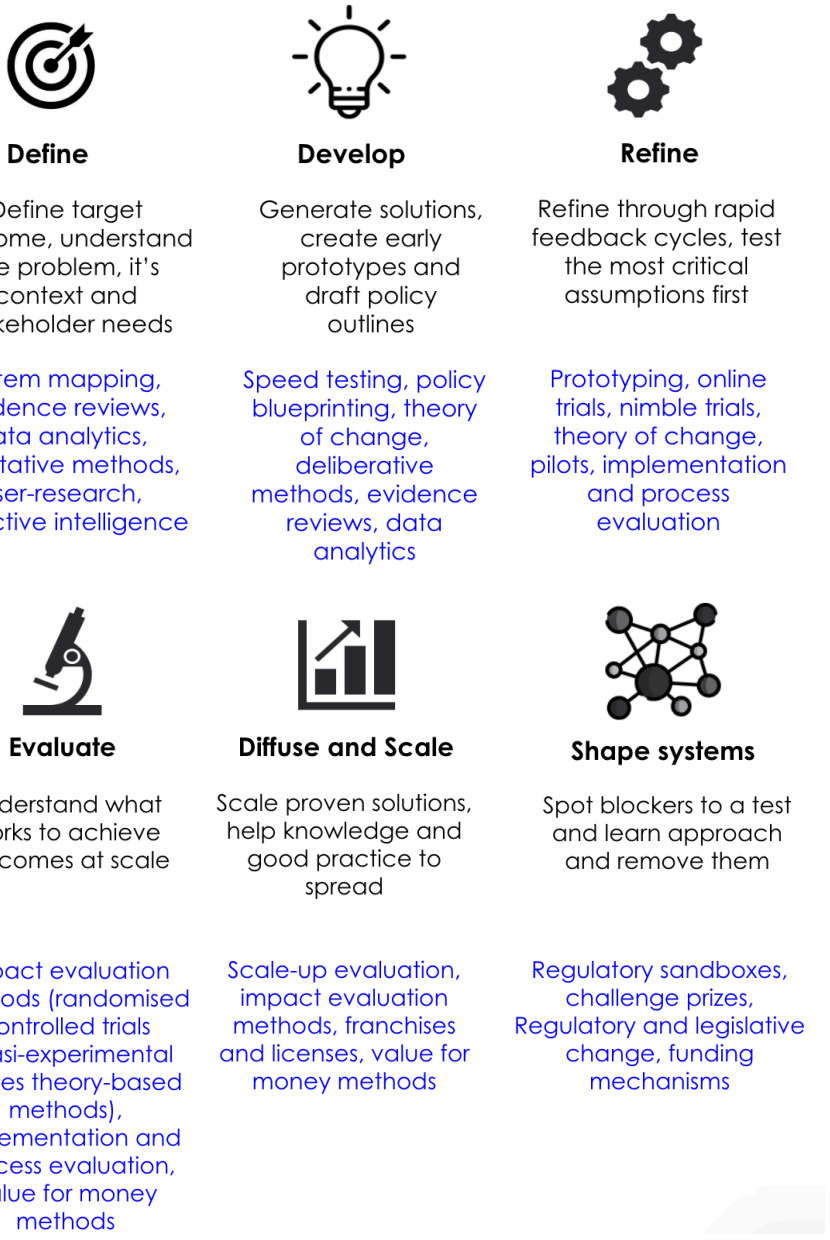
applying it in specific local contexts can be particularly valuable, allowing for deeper engagement with communities and more tailored, context-aware solutions. By embedding these principles, we not only make test and learn effective but also create policies and services that are more responsive, resilient, and deeply connected to the people they serve.

Occasionally people need long-term certainty. For example, businesses may want the government to provide a consistent industrial strategy. In some instances, providing certainty about direction is enough – we hear this a lot from the energy industry on issues like the future of heating. In other instances, stakeholders need certainty on delivery details. For example, companies delivering government energy efficiency schemes need details on the duration of funding and the eligibility criteria. The need for certainly is obviously an important consideration. However, even in these instances, there may be opportunities to leverage some elements of test and learn, such as local experimentation or running phased rollouts to refine policies before full-scale implementation.

The test and learn methods toolkit

Test and learn draws on a toolkit of methods and techniques from different disciplines. These are combined flexibly to address different policy challenges and inform different stages of the policy, programme or service life-cycle. Figure 3 below shows which methods might be helpful at which stages of the Test and Learn Framework.

Figure 3: Test and learn methods toolkit



The toolkit builds on evaluation methods outlined in the HM Treasury's Magenta Book, which provides detailed guidance on assessing impact – an essential part of test and learn. However, test and learn extends beyond impact evaluation, offering a broader set of methods and techniques that support policy development, refinement, and scaling.

The toolkit is also continually evolving and the test and learn mindset should embrace the challenge of constantly improving our own ways of working. A great example of this is the fact that almost all of our methods can be augmented or accelerated by harnessing AI tools – from using AI agents to conduct large-scale qualitative research to AI-led evidence discovery and analysis, and AI-enhanced ideation and solution design. AI is increasingly a core component of the test and learn toolkit at each stage, and we should expect to iterate and improve on the value that we derive from it.

As so many methods can be used for test and learn, the toolkit will inevitably not be comprehensive. We outline some of the most common and most important methods below.

Defining the challenge

Prior to designing a policy or programme, it is important to develop a deep understanding of the context. This will involve evidence, data and **engagement with local and/or frontline partners**, which is critical to ensure alignment on goals, a shared understanding of the problem and context, and buy-in from all involved. A number of methods can be used at this stage to lay the groundwork for solutions.

Qualitative methods and user research

Methods such as interviews, focus groups, observations or ethnography can provide an initial diagnosis, offering a deeper understanding of the challenge, the underlying barriers, and opportunities for intervention.

Rapid evidence reviews

These help synthesise existing evidence, balancing rigour with speed to provide timely, reliable insights for decision-making. They can be used to quickly understand the context, build an up-to-date evidence base on potential solutions and inform programme design and delivery.

Data analytics

Data-driven approaches harness existing administrative and large-scale datasets. They offer insights into trends that can inform programme design and enable monitoring of "leading indicators" of successes and challenges with delivery.

System mapping

This visualises complex interconnections within systems – identifying actors, behaviours, structural influences, and feedback loops that shape outcomes. Co-creating the map enhances understanding of the challenges and opportunities to intervene to achieve scalable and sustainable change.

Collective intelligence methods

These are outlined in Nesta's Collective Intelligence Design Playbook and include methods such as **crowdsourcing, deliberation, and data collaboratives**. They bring together diverse groups of people, data and technology to help surface diverse perspectives, uncover blind spots, and build a more comprehensive understanding of the challenge. They can help ensure the challenge is framed accurately and that potential solutions address the root causes rather than just the symptoms.

Case study: data analytics

York ward profile data dashboard

Nesta worked with the City of York Council to explore how linking administrative datasets could help decision makers define the challenge so as to better target resources. The team built a data dashboard that linked and

visualised ten different data sources covering education, health and other local authority data. They analysed the linked dataset through a number of different lenses and created novel data visualisations that York can use to easily interpret their data. For the first time, York could see the journey of a child through the whole early years system, how the demographics of children impact their outcomes and how these differ at ward level. The dashboard allowed York's team to make better-informed decisions about where to target enhanced support for their 0-5 population. York used their Raise York family hubs and their award-winning Early Talk for York programme as part of this network of support.

Case study: behavioural system mapping

Tackling discharge delays at Homerton hospital

BIT partnered with the Homerton University Hospital NHS Foundation Trust to tackle patient discharge delays at Homerton Hospital. The complexity of the discharging process meant that a single behavioural intervention would be unlikely to make much impact. Recognising that the discharging process is a collaboration between multiple interconnected actors and actions that are embedded within the organisational structure and culture of the NHS, the team took a systems approach to define the challenge and identify opportunities for intervention. They co-produced a behavioural systems map with the key partners involved in the discharge process. This helped to identify leverage points where small changes such as pharmacy cut-off times or changes to induction training were likely to have a broader impact across the system. The proposed solutions are now being implemented as part of a hospital-wide action plan.

Developing solutions

In the early idea and development stages of a programme, adopting agile, iterative approaches can help spot, refine, and validate new ideas and improvements before committing to large-scale implementation and evaluation. These methods look for **"leading indicators"**[1], measurable metrics that signal potential success or

challenges and indicate the likely impact of a programme. **Engaging stakeholders and incorporating co-design** ensures that solutions are grounded in real-world needs and perspectives, leading to more effective and sustainable outcomes.

Theory of change

A theory of change can be used to articulate (often implicit) causal links through which a proposed policy or programme is expected to achieve its intended outcomes. It helps identify critical assumptions that need to be tested early to refine a solution. Best developed through multidisciplinary workshops involving a broad range of stakeholders, they should be living documents that are regularly reviewed and refined as more is learnt about delivery and assumptions.

Policy Blueprinting

Nesta-developed Policy Blueprinting is a tool and a collaborative process for the design of a policy proposal or portfolio of interventions that address large-scale policy challenges or systemic issues. It brings together stakeholders across the policy lifecycle to integrate policy and delivery considerations in a collaborative design process, using visual mapping to simplify complexity and highlight key interactions and dependencies. (See case study on Coordinated transition to low-carbon heat)

Speed testing

This is an approach for rapidly testing new ideas with real-world feedback. They can be used rapidly and cheaply to spot, refine or discount promising ideas, particularly in contexts of high uncertainty. Interdisciplinary teams come together around a highly focused challenge and rapidly test and validate the most critical assumptions based on real-world feedback. This can involve a range of activities, from stakeholder and user interviews to running elements of a service or building a minimum viable product. Even a small-scale trial among colleagues can provide valuable insights to refine and improve the solution before wider implementation.

Deliberative methods

These include citizen assemblies, deliberative fora and deliberative polling, and can further enrich the solution development process by bringing diverse voices into decision-making and strengthening public trust in policy development.

Case study: speed testing

Increasing heat pump adoption at Nesta

A multi-disciplinary team at Nesta speed-tested a range of promising ideas for increasing heat pump adoption in order to quickly understand their viability. One of these explored the idea of providing temporary 'interim boilers' to households facing heating system failures. This approach aims to offer immediate relief by installing a temporary gas boiler, allowing homeowners the necessary time to transition to low-carbon alternatives like heat pumps without the pressure of immediate replacement. Our investigation included developing a business model, consulting with heating engineers to understand costs and barriers, and engaging with consumers to gauge interest. Findings indicated that while the service could facilitate a smoother shift to sustainable heating solutions, there were challenges such as high rental costs, regulatory challenges and supply chain limitations. In the end, speed-testing helped us quickly rule out interim boilers as a solution for enabling heat pump adoption - and pivot to other potential solutions early.

Refining and iterating solutions

Once initial promising solutions are identified, test and learn methods can be used to help refine these in repeated iterative cycles. This enables policies, programmes and services to adapt, pivot and effectively scale in response to early and ongoing data and evidence. Using these methods allows for quicker innovation cycles, reduces risks associated with major policy shifts, and ensures that full-scale evaluations are conducted on more mature, field-tested policy concepts.

Iterative prototyping

This can be used to develop, test and improve an early (minimum viable) version of a policy, programme or service before committing a lot of resources to it or making large-scale changes. Iterative prototyping is best for the early-stage development of policies, services, or interventions. They are useful for exploring new ideas, testing different approaches quickly, and refining these based on user feedback.

Online experiments

Experiments like these can generate quick empirical evidence to inform (changes to) policy or programme design by testing in tailored test environments using visual, audio, or video stimuli or simulated environments. They are particularly useful for testing the assumptions in policy design, building evidence on whether assumptions made in the programme theory of change are likely to be valid in a given context. They are also useful for informing programme communication and engagement strategies, and to inform digital service design, including optimizing user experiences on government platforms.

Nimble trials

These are quicker, cheaper randomised controlled trials that can be conducted within weeks or months. They can optimise a programme by testing critical elements of design such as recruitment and engagement strategies, using metrics that are routinely collected and focusing on easily measurable proximate outcomes. As such, they are particularly well-suited for policy or service optimisation where controlled experimentation can provide robust evidence of impact.

Case study: iterative prototyping

Visit a Heat Pump

Nesta's Visit a Heat Pump service demonstrates an iterative approach to prototyping and scaling solutions that promote clean heating adoption. We started with the idea of a service that would enable people to see a heat pump in a real home. We mocked up some basic prototypes and showed them to a few homeowners to test whether they'd be interested in services like this, and a few heat pump owners to see what their response would be to having someone visit their home. This helped us choose between a few potential options (visiting homes or visiting showrooms) and flagged issues that would be important to consider when designing a service. We found the concept had legs so we launched a service with a minimal set of features. The digital platform connected people interested in heat pumps with hosts who already have them, enabling visitors to book viewings and ask questions. The team has been refining and expanding the service in a user-centred way ever since, with more than 200 iterations, eg, adding calls to action in response to

feedback from both visitors and hosts. Since its national rollout in April 2024, over 400 heat pump hosts have joined, and there have been over 1,000 visits, reflecting the success of this phased, user-centred development approach.

Case study: online experiments

Testing how AI tools can improve public services

Using their in-house experimental platform Predictiv, BIT conducted an online experiment with 5000 adults to assess public engagement with AI-powered chatbots on government websites. Participants were tasked with scenarios such as resolving rent disputes and identifying child health issues. Participants were randomised so that some had access to different versions of an AI chatbot, whilst others did not. The results were recorded and analysed within just a few days. Those with access to the chatbot reported finding the task easier. They also showed higher support when asked if AI should be used to help citizens, indicating that exposure to AI increases acceptance. However, less than half of the people engaged with the chatbots, and their presence did not significantly enhance task accuracy or speed, suggesting that while AI chatbots are generally welcomed, their practical effectiveness depends on thoughtful implementation and complexity of the tasks they address.

Case study: nimble trials

Improving the National Tutoring Programme

In partnership with the Education Endowment Foundation, BIT and NFER conducted nimble trials within the first year of the National Tutoring Programme to quickly test and refine elements of the programme. BIT developed and tested three different strategies to improve school uptake of the programme and pupil participation at NTP tutoring sessions. These included different approaches to building tutor-pupil rapport such as a brief survey that helped identify shared interests and preferences, relationship-building activities and behaviourally informed reminder messages. The quick survey increased attendance at tutoring sessions by 4.2 percentage points. This increase would have translated to 1,600 more tutoring sessions attended, had the changes been rolled out to the control group.

Evaluating and assessing impact

Once policies and programmes are optimised using iterative cycles of development, robust impact evaluation methods provide the "missing" signal about whether policies and programmes address unmet need and add value over and above standard practice. They assess the contribution that policies make to long-term goals, uncover any unexpected outcomes, and help guide improvements prior to scale-up. The HM Treasury's Magenta Book provides a comprehensive and authoritative source on these methods. The section below briefly outlines some of the key methods.

Randomised controlled trials and quasi-experimental designs

These evaluate how a programme performs compared to an alternative, such as a different intervention or standard practice. They are particularly useful for programmes that can be delivered and scaled up in a relatively standardised way across a large number of settings and/or individuals. An accompanying set of evaluation methods helps explore not only whether programmes have an impact, but also how, why and for whom, and whether they are cost-effective. Since BIT (then part of the Cabinet Office) published the Test, Learn and Adapt report in 2013 which set out the case for RCTs in public policy, hundreds of interventions have been evaluated by UK government departments, the network of What Works Centres, and organisations like BIT and Nesta, with some great examples in government collated by the UK Evaluation Task Force.

Theory-based evaluation methods

For system-level policy changes or complex, multi-component or place-based programmes, theory-based evaluation methods can be used to refine programme design, assess programme contribution to outcomes and inform scale-up.

Case study: theory-based evaluation

Reducing rough sleeping through time-limited accommodation

BIT partnered with the Centre for Homelessness Impact to test a programme offering time-limited accommodation and immigration advice to non-UK nationals in order to resolve the immigration status and reduce rough sleeping. The programme has multiple interacting components and delivery is tailored within each local authority. Given the complex nature of the programme and limitations in rough-sleeping data, the team is using contribution tracing – a theory-based evaluation approach – to evaluate the impact of the programme. The team worked closely with partners to develop a detailed theory of change that outlines how the programme is expected to reduce rough sleeping. They are now collecting data to assess whether the programme was implemented as intended, examine the key causal mechanisms at work, and explore the influence of other factors on rough sleeping. Through iterative evidence collection and analysis and systematic assessment of the weight of the evidence, the team will build a robust understanding of the programme's contribution to rough-sleeping reduction (see the protocol for this study).

Diffusing and scaling

Test and learn also plays a critical role in diffusing and scaling proven programmes. Further iterative experimentation can unpack how programmes need to adapt in response to variation in contexts and implementation models, followed by further robust evaluation where possible to understand impact at scale.

Scale-up evaluations

These identify specific insights to support a programme to scale effectively as well as general lessons about what it takes to scale similar programmes. Scale-up evaluations use mixed methods to understand what structures, resources, contexts and enabling factors allow programmes to be scaled up.

Franchises and licenses

One way of scaling programmes is through building a delivery network. Rather than doing all the heavy lifting themselves governments can under the right circumstances do this by granting franchises and licenses to civil society or industry. Franchises and licenses can incorporate test and learn by granting them to many different organisations and seeing which works out best.

Case study: scale-up evaluation

Scaling a professional development programme for teachers

BIT partnered with the Education Endowment Foundation and the Schools, Students, and Teachers Network to support and evaluate the scale-up of a 2-year intensive professional development programme for teachers. While there was strong evidence (from a randomised controlled trial) that the programme improved pupil attainment, implementing it well at scale was not a given. Only a small number of successful educational programmes had been scaled up, with the process of scaling being formally evaluated. The team evaluated the scale-up over four years using a mixed-methods approach that involved school case studies, surveys, interviews, and data reviews. They assessed scaling strategies, reach, fidelity, costs and sustainability. Insights from the evaluation informed and supported the scale-up efforts and helped build an understanding of what it takes to effectively scale up educational interventions in the English state school system. The team found a number of adaptations that made scaling easier – these included automating processes, expanding mentoring support, and refining the programme marketing strategy. A school subsidy also played a critical role in facilitating uptake.

Shaping systems

For test and learn to have maximum impact, it needs to be deployed in a conducive environment. This requires understanding and removing blockers, and incentivising innovation. Systems need to be shaped to give test and learn the best possible chance of success. Methods to help might include:

Challenge prizes

Competitions that reward whoever solves a problem first or most effectively. Pre-defined criteria describe success, without indicating how exactly it should be achieved. Prizes reach beyond the usual suspects to engage innovators that other sorts of funding might miss. They can be used to help shape systems by providing an incentive for new players to tackle problems in different ways or by spotlighting a particular problem or solution as a way of aligning innovators around that goal.

Regulatory sandboxes

These provide regulators and innovators with a managed and often temporary opportunity to test ideas under live conditions, usually with real customers. They allow managed waivers to regulations that allow both companies and regulators to learn about new solutions and technologies. Regulatory sandboxes provide a space in which changes to systems can be tried through a process of testing and learning.

Used in combination, these methods provide a powerful toolkit for government and its missions to test and learn their way to improved outcomes.

Case study: challenge prize

The Longitude Prize on Antimicrobial Resistance

In 2014, Nesta's Enterprise Challenge Works launched the £8 million Longitude Prize for anyone who could develop a rapid, affordable, and simple test to determine whether an infection requires antibiotics. The goal was to reduce the number of tests patients needed and prevent doctors from prescribing antibiotics as a precaution. This would help tackle antibiotic resistance which is a major threat to global health. The winner, Sysmex Astrego, announced in 2024, developed a test which can provide results in 45 minutes. The prize shaped the system through test and learn by providing an incentive for innovators to try new ways of tackling the problem. Hundreds of teams competed which brought many new solutions closer to market.

How test and learn can address key challenges for government

While test and learn can be applied in lots of ways there are common patterns for how they can be used to help address key challenges the government faces. Below we briefly outline three of these:

1. **Valuable service challenge** – we need to improve existing services (sometimes quite drastically) whilst keeping them running, often with limited capacity and capabilities to do so.
2. **What works challenge** – we need to understand which services, programmes and policies (or elements of them) work so we know which options to pursue and what to spread into new contexts.
3. **Discovery challenge** – we need to develop and build confidence in ambitious new policies and services to address issues where it's not possible to understand the whole problem or predict the effect of interventions.

1. Valuable service challenge

Challenge: We need to improve existing services (sometimes quite drastically) whilst keeping them running, often with limited capacity and capabilities to do so.

What to look for: Existing services and policies that need fixing because they do not meet the needs of service users or providers. While big changes may be needed the main goals of the policy or service remain the same.

What it looks like:

	Stage	Methods to consider
1	**Discovery and prioritisation** Identifying areas for improvement on existing services and making rapid progress on	Where possible, bring together and analyse **data** to gain granular insights into the challenges and help target service improvements. Techniques like user journeys and **system mapping** can help map out more complex services and identify the

	addressing them, while building team confidence and capabilities.	most promising points of intervention. Use a structured process like **speed testing.** This can help interdisciplinary teams test critical assumptions and make improvements through rapid learning loops based on experience from the real world. **Rapid evidence** reviews can be integrated into the process to ensure development is informed by knowledge of what has worked from elsewhere.
2	**Iteratively deliver value** Changing (sometimes quite drastically) existing services so that they meet people's needs, have the desired results, and work within operating constraints.	Use **iterative prototyping** to develop a mock-up before committing a lot of resources. This should happen alongside live testing and development using techniques like **online experiments** and **nimble trials**. As service improvements are scaled, evaluation techniques like **randomised control trials** and **quasi-experimental methods** can be used to assess overall impact.
3	**Continuously improve** Building habits that: 1) grow and optimise existing services; and 2) discover new ways to deliver value.	Once services and policies are consistently meeting the core needs of citizens the focus should move to continuous improvement and adaptation. Building team confidence with product/delivery techniques can be helpful including the capability to analyse and visualise **data** to monitor performance and support continuous experimentation. If they haven't been done already, evaluation methods such as **randomised controlled trials** and **quasi-experimental methods** can be undertaken to improve understanding of impact and identify opportunities for improvement. Finally, returning to approaches like **rapid prototyping** and **speed testing** can be helpful to discover new ways to deliver value and ensure services can respond to changing needs.

Case study: improving midwifery and health visiting services in Stockport

What happened?

In Stockport, Nesta partnered with the Local Authority and used test and learn to improve antenatal care by better-linking midwifery and health visiting.

How was test and learn applied?

The team gathered insights from parents and practitioners to spot key problems such as poor communication, uncoordinated support, and missed early help. They co-designed and tested a new joint visiting model, then through **iterative prototyping** refined visit structure and information sharing.

Strict rules and key performance indicators meant it was not possible to increase the number of visits, so instead the team improved digital record-keeping, created shared training, and provided consistent family assessments to prioritise joint visiting for families who would benefit most.

What was the impact?

A mixed-methods evaluation used interviews and analysed data to measure progress, gathering parent and practitioner feedback. It showed how iterative prototyping improved services in line with the priorities of the service providers, streamlined processes, created buy-in and improved support for families.

What lessons were learned?

- Test and learn can show how to improve even well-functioning services.
- More effort is required to persuade people to try new approaches like test and learn with existing policies and services that they are mandated to deliver.
- Strong cross-team ties and aligned goals are key to collaboration.
- Ongoing parent and practitioner feedback helped balance professional and family needs.

2. What works challenge

Challenge: We need to understand which services and programmes (or elements of them) work so we know which options to pursue and what to spread into new contexts.

What to look for:

- A need to understand whether and how services or programmes work to inform decisions on what to fund and whether and how to scale.
- Similar problems that have already been solved elsewhere from which we can learn lessons and make predictions.
- Cases where more proof is needed to take action.

What it looks like:

	Stage	Methods to consider
1	**Diagnose and Design** Identifying and designing evidence-based solutions to address your challenge.	Start by defining your challenge and the outcome(s) you want to achieve. Conduct **qualitative research** and **consult stakeholders** to understand the context and the key barriers. **System mapping** can be used to understand how different actors, behaviours, structural influences, and feedback loops shape the outcome and to identify opportunities for intervention. A structured process like **speed testing** can help generate potential solutions that could be tested. A **rapid evidence review** can be used to help understand previous research and identify opportunities to adapt previously evaluated solutions from elsewhere to your context. Where data is available, **data analysis** can help identify trends to inform intervention design.
2	**Refine and evaluate** Using rapid feedback loops to refine solution design and improve intervention	Once you have a proposed intervention design, a **theory of change** can be used to identify critical assumptions.

	promise prior to robustly evaluating whether it works.	**Prototyping** or piloting of the intervention at a small scale can help test these critical assumptions early and adapt the intervention before a large-scale evaluation. Rapid evaluation methods like **nimble trials** can be combined to optimise interventions before they are robustly evaluated through a **randomised control trial** or **quasi-experimental study.**
3	**Replicate and evolve** Sharing and spreading proven solutions, supporting adaptation and scaling, and building the capacity and conditions for others to adopt them and contribute to learning.	Proven interventions are further adapted and refined, informed by insights from **scale-up evaluations** as they are adopted in new contexts. The results and good practice can then be shared, and the service or programme re-tested at scale using **randomised controlled trials** or **quasi-experimental methods** where possible. Capacity building will be needed to help others adopt interventions that are developed in this manner.

Case study: helping young people at risk of not being in education, employment or training

What happened?

BIT has partnered with the Youth Futures Foundation (YFF) and Football Beyond Borders (FBB) to help secondary school pupils at risk of not being in education, employment or training (NEET). Using a test and learn approach BIT is developing and evaluating a scalable mentoring programme "Building Futures" that could be delivered by many different organisations.

How was "test and learn" applied?

First the team worked with partners to map how the programme might create impact using a **theory of change**. This gave a clear picture of the assumptions being made, and where changes might improve outcomes.

Subsequently, the team is using **iterative prototyping** to refine and adapt different elements of the intervention with a focus on leading indicators of success – measures that speak to whether the programme is appealing, scalable and addresses young people's needs.

In parallel the team are running a **feasibility study** and a **pilot evaluation** to ensure that the final model can be properly assessed using a **randomised controlled trial**. Regular reflection points help us adjust both the service design and the most suitable way to evaluate.

What was the impact?

While the impact of this programme is still ongoing, it has already transformed ways of working by embedding a test and learn approach from the outset. The team worked with YFF and FBB to identify key uncertainties during the project kick off and established a structured, iterative way of working – tailoring research activities, maintaining a dynamic backlog of test and learn initiatives, and adapting based on emerging evidence. This has fostered a culture of agility, collaboration, and data-driven decision-making for programme impact and scaling.

What lessons can be learnt?

- **Leading indicators help signal when the project is off-track** – Instead of waiting until the end of a pilot, early signals – such as low sign-ups rates – are actively monitored to inform timely pivots. This is especially critical in a project with overlapping delivery waves, where decisions for future delivery waves must be made early in the current cycle, requiring agile adjustments based on emerging data.
- **A detailed theory of change is an essential roadmap** – it enables the team to anticipate how intervention changes might affect outcomes and how unintended impacts might be mitigated. It also underpins a structured, iterative evaluation approach, accelerating learning within tight time constraints.
- **Flexibility and quick adaptation are key** – As intervention models evolved significantly between the first two delivery waves, the team shifted the approach to learning and optimisation. By working with FBB to integrate learning activities within delivery, they built a flexible system that enables the team to adapt the programme as new insights emerge.

3. Discovery challenge

Challenge: We need to develop and build confidence in ambitious new policies and services to address issues where it's not possible to understand the whole problem or predict the effect of interventions.

What to look for: Big complex, uncertain challenges like climate change or inequality often require new services, policies and ways to address them. These issues usually cut across many areas of government and require coordination across services, organisations and sectors. The focus is less on solving a single problem and more on shifting an entire system to a preferred state.

What it looks like:

	Stage	Methods to consider
1	**Define** Understanding the system in which the service or policy sits to allow a preferred new state to be co-developed. This will involve surfacing critical assumptions to build confidence.	Big, complex, ambitious challenges often need a whole system approach. **Policy blueprinting,** a collaborative process for designing interventions to address large-scale policy challenges and systemic issues, can be a helpful way to start. For more open-ended situations other **system mapping** techniques can be used to understand the situation and where to intervene. These can be complemented by more traditional policy analysis and consultation, as well as strategic foresight techniques.
2	**Demonstrate** Iteratively developing, testing and delivering a cross-cutting new policy or service system either directly or by facilitating those closest to the challenge to contribute to shared outcomes.	After understanding the system and spotting critical uncertainties or opportunities to intervene, **iterative prototyping** can be used to deliver demonstrators that build confidence in critical elements of the new system and make progress towards shared outcomes. Where solutions already exist, good practice can be widely spread and used to update strategy. **Theory-based evaluation methods** can be used to help assess and observe if the desired changes are occurring.
3	**Embed and steward** Building capacity across	As confidence in the policy builds through prototyping and demonstrators, attention should shift to capacity

	the entire policy or service system to work together to deliver shared public outcomes, learn from each other and continually improve delivery.	building and stewarding further scale-up. **Data analytics** and **theory-based evaluation** methods for complex systems are used to help monitor the state of the system and facilitate continual adaptation.

Case study: coordinating the transition to low-carbon heat

What happened?

Nesta is testing a bold approach to planning and delivery of low-carbon heating by shifting the entire system.

How was test and learn applied?

The team began with **systems mapping** to help define the problem. They then applied **policy blueprinting** to develop an ambitious plan for a coordinated approach to decarbonising homes.

The team mapped the processes, structures and stakeholders that would be required to successfully plan and deliver street-by-street adoption of sustainable heating. This was followed by engagement with community groups, contractors, policymakers, local and national governments and manufacturers so they could contribute, critique and build on the proposed plans.

The feedback and insights helped define a series of **speed-testing** projects that will test key questions, riskiest elements and challenges around the coordinated switching journey. It also helped identify potential partners for **small local pilots** of collective switching and built a network of local bodies and other practitioners for shared learning and capability building.

What was the impact?

The test and learn approach helped turn the concept of street by street home decarbonisation into an implementable policy with actions owned by specific organisations.

The technique helped build a network around the policy of more than a hundred stakeholders from the heating industry, civil society and local government whose efforts will be needed to take it forward.

In a related piece of work the team have been building open source data tools to help other local areas replicate this sort of work. This could help reduce costs as building on what has been done before requires fewer resources than starting from scratch.

- **What lessons can be learnt?**
- **Don't let the perfect be the enemy of the good** – systems are often complex so aiming to fully understand them is not practical.
- **Knowledge of systems is often not codified** – key insights can be lost when people leave, as happened during this project. Carrying forward tacit knowledge through consistency of the team is important.
- **Use the process to build a coalition of the willing** – Engaging the system early increases, helps to promote collaboration across the system and the likelihood of test and learn partnerships.

How to identify test and learn opportunities?

Test and learn approaches can be applied to design, evaluate and scale most government policies and programmes. They help de-risk decisions and funding, and help navigate policy problems which are almost always complex. They can be used both when there is investment to support new programmes or services, and in instances where there is no additional funding to help deliver better outcomes for less. They are particularly useful for addressing:

1. **Priority challenges** such as those looking to deliver on the Plan for Change
2. **High level of uncertainty** or risk that could be reduced through small-scale testing and iterative learning
3. **Cross-cutting** challenges that require cross-departmental collaboration and partnership between central and local government and wider partners.
4. Challenges where there is **limited evidence on what works**.

For quick wins, the government should look for a number of criteria that make an opportunity particularly amenable to a test and learn approach. These are:

1. **Data availability** – easily available or easy to collect data, particularly on leading indicators that signal the likelihood of success and ideally also on outcome data to facilitate robust evaluation.
2. **Tractability** – opportunities where we already have a basic understanding of the problem and mechanisms of change to enable meaningful testing, and where the policy and regulatory environment allows us to experiment and adapt.
3. **Established programme or service provider(s)** – able to deliver the programme or services, with access or ability to generate the resources needed to operate the programme or service or where there is sufficient funding available to support delivery.

4. **Delivery infrastructure to scale** – potential to scale to a significant number of people, businesses or institutions beyond initial delivery. This might be existing mechanisms to scale, or where a programme already has reach and there is an opportunity to test and refine how it is delivered.

5. **Recognition and appetite to solve the problem** – especially by key partners such as frontline staff where most of the delivery capacity sits, with the opportunity clearly addressing user and partner needs and priorities.

These rules of thumb help identify contexts where test and learn approaches can provide the greatest opportunities for quick wins. However, some of the most transformative opportunities might not have all these hallmarks, and may require a longer-term strategic commitment and sustained effort to unlock test and learn's full potential.

How to embed test and learn?

The UK government already has many of the necessary building blocks required to work in a test and learn way. The civil service employs designers, economists, analysts, data scientists, evaluators, operational researchers and policy specialists. Government departments increasingly have evaluation strategies and use evaluation methods to understand the impact of government programmes and yield on returns relative to cost. There is well-established precedent for using design principles by the Government Digital Service over the past 10 years. And there are some good examples of iterative policy design, such as DWP's Universal Credit or Defra's Future Farming and Countryside programme. However, the use of these approaches is still rare, and there are very few examples where iterative policy design has been successfully combined with a robust evaluation of impact.

As we outline in the Nesta / Public Digital The Radical How report, embedding a test and learn approach sustainably requires radical new ways of working. These include:

- **Outcome-based accountability** – teams tasked with delivering outcomes not policies, activities or outputs, firmly defining goals while remaining flexible in the ways to reach them.
- **Agile ways of programme delivery,** shifting the traditional mindset of major programmes, which often divorces policy from delivery and relies on rigid, linear "waterfall" approaches that try to predict, plan, implement, test and launch a programme in a series of linear sequential stages of work.
- **Changes to funding rules** including those set out in the HM Treasury's Green Book that would enable funding that increases when risk decreases as more is learnt about programmes, rather than requiring a full economic appraisal upfront that locks in untested assumptions. This could be achieved through dynamic, stage-gated funding mechanisms alongside more locally controlled funding.
- **Changes to procurement systems** to incremental or flexible contracting mechanisms that anticipate and factor in uncertainty and focus on delivery of desired outcomes rather than ticking off deliverables.

- **New units of delivery** in the form of a multi-disciplinary team – accountable for outcomes, not outputs.
- **Improved data infrastructure** at both central and local government levels that can be used across government for real-time monitoring and decision-making.
- **Local capacity and capability** to innovate, monitor, evaluate and iteratively adapt.
- **Adaptive legislative frameworks** that create sandboxes for local experimentation.
- **Working in the open,** sharing insights, successes, and failures and openly and frequently publishing data on progress towards achieving outcomes.

Conclusion

Test and learn can help mitigate risky policy and funding decisions, avoid the 'sunk cost' fallacy, and achieve impact and change at scale quicker and more cost-effectively than current ways of working. While some policy reforms will inevitably proceed in a traditional fashion, progress could be made more quickly in many priority areas by embedding test and learn methods, building an up-to-date evidence base, and enhancing innovation capabilities. The UK Government has a unique opportunity to try these methods within the missions – they have the permission to break things and work differently in order to deliver. They could act as incubators for a new way of working.

Endnotes

1. A leading indicator is a measurable metric that changes in advance of, and helps predict, the eventual impact or outcome of a programme. Compared to impact measures, leading indicators appear earlier in the implementation process, often tracking implementation fidelity, intermediate processes, and early user responses. They suggest probable success or failure with varying levels of reliability. They are usually programme specific, tied to the theory of change and are primarily used for ongoing programme management and adjustment, whereas impact evaluation measures determine overall effectiveness and attribution.

58 Victoria Embankment
London EC4Y 0DS
+44 (0)20 7438 2500

information@nesta.org.uk
www.nesta.org.uk

info@bi.team
www.bi.team

Nesta is a registered charity in England and Wales with company number 7706036 and charity number 1144091. Registered as a charity in Scotland number SCO42833.

Registered office: 58 Victoria Embankment, London EC4Y 0DS.