



University  
of Exeter

# Testing, Learning and Adapting: an introduction to the new Magenta Book annex

Gloria Crabolu, University of Exeter  
Senior Lecturer in systems thinking and sustainability  
[g.Crabolu@Exeter.ac.uk](mailto:g.Crabolu@Exeter.ac.uk)



# Why 'Test & Learn'?

“Everyone agrees that the status quo is not working. The public, politicians and civil servants are all frustrated by the pace of change.”

Rt Hon Darren Jones MP



“Test it. Fix the problems. Change the design. Test it again. Tweak it again... 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?'”

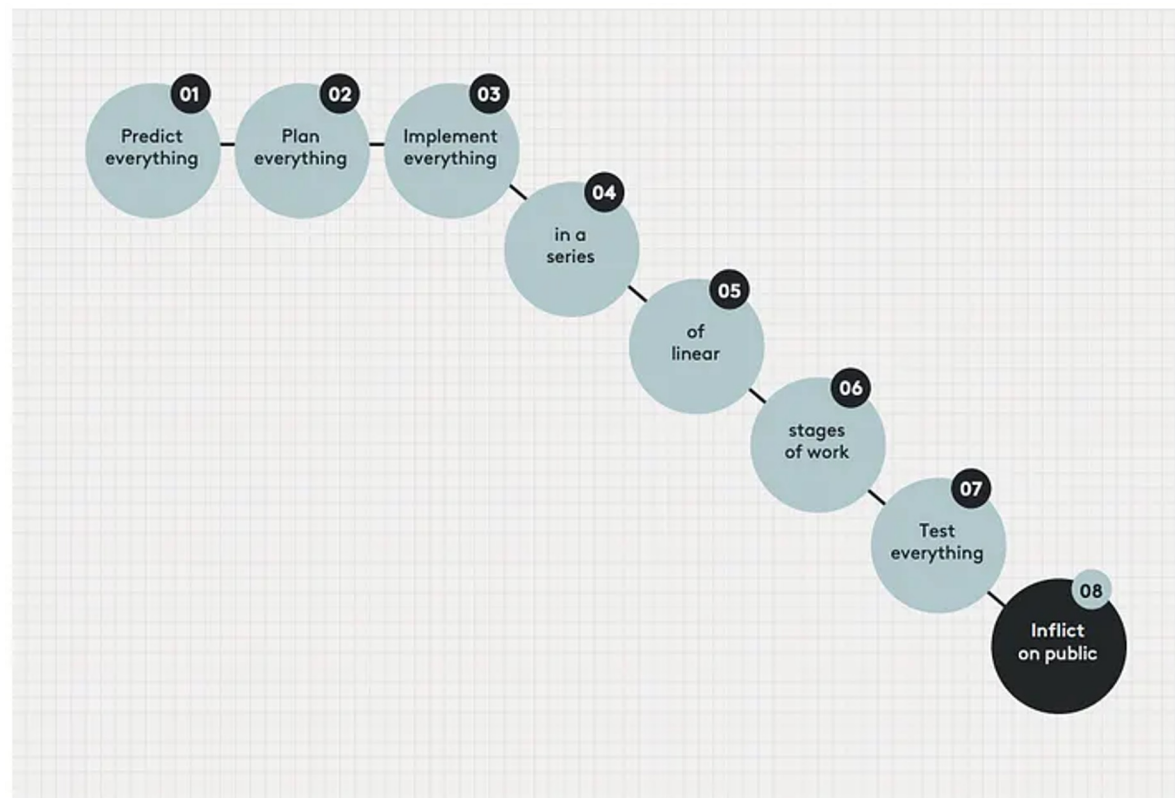
**Rt Hon Pat McFadden MP**



# Waterfall policymaking

Presumes:

- It is possible to collect all the information we need at the start of the process
- It is desirable to move through stages linearly without returning to previous stages
- Trying out interventions in the 'real-world' should be done at the end when everything is 'known'



Credit: Public Digital's [The Radical How](#)

# Test & Learn

## What it is

A way of working that uses structured testing, feedback and adaptation to improve how policies and services are designed and delivered , integrating research and evaluation thinking throughout the policy cycle rather than at the end of it

## What it is not

- ✗ Replacement for formal evaluation
- ✗ Static
- ✗ No clear outcomes

# Key features of Test & Learn



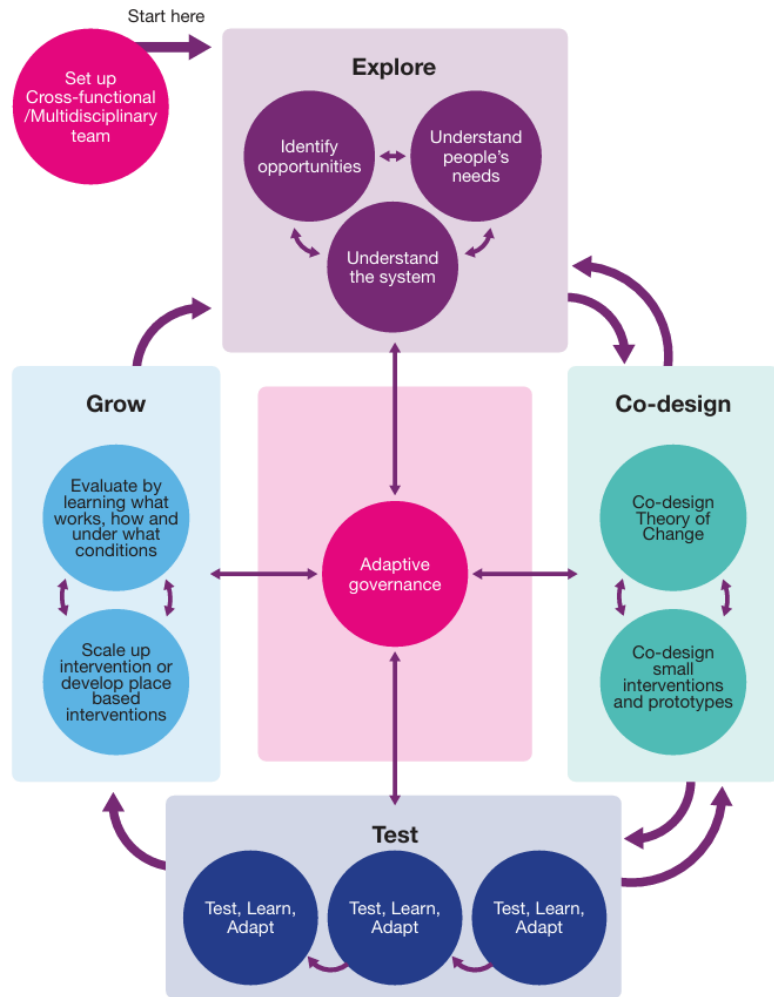
1. Delivered by multidisciplinary teams
2. Starts with shared, measurable outcome/s
3. Based on an evidence-based understanding of the challenge
4. Tests assumptions where there are greater uncertainty and risk
5. Iterative
6. Supports both scaling and local adaptation
7. Transparent and open

# Waterfall vs Test and Learn



<b>Waterfall approach</b>	<i>Dimension</i>	<b>Test and Learn</b>
Defined upfront by experts	<i>Problem framing</i>	Co-created with stakeholders
Taken for granted, untested	<i>Assumptions</i>	Identified, prioritised, tested early
Fixed before implementation	<i>Design process</i>	Iterative, adapted through cycles
Full rollout from the start	<i>Scale of delivery</i>	Start small, scale with evidence
Gathered retrospectively	<i>Evidence use</i>	Continuous, feeds back into design
Siloed, sequential handoffs	<i>Teams</i>	Cross-functional, multidisciplinary
Discovered late, costly to fix	<i>Failure</i>	Surfaced early, low cost to adapt
End-stage, often too late	<i>Evaluation</i>	Integrated throughout, then robust

# The Test and Learn framework



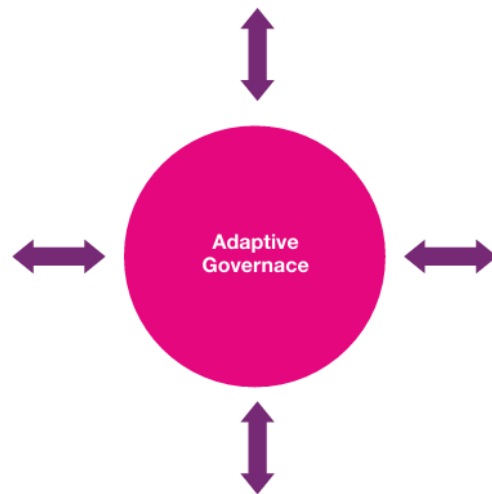
- Test and Learn Playbook – NESTA
- The human learning system – Centre for public impact
- Human Systems Dynamic – Glenda Eoyang
- Academic literature on adaptive approaches
- Internal government experience

# Why interdisciplinarity and adaptive governance matter



## Why interdisciplinarity matters

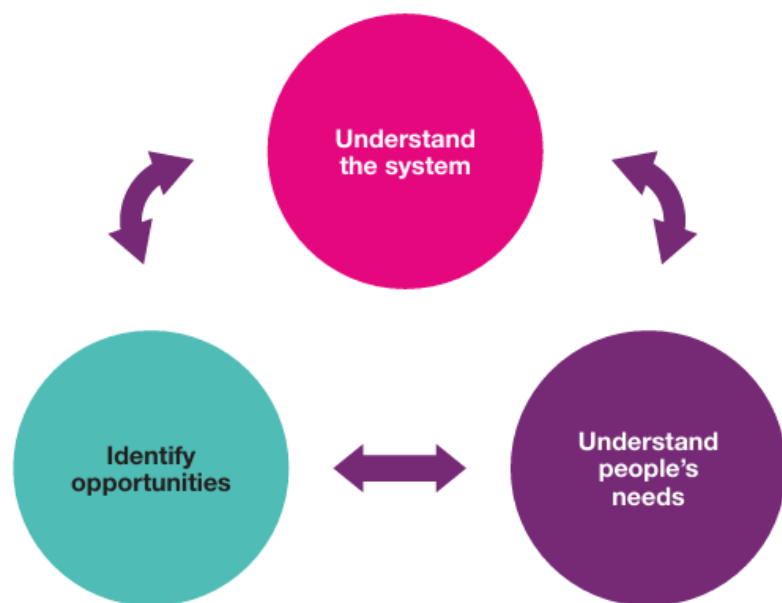
- Complex challenges span multiple stakeholders, departments and sectors, no single expertise can represent the full system.
- Teams drawing knowledge from across government departments, professions, local authorities, including a range of methodological expertise, possibly with external experts e.g. charities, academia and business tailor interventions to local context, reflecting the realities of delivery and the diversity of user needs.



## Why adaptive governance matters

- Adaptive governance conceptualises governance as a nested, multi-level set of institutions and networks that connect local, regional, national and transboundary actors. Rather than relying on fixed plans, it supports experimentation, negotiation and shared learning across jurisdictions and organisations, allowing rules, institutions and practices to adjust as new evidence emerges through the Test and Learn cycle.
- Rigid funding models, siloed decision-making and risk-averse cultures hinder adaptation. Adaptive governance addresses these by devolving authority and keeping oversight proportionate.
- It creates space for experimentation and accepts that not all tests will succeed making it possible to learn quickly, fail safely and scale what works.

# EXPLORE: Building a system understanding



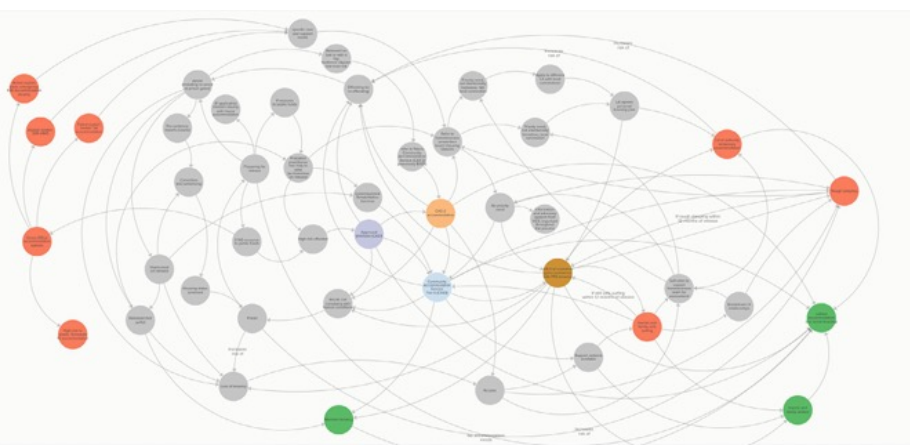
- **Systems aware:** Who are the system's actors, relationships, feedback loops and contextual factors affecting our main challenge? systems mapping, soft systems methodology or actor mapping
- **Participatory:** co-create a shared understanding by engaging those closest to the issue: service beneficiaries, policy teams, delivery partners, and academics
- **Identify opportunities and leverage points:** pinpoint where well-designed interventions could have disproportionate positive impact

# EXPLORE phase methods



Systems Approaches	e.g. systems mapping, actor mapping, SSM	Handling complexity guidance and systems thinking for civil servants
Other participatory, qualitative approaches	e.g. Stakeholder participatory workshops, interviews, focus groups	Handling complexity guidance and Main Magenta Book
Data and analytical approaches	e.g. Social network analysis, topic mapping	Handling complexity guidance and Main Magenta Book
Behavioural	e.g. Behavioural diagnosis	Achieving behaviour change – a guide for national government
Evidence synthesis	e.g. Rapid Evidence Reviews	Main Magenta Book

# Case study: a system wide evaluation of Homelessness and rough sleeping



[Systems-wide evaluation of homelessness and rough sleeping: Social housing allocations - GOV.UK](#)

- It brought together evidence across criminal justice, housing and homelessness systems to build shared understanding before any solutions was designed or tested. This gave policymakers robust evidence they could reference in strategy documents and spending review bids.
- The process exposed hidden feedback loops, unexpected pathways, and cross-sector dependencies, enabling more realistic intervention design and modelling of potential impacts

## What the analysis found

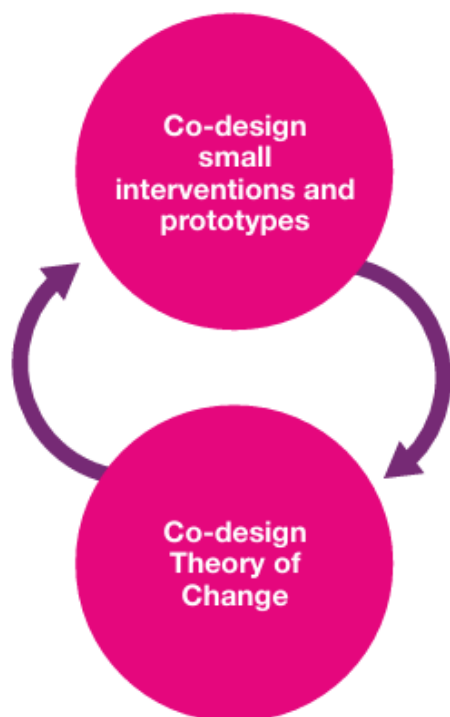
- Made visible the fragmentation between services and the feedback loops through which homelessness and reoffending reinforce one another
- Identified gaps in support at key transition points — particularly for women, those on remand and those serving shorter sentences

<b>Hidden Feedback Loop</b>	Funding overlaps caused recruitment competition, slowing programme implementation.
<b>Reinforcing Crisis Cycle</b>	Focus on crisis relief reduces investment in prevention, increasing future demand.
<b>Unexpected Pathway</b>	External factors (e.g., welfare, mental health, housing supply) funnel into homelessness trends beyond MHCLG's control.

## Impact

- Findings have already informed the National Plan to End Homelessness including changes to funding models and a review of housing allocation legislation

# The Co-Design Stage



- **Participatory and collaborative by design:** The participatory element is maintained throughout all phases. Engagement with stakeholders ensures that prototypes are enriched with lived experience, revealing assumptions, practical barriers and opportunities not visible in the data alone.
- **Generative and explorative:** the T&L team generate multiple ideas that address the key challenge identified in the Explore stage. Several options are explored in parallel. Small interventions might target different dimensions of the same issue or/and address different but connected challenges within the wider system. At any given point, a team may be progressing several prototypes simultaneously.
- **Theory of Change and iteration:** Co-design outputs directly inform the Theory of Change, setting out the causal pathway, key assumptions, risks and mechanisms to provide a robust framework for the Test phase. Prototypes showing potential are refined and retested against agreed success criteria; those that do not meet criteria are stopped, with learning documented. If ideas consistently fail to produce the intended behavioural effects, this should trigger a deliberate return to the Explore stage.

## Checklist: Practical steps for the Co-design stage

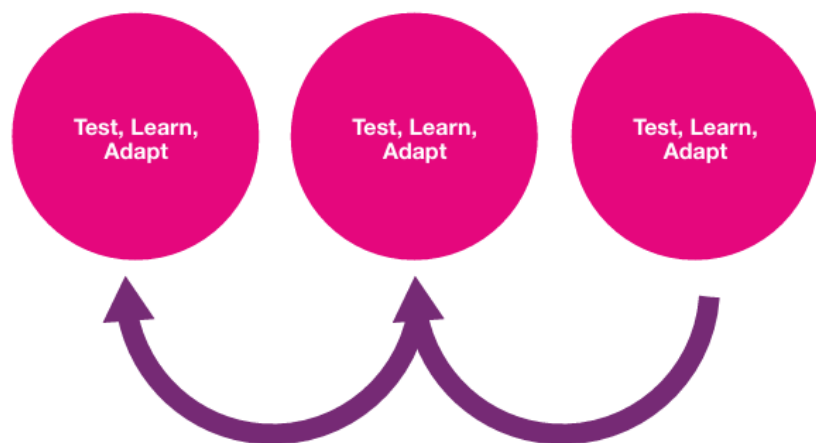
- ✓ **Engage key participants:** Involve frontline staff, service users, and partner organisations from the outset.
- ✓ **Generate ideas:** Use evidence from the Explore stage (data, evaluations, research, best practice) and inspiration from other sectors to create multiple potential solutions.
- ✓ **Create simple prototypes:** Develop low-cost, low-effort models (e.g. storyboards) to make ideas tangible.
- ✓ **Gather early feedback:** Test prototypes informally with stakeholders to uncover assumptions, barriers and opportunities.
- ✓ **Narrow down options:** Select the most promising prototypes based on feedback for further development.
- ✓ **Test in real-world settings:** Trial selected prototypes in a controlled, time-limited environment (offline or small-scale live test).
- ✓ **Observe and collect feedback:** Capture qualitative insights and quantitative data on usability, feasibility and impact.
- ✓ **Assess against success criteria:** Evaluate against agreed measures such as user satisfaction, operational feasibility, cost-effectiveness and policy alignment.
- ✓ **Refine or stop:** Improve and retest prototypes with potential – stop those that do not meet criteria, recording lessons learned and exploring explanations.
- ✓ **Develop and refine a Theory of Change:** Set out the causal pathway, assumptions, risks and mechanisms to inform the Test phase.

# Co-design phase methods



- **Observation and shadowing:** Watch how users or staff interact with the prototype in real time
- **Think-aloud testing/Interviews:** Ask participants to narrate their thoughts while using the prototype, revealing hidden assumptions and usability issues. Capture immediate reactions and insights from users or delivery partners
- **Focus groups or group feedback sessions:** Bring together users or frontline staff to discuss what worked, what didn't, and how the prototype could be improved.
- **Quick surveys or feedback forms:** Gather structured data on satisfaction, ease of use, or perceived impact;
- **Journey mapping:** Work with participants to chart their experience step by step, highlighting pain points, enablers and emotional responses.
- **Service walkthroughs/role-play:** Rehearse how the service would operate end to end, with users or staff playing roles to expose gaps and opportunities
- **Data analysis:** Include simple data analysis, such as tracking take-up, engagement levels or drop-off points to spot patterns quickly.

# The Test, Learn, Adapt Stage



## Purpose

- It focuses on testing the most critical or uncertain parts of the Theory of Change in a real-world setting, isolating specific components and assumptions rather than testing the entire intervention at once
- It focuses on whether the short-term outcomes and behavioural signals underpinning the Theory of Change are beginning to emerge.

## The approach

- **Start with the riskiest assumptions.** Prioritise components that are most uncertain or vital to success, building robust evidence about what works before investing in large-scale implementation.
- **Decide in advance.** Define success criteria and the evidence needed to justify continuation before testing begins, guarding against bias and over-commitment to approaches simply because time and effort have already been invested.
- **Learn systematically.** Feed evidence back into the Theory of Change iteratively, refining assumptions, risks and causal pathways throughout each funding cycle.
- **Only move to Grow once stable.** Design, delivery and core processes must be operating as intended before shifting focus from refinement to assessing overall effectiveness, impact and value for money.

# Test Learn and Adapt phase methods



- Qualitative Sprints
- A/B Testing
- Ecological momentary assessment (EMA)
- Systems based methods
- Nimble RCTs,
- User Testing
  
- ....

# GROW: Robustly evaluating to inform decisions on potential expansion

## Grow

Evaluate by learning what works, how and under what conditions

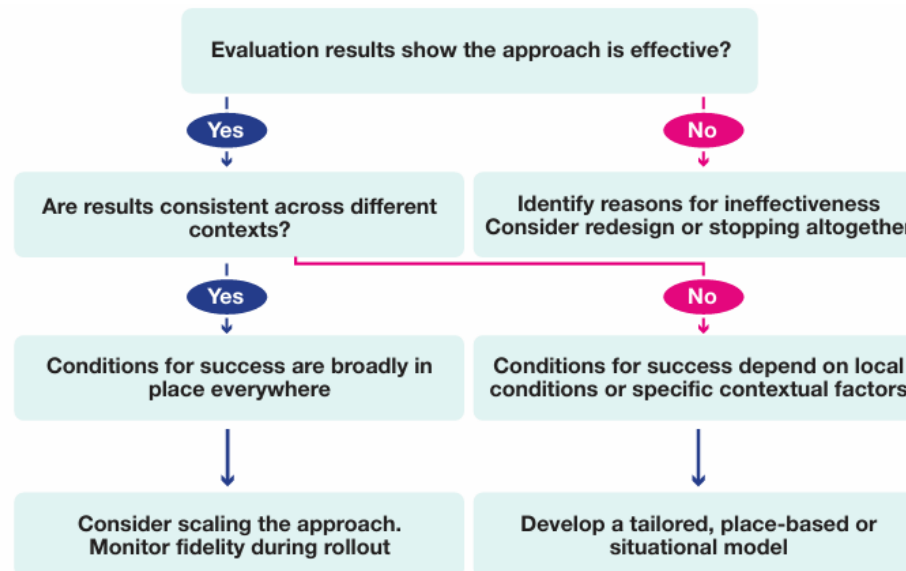


Scale up intervention or develop place based interventions

## Purpose

- Robustly evaluate whether an optimised, stable intervention achieves its intended outcomes, shifting focus from refining the approach to assessing effectiveness, impact and value for money.
- Test and Learn prepares interventions for this stage: stable design, clarified mechanisms and early feasibility evidence mean formal evaluation is applied to something well understood and evaluable by design

## The decision to scale or localise



# Test, Learn & Grow Programme

**Test and Learn** is a way of working, used to deliver user-centred services that improve over time

**Test, Learn and Grow** is a programme that applies test and learn at scale

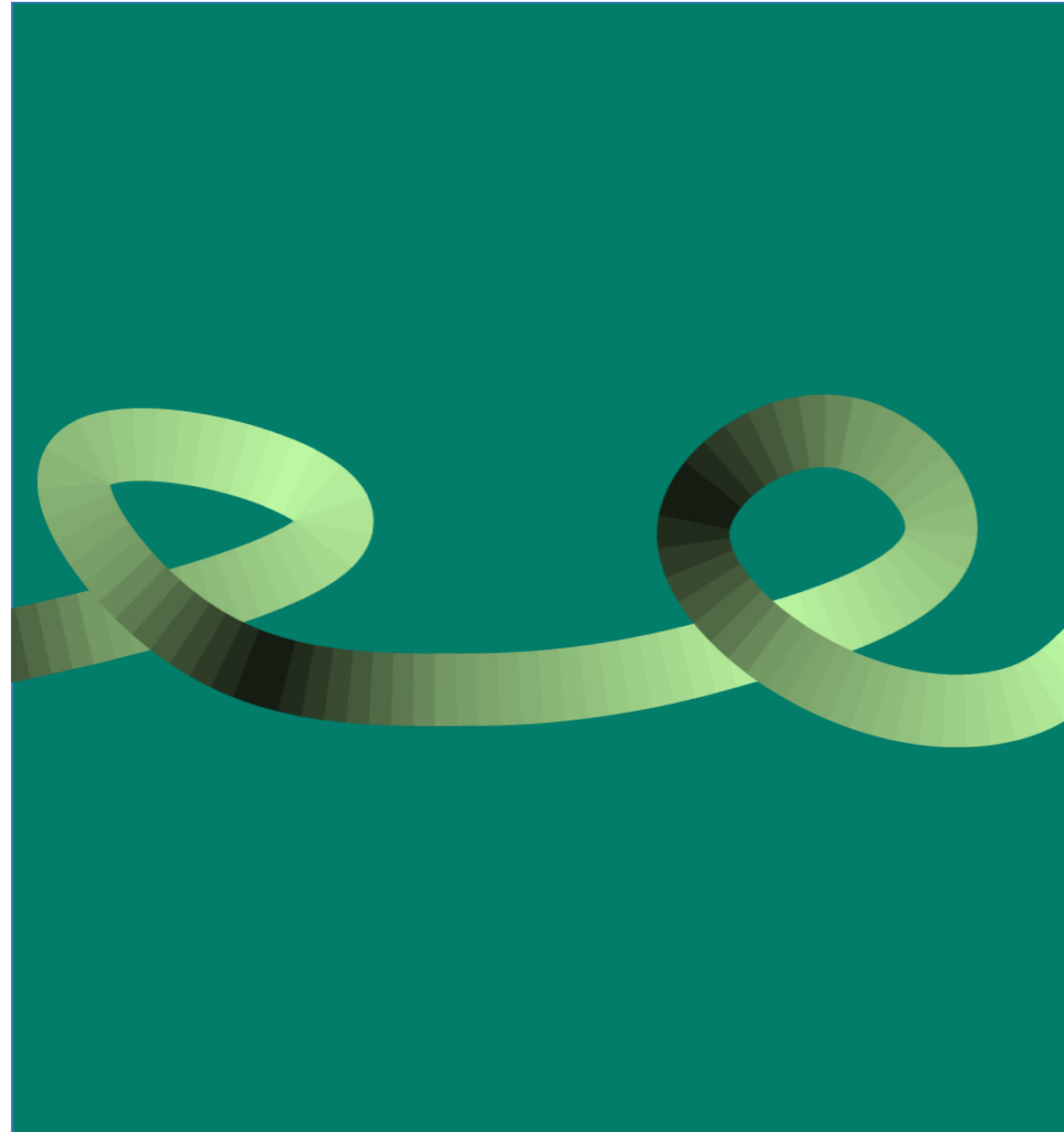
Slide courtesy of Rowena Bermingham



## **Test, Learn and Grow is here to:**

1. Show how test and learn works
2. Support people to test and learn
3. Remove barriers to using test and learn

Slide courtesy of Rowena Bermingham



**3**  
years

To design, test and scale solutions and the Test and Learn approach across government

**10 to 12**  
places

We will work in partnership with Places, where services meet communities

**6 to 10**  
issues

Spanning a range of challenge areas that are key to mission delivery and public service reform

**X**  
barriers

We will tackle barriers to innovation and reform at the centre of government – based on what we learn



University  
of Exeter

Slide courtesy of Rowena Bermingham

# How it will create this change

## 1. Show how test and learn works

### Accelerators

Use test and learn approaches with departments and local places to improve outcomes

## 2. Support people to test and learn

### Movement

Inspire through real examples and tell great stories and build the 'scaffolding' for others

## 3. Remove barriers to using test and learn

### Foundations

Shift underlying conditions so Test and Learn becomes normal

## 4. Evaluate impact of TLG

# Feedback loops



**Accelerators**

**Movement**

**Foundations**

## **We learn as we go**

What works locally informs national conditions, which enable more local changes

The approach will evolve based on what is learned

# Accelerators

Purpose: show test and learn methods work by:

Using test and learn methods in partnership with places, to improve outcomes

Sharing results and learnings to inform change nationally

Slide courtesy of Rowena Bermingham



## Locations

- Neighbourhood Health | Plymouth & Liverpool
- Best Start in Life | Northumberland & Manchester
- Violence Against Women & Girls | Islington
- Economic Inactivity | Wakefield
- AI & Data | Barnsley & Essex
- SEND | Nottingham & Sandwell

# Neighbourhood Health | Liverpool



In Liverpool, the work is focused on testing a **data-led respiratory health intervention** for children aged 0–5.

Using **targeted clinics** and referrals into wider support, the team will explore whether local partners can improve symptom management, address social and environmental drivers of poor respiratory health (e.g. housing conditions) and **reduce avoidable reliance on acute services** over time.

Kicked off with  
Liverpool: built  
relationships in  
place

Identified and  
onboarded  
experts into  
multidisciplina  
ry team

Used insights  
from place  
partners to  
select solution  
to test

Agreed test  
plan with place  
partners

Ran workshops with  
services to  
understand issues

Ran idea  
generation  
sessions to  
explore solutions

Co-created  
Theory of  
Change for  
solution with  
place partners



University  
of Exeter

# Thank you!

**Gloria Crabolu**  
**Senior Lecturer in Systems thinking and**  
**sustainability**  
[g.crabolu@exeter.ac.uk](mailto:g.crabolu@exeter.ac.uk)

**Last slides courtesy of Rowena Bermingham**  
**Head of Analysis and Evaluation**  
**Test, Learn & Grow Programme – Cabinet**  
**Office**