Department for
Business, Energy
& Industrial Strategy

Alternative policy evaluation frameworks and tools

Exploratory study

BEIS Research Paper Number 2020/044

October 2020

Acknowledgements

This report is prepared by the UCL Institute for Innovation and Public Purpose for the Better Regulation Executive at the Department for Business, Energy and Industrial Strategy.

The UCL Institute for Innovation and Public Purpose (IIPP) team was led by Mariana Mazzucato and Rainer Kattel, and included Sarah Albala, George Dibb, Martha McPherson and Asker Voldsgaard.

About the Institute for Innovation and Public Purpose

The UCL Institute for Innovation and Public Purpose (IIPP) is a department within University College London (UCL) — founded in 1826 to solve grand challenges — and part of The Bartlett faculty, known internationally for its radical thinking about space, design and sustainability. We apply our critical thinking to research and thought leadership; teaching and training; influencing public policy; and engaging the broader public.

ucl.ac.uk/iipp

Any enquiries regarding this publication should be sent to us at:
enquiries@beis.gov.uk

Contents

Executive summary 6
Recommendations 8
Introduction: Challenge-driven policies and evaluation practices 10
Aims and methodology of the study 20
Mapping policy evaluation and appraisal frameworks 21
Results 22
Overall results 22
Discussion of academic literature 23
Examples of alternative methodologies in academic literature 26
Social fabric matrices 26
Living labs 27
Public value mapping 28
Discussion of policy case studies 29
Examples of alternative methodologies from policy case studies: three pillars of evaluation 31
Overall discussion 33
Conclusion 35
Appendix 1. Literature review methodology in detail: Data selection process 36
Data sources 36
Data search 37
Data screening 37
Eligibility 39
Coding of eligible articles (academic and policy) 40
Definitions 40
Appendix 2. Literature review flow diagram 42
Appendix 3. MOIN survey 43

“We should evidently divide the science of measurement into two parts in accordance with what has been said. One part comprises all the arts which measure number, length, depth, breadth, and thickness in relation to their opposites; the other comprises those which measure them in relation to the moderate, the fitting, the opportune, the needful, and all the other standards that are situated in the mean between the extremes.”

(Plato, Statesman)

Executive summary

21st century politics is increasingly defined by the need to respond to major social, environmental and economic challenges. Sometimes referred to as 'grand challenges', these include threats like climate change, demographic, health and wellbeing concerns, as well as the difficulties of generating sustainable and inclusive growth. Governments are increasingly confronting the realisation that to solve these challenges they need to go beyond market-fixing approaches and instead consciously co-shape and co-create markets.

Policy design and evaluation frameworks matter because in many ways these frameworks constitute policy reality. The 'mainstream' approach to policy evaluation is derived from neoclassical economic theory, in particular microeconomic theory and welfare economics. It emphasises the idea that, given certain assumptions, individuals pursuing their own self-interest in competitive markets gives rise to the most efficient and welfare-maximising outcomes. Efficiency is understood in a utilitarian sense, whereby an activity is efficient if it enhances someone's welfare without making anyone else worse off (so-called Pareto efficiency). Under these conditions, the role of government intervention should be limited to addressing instances where the market is unable to deliver Pareto-efficient outcomes. In policy evaluation, this approach is epitomised by appraisal and evaluation techniques of static ex ante cost-benefit analysis (CBA). Costs (including the costs of potential government failure) are usually defined by their opportunity cost, i.e. the value which reflects the best alternative use a good or service could be put to (including a do-nothing/business as usual option) with market prices usually the starting point for the analysis. Policy evaluation, after the policy intervention, then seeks to verify whether the estimates were correct and whether the market failure was addressed.

The focus on market-fixing creates a particular orientation towards innovation, industrial policy and structural economic change. While certain elements of innovation policy, in particular early-stage R&D, can be considered to be a public good and thus justify public policy provision, the private sector is mainly assumed to be the more efficient innovator, possessing greater entrepreneurial capacity than the public sector, and better able to take risks given the pressure created by competition. In contrast, the state is viewed as risk-averse and in danger of creating government failure if it becomes too involved in industrial policy by 'picking winners'. Thus, the 'mainstream' policy framework of fixing market failures creates a relatively small envelop into which government 'performance' has to fit.

We propose an alternative approach, which begins with the notion of public value as collectively generated by a range of stakeholders, including the market, the state and civil society. Key here is the emphasis on value creation at the core: not 'public' value but value itself – with a clear delineation of the role of the different actors that are central to its formation. While in mainstream economics value is, in essence, created inside businesses and only facilitated by the public sector, in this view value is co-created and requires a stakeholder understanding of capitalism itself. If value is created collectively, what are the methods to evaluate value creation processes? This exploratory study seeks to map out state-of-the-art alternative methods of policy appraisal and evaluation, and how such evaluations deal with issue of value, market-fixing and market-shaping.

The study is based on extensive systematic literature review of academic discussions in policy appraisal and evaluation over the past five years up to 2019. The review covers academic publications, both mainstream and alternative economics and policy journals; and specialised policy analysis journals. The literature review is supplemented by selected international policy cases, including from the European Commission and a survey of the UCL IIPP Mission Oriented Innovation Network (MOIN). A workshop on alternative evaluation methods was held at the annual meeting of MOIN in Bellagio, Italy, in March 2019 which contributed to the initial review of alternative evaluation methods. The literature review is based on our innovative explorative conceptual framework, which enables us to create a matrix mapping market-shaping against market-fixing, and private value against public value.

The key conclusions of the study are:

- Most public organisations operate in a context of multi-dimensional appraisal and evaluation practices, in which frameworks and tools originate from various levels of governance and do not rely only on external data and evaluations, but are increasingly creating in-house datasets and evaluation tools. The emergence and rise of public digital and design agencies has led to an influx of diversity of new evaluation methodologies around user experience and service design. These methodologies avoid the idea of aggregated indicators or averages as measurements and focus on mapping a multitude of contextual as well as individualised factors. These agencies are also pioneering real-time monitoring practices, such as dashboards, and are challenging 'theatre of evaluation' that relies on single indicators or overly complex modelling.

- Public organisations are increasingly aware of the need to combine retrospective research and evaluation methodologies with prospective methods which emerge from design-oriented research and evaluation activities. Thus, policy evaluation is as much about understanding past actions as it is about shaping future policy frames and expectations.

- Public organisations are increasingly concerned about creating reflexive capabilities, particularly in relation to how these organisations create, maintain and evaluate public value. There is a clear need to significantly deepen our understanding of the various public value frames (e.g. economic, legal, organisational, etc) that can be used in evaluation practices. Reflexive capabilities connect evaluation and learning practices within organisations.

- Overall, we were able to identify 80 different policy evaluation and appraisal methodologies across all items listed in our review. These ranged from very traditional – cost-benefit analysis, cost-efficiency analysis, general equilibrium modelling and semi-structured or elite interviews (in the form, for instance, of commissions) to much less orthodox approaches – social fabric matrices, asset mapping, and public value mapping.

- Of all literature and policy case studies surveyed, 35% referred to market-fixing frameworks, while 46% referred to market-shaping frameworks and 24% referred to non-market-oriented frameworks.

- In academic literature, a larger percentage of the market-fixing frameworks was discernible in more orthodox economic publications, where 55% of studies analysed were using market-fixing frames; in heterodox economic journals, the proportion was 44% market-fixing and 44% market-shaping frameworks. In specialised evaluation literature, we saw that 82% of all articles were on public value-led policies, compared to

36% in heterodox and 43% in orthodox journals. This implies that alternative evaluation methods are being taken seriously by the academic and specialist community.

- As anticipated, market-fixing frameworks are heavily reliant on quantitative models; analytical methods for market-shaping are, on the other hand, much more diverse, ranging from quantitative approaches, such as agent-based modelling, to qualitative approaches, such as living labs.
- While alternative approaches to appraisal and evaluation are increasingly being developed by innovative agencies, and are more frequently seen in orthodox academic literature, they are far from standard tools for policy-makers. The study shows that, while not averse to new, public-value oriented evaluation approaches, practitioners often prefer to employ multiple mainstream methodologies. This offers diversification and triangulation, but is not an ambitious leap in terms of alternative approaches.

Recommendations

- As governments are focusing on grand challenges, current practices of policy appraisal and evaluation need to be supplemented by alternative approaches. For instance, one pragmatic change would be for the UK's Green Book and Magenta Book to include in their annexes the identified typologies of appraisal and evaluation (for instance, the categories of economic evaluation, empirical impact evaluation and process evaluation), and to use a research-informed practice model and incorporate some of the known innovative policy assessment techniques and their uses in the guidance. Such typologies could be reflected in the business cases departments and agencies they are preparing for funding decisions.
- Our study advocates for both a wider user of alternative evaluation and appraisal methodologies, as well as a synthesis of a mixture of different techniques. For instance, the Green Book claims to focus on improving social welfare or wellbeing (social value). However, since this metric is currently based on standard economic evaluation techniques, it can only capture social value to a limited degree. Complementary tools must be integrated as standard procedure to provide a fuller picture for decision-makers.
- Further, policy appraisal and evaluation should be held to a higher standard than a do-nothing approach or business as usual. This framework inherently sets the standard fairly low in terms of policy outcomes as it will only focus on a single singular parameter changing and will only seek a marginal improvement. Business as usual does not foster an innovative policy landscape as the evaluation and appraisal toolkits are set at this level. Also, business as usual is a false premise. The contextual policy environment is always in flux and thus the aim should be for appraisals and evaluations to best capture the changing environment, and further advocate for active market-shaping, as markets are inherently being altered in any case.
- The study suggests that not only do the policy appraisal and policy evaluation frameworks have to be altered, but so too should there be consideration for evaluation and appraisal of regulatory practices. Policy is only one tool that can create change for the betterment of the economic, public and social spheres. For instance, the UK's Better Regulation Framework is specifically focused on regulations that have the greatest potential impact on business and civil society organisations. Evidence, transparency and proportionality are quoted as being essential to the formation of appropriate regulatory mechanisms. Regulatory provision in relation to business activity is a statutory provision which imposes or amends requirements, restrictions or conditions; or sets or amends

guidance in relation to the activity or the securing of compliance with, or enforcement of, requirements, restrictions, conditions, standards or guidance for the activity. Similar to the general strategies for policy evaluation and policy appraisal, the Better Regulation Framework is also based on the ROAMEF cycle (Rationale, Objectives, Appraisal, Monitoring, Evaluation and Feedback) and uses cost-benefit analysis. This is clearly a private value-oriented approach; it is pivotal that such frameworks to analyse regulatory impact are complemented by public value and market-shaping frameworks and tools to enable better decision-making.

- Finally, we recommend the Better Regulation Executive analyses and tests ways in which mainstream and alternative appraisal and evaluation methodologies could be brought together into a unified guidance for policy-makers. Formation of a community of practice around appraisal and evaluation methodologies could be a first step in this direction.

In summary, the existence and uptake of challenge-driven policy frameworks make it clear that the nature of evidence, evaluation and performance in policy-making is changing. First, rather than focusing on single instruments and representative agents, there is increasing consensus that we should analyse and evaluate policy mixes, as the impact of various instruments is rarely linear and can indeed be contradictory, with significant time lags and spillover effects. Second, policy-making takes place in a multiplicity of feedback channels that generate diverse evidence and perspectives on evidence. That is, policy appraisal and evaluation should take account of various contested values inherent to policy-making.

Introduction: Challenge-driven policies and evaluation practices

- 21st century politics is increasingly defined by the need to respond to major social, environmental and economic challenges. Sometimes referred to as 'grand challenges', these include threats like climate change, and demographic, health and wellbeing concerns, as well as the difficulties of generating sustainable and inclusive growth. Governments are increasingly confronting the realisation that to solve these challenges they need to go beyond market-fixing approaches and instead consciously co-shape and co-create new markets, as well as upgrade existing ones.

- A compartmentalised public sector is not fit for purpose to tackle highly interconnected and cross-generational challenges such as climate change and instead is turning to challenge-driven policy frameworks. These frameworks are being applied in research and development (R&D) and industrial and innovation policy areas as mission-oriented policies (e.g. the European Union's Horizon Europe 2021-2027[1]; the UK's Industrial Strategy from 2017[2] (see Box 1 below)) and in fiscal policy as cross-ministerial goals (based on Sustainable Development Goals (SDGs) as in Scotland) or on wellbeing (as in Wales and New Zealand). At the core of such new policies and agendas is the realisation that new policy frameworks need to be based on 'new economics' that enable us to develop sensitive and targeted tools and metrics for challenge-driven policies.[3] Such tools include quantitative approaches (e.g. dynamic efficiency measurements in the form of multiplier effect and crowding in of R&D investments) and qualitative approaches (e.g. systems mapping, public value mapping).

- Challenge-driven policy frameworks such as the ones mentioned above are emerging in parallel to well-established modernisation,[4] competitiveness[5] and performance management[6] frameworks. This is also reflected in the increasing diversity of the policy analysis landscape. While modernisation, and in particular competitiveness. frameworks rely on the idea that government should first and foremost fix market failures,[7] a challenge-driven agenda does not have such clearly defined theoretical origins and analytical lenses. Rather, challenge-driven policies are focused on specific 'wicked' socio-economic problems, such as air pollution in cities, which do not have single

[1] Mariana Mazzucato (2018), *Mission-Oriented Research & Innovation in the European Union: A problem-solving approach to fuel innovation-led growth.* Report for the European Commission: https://ec.europa.eu/info/sites/info/files/mazzucato_report_2018.pdf.
[2] UCL Commission on Mission-Oriented Innovation and Industrial Strategy (MOIIS), co-chaired by Mariana Mazzucato and David Willetts (2019). A Mission- Oriented UK Industrial Strategy. UCL Institute for Innovation and Public Purpose, Policy Report Working Paper Series (IIPP WP 2019-04): https://www.ucl.ac.uk/bartlett/public-purpose/wp2019-04.
[3] Mariana Mazzucato (2016), From market-fixing to market-creating: A new framework for innovation policy. *Industry and Innovation*, 23(2), pp. 140-156.
[4] See, for instance, the Blair government's *Modernising Government* white paper from 1999: https://www.civilservant.org.uk/library/1999_modernising_government.pdf.
[5] The EU's Lisbon Strategy from 2000 is perhaps the best-known example of this. See: https://portal.cor.europa.eu/europe2020/Profiles/Pages/TheLisbonStrategyinshort.aspx.
[6] Geert Bouckaert and John Halligan (2008), *Managing Performance. International Comparisons*, Routledge; and Allen Schick (2003), The Performing State: Reflection on an Idea Whose Time Has Come but Whose Implementation Has Not, *OECD Journal on Budgeting*, 3, 2, pp.71-103.
[7] On competitiveness, see Erik S. Reinert (1995), Competitiveness and its predecessors – a 500-year cross-national perspective, Structural Change and Economic Dynamics, 6, 1, pp. 23-42: https://doi.org/10.1016/0954-349X(94)00002-Q.

solutions or analytical approaches within current value frameworks. This leads to policy evaluation frameworks and practices that are increasingly diverse and formed through the confluence and direction of multiple inputs and actors. As argued by Kroll, "Compared to traditional programme evaluation, the combined assessment of innovation strategies and strategically embedded innovation policy mixes poses a substantially more difficult challenge – for reasons more fundamental than mere complexity. As is broadly accepted, political decisions emerge in unpredictable, iterative processes so that the transfer of grand ideas to concrete instruments constitutes an iterative and in many ways open-ended process of learning in practice."[8]

Box 1: What is mission-oriented innovation?

Mission-oriented policies tackle complex challenges – whether technological or social – in bold and ambitious ways that spark the development of many different solutions. Missions informed the structure of the EU's Horizon 2020 programme and the UK government's Industrial Strategy.

A mission is a concrete target and an achievable step within a societal challenge that acts as a framing and a stimulus for innovation projects. Using missions to drive industrial strategy or innovation policy means focussing less on sectors and more on problems that multiple sectors need to solve together.

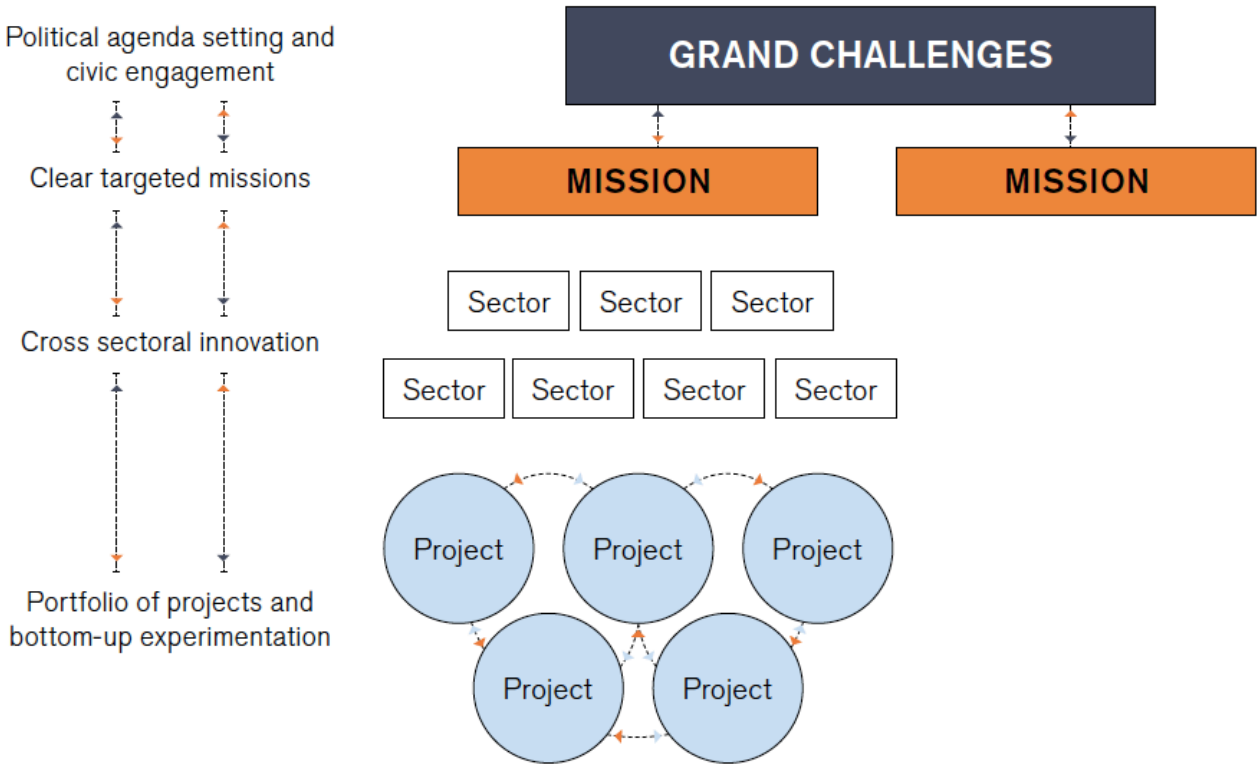
Missions are inspired by challenges, but are more granular and concrete in that you can answer whether you have or have not achieved them. How do you go from a challenge to a mission? To do so the mission must set clear objectives that can only be achieved by a portfolio of projects and supportive policy interventions.

We have set out five criteria for the development of missions. They should:

- Be bold and inspirational with wide societal relevance
- Set a clear direction, and be targeted, measurable and time-bound
- Be ambitious but realistic
- Be cross-disciplinary, cross-sectoral and cross-actor
- Involve multiple, bottom-up solutions

[8] Henning Kroll (2019), How to evaluate innovation strategies with a transformative ambition? A proposal for a structured, process-based approach, *Science and Public Policy*, 2019, pp. 1–13: https://doi.org/10.1093/scipol/scz016

Grand challenges and missions

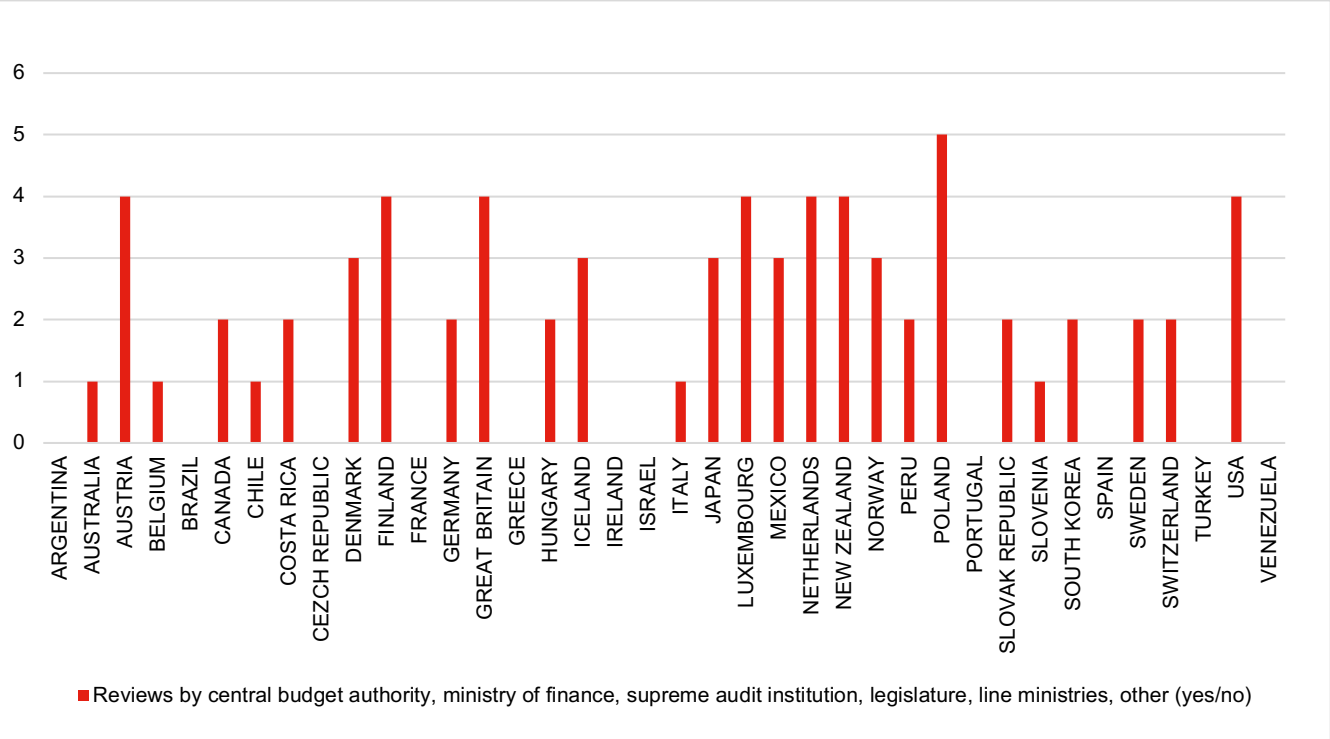


Source: Mariana Mazzucato (2018), Mission-Oriented Research & Innovation in the European Union. European Commission: https://ec.europa.eu/info/sites/info/files/mazzucato_report_2018.pdf.

- The modernisation and competitiveness-focussed policy frameworks of the1990s and 2000s relied heavily on performance targets in both outputs and outcomes, and on related appraisal and evaluation practices.[9] For instance, by the mid 2000s, most OECD countries had installed efficiency and cost-effectiveness reviews on multiple levels of governments (e.g. central budget authorities, line ministries, audit offices – see Figure 1 below).

[9] For a useful discussion, see Christopher Hood (2007), Public Service Management by Numbers: Why Does it Vary? Where Has it Come From? What Are the Gaps and the Puzzles?, *Public Money & Management*, 27:2, 95-102: https://doi.org/10.1111/j.1467-9302.2007.00564.x. For a discussion of productivity and performance targets, see Rainer Kattel, Alex Cepilovs, Wolfgang Drechsler, Tarmo Kalvet, Veiko Lember and Piret Tõnurist (2013), Can we measure public sector innovation? A literature review, LIPSE working paper: https://www.researchgate.net/publication/273887292_Can_we_measure_public_sector_innovation_A_literature_review.

Figure 1: Efficiency and/or cost-effectiveness reviews in OECD countries, 2007



Source: OECD Survey on Budget Practices and Procedures 2007: https://www.oecd.org/governance/budgeting/43829277.xls.

- Within the UK, government's Green and Magenta Books, which cover the topics of policy appraisal and policy evaluation, illustrate the focus on efficiencies and cost-effectiveness. The most highly valued analyses include cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA). Despite being held up to the highest gold standard of evaluation, these appraising techniques, as outlined in the Green Book, are acknowledged to be limited as they rely on the assumption that the broad environment remains unchanged as a result of the intervention. They do poorly in handling dynamic policy actions and rather can only capture marginal changes when conditions are thought to remain stable.[10]

- Further, the focus on CEA and CBA holds up market value evaluations, and even when appraising policies which entail changes to non-market goods or instances of unclear market values, limited policy evaluation and appraisal tools effectively straitjacket policy outputs and objectives to be phrased within market evaluation of comparable products. As part of the ROAMEF cycle (Rationale, Objectives, Appraisal, Monitoring, Evaluation and Feedback), a hierarchy of appraisal and evaluation techniques is mentioned whereby, in instances where market values cannot be assessed, there is a schema on how best to assign such a market value. This includes comparing the value of similar costed items, linking price to a pre-determined transferable price as set in the Green Book price reference table, and using revealed preference pricing on how that good would behave in similar markets. The next tier of rigour when determining the value of policy output or programme includes assessment and modelling of willingness to pay and willingness to accept loss. At the bottom tier of credible and suggested policy appraisal methodologies are those that take into account direct wellbeing based on

[10] The Green Book: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf.

responses, and valuing the output based on stakeholder interaction or other dynamic processes. A wellbeing metric that is a composite of multiple ways of assessing a policy is viewed as substandard, whereas assessing policy performance using existing market value indicators that are 'clear cut' are placed at the highest priority level, as noted in the Green Book. Such rejection of a valuing practice that includes incorporating either multiple vantage points of a policy's value, or considers policy output within a dynamic and evolving context, demonstrates some of the existing weaknesses in what is the most up-to-date policy appraisal document for the UK.

- The Magenta Book, which focuses more on policy evaluation, uses the broad categories of process evaluation, theory-based evaluation, economic evaluation and empirical impact evaluation. While the Magenta Book claims to offer more insight into alternative policy evaluation methodologies, which ought to be beneficial for promoting meaningful and varied tools for evaluating policy, the lack of any specific methodological tools beyond the broad suggested categories of evaluation genres is ineffective at promoting the use of innovative and dynamic policy evaluation methods. By not providing a clear framework with either case studies or specific examples of particular policy evaluation methodologies, a policy practitioner is left with little specific guidance or support to use alternative and differing methods, and will consequently default to reinforced market-focused policy evaluation toolkits, such as economic impact evaluations, ergo CEA and CBA.[11]

- The Green and Magenta Books' approach to policy appraisal and evaluation is derived from neoclassical economic theory, in particular microeconomic theory and welfare economics. It emphasises the idea that, given certain assumptions, individuals pursuing their own self-interest in competitive markets gives rise to the most efficient and welfare-maximising outcomes. Efficiency is understood in a utilitarian sense, whereby an activity is efficient if it enhances someone's welfare without making anyone else worse off (so-called Pareto efficiency). Under these conditions, the role of government intervention should be limited to addressing instances where the market is unable to deliver Pareto-efficient outcomes.[12]

- Policy design and evaluation frameworks matter because in many ways these frameworks constitute policy reality.[13] The focus on efficiency only creates a particular orientation towards innovation, industrial policy and structural economic change. While certain elements of innovation policy, in particular early-stage R&D, can be considered to be a public good and thus justify public policy provision, the private sector is mainly assumed to be the more efficient innovator, possessing greater entrepreneurial capacity than the public sector and better able to take risks given the pressure created by competition. In contrast, the state is viewed as risk-averse and in danger of creating government failure if it becomes too involved in industrial policy by 'picking winners'. Its role is to level the playing field for commercial actors – mostly through supply-side inputs such as better skills or the removal of market frictions – and then get out of the way. However, retrospective analysis of how radical technologies such as the Global

[11] The Magenta Book: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf.
[12] For a more detailed analysis, see Rainer Kattel, Mariana Mazzucato, Josh Ryan-Collins and Simon Sharpe (2018), The economics of change: Policy appraisal for missions, market-shaping and public, IIPP Working Paper, no. 2018-06: https://www.ucl.ac.uk/bartlett/public-purpose/wp2018-06.
[13] B Guy Peters, Giliberto Capano, Ishani Mukherjee, Meng-Hsuan Chou, Michael Howlett and Pauline Ravinet (2018), *Designing for Policy Effectiveness: Defining and Understanding a Concept* (Elements in Public Policy), Cambridge University Press: https://doi.org/10.1017/9781108555081.

Positioning System (GPS) and the internet were developed presents a very different view of how public and private actors interact to generate technologies with enormous technological and economic spillovers.[14]

As argued by Richard Nelson in 1959, welfare economics-driven market failure approaches are good at identifying problems, such as areas with under-investment in R&D, but quite a poor guide for identifying areas with the potential highest 'social profit'.[15] As Kenneth Arrow aptly put it: "Formally, of course, resources should be devoted to invention until the expected marginal social benefit there equals the marginal social benefit in alternative uses, but in view of the presence of uncertainty, such calculations are even more difficult and tenuous than those for public works. Probably all that could be hoped for is the estimation of future rates of return from those in the past, with investment in invention being increased or decreased accordingly as some average rate of return over the past exceeded or fell short of the general rate of return."[16] Thus, challenge-driven policies such as mission-oriented innovation policies are a conscious decision to go beyond a market failure approach to policy-making and necessarily create a need for new alternative policy appraisal and evaluation tools.

- We propose that evaluation of challenge-driven policies should begin with the notion of public value as collectively generated by a range of stakeholders, including the market, the state and civil society. Key here is the emphasis on value creation at the core: not 'public' value but value itself – with a clear delineation of the role of the different actors that are central to its formation.[17] While in mainstream economics value is, in essence, created inside businesses and only facilitated by the public sector, in this view value is co-created and requires a stakeholder understanding of capitalism itself. If value is created collectively, what are the methods to evaluate value creation processes? This study therefore seeks to map out state-of-the-art alternative methods of policy appraisal and evaluation, and how such evaluations deal with issues of value, market-fixing and market-shaping.

From market failure to market-shaping and public value

- CBA-type analyses are concerned with allocative or distributive efficiency, which involves making the best use of (fixed) resources at a fixed point in time. Market-shaping policy and challenge-driven innovation, on the other hand, are focused on making the best use of resources to achieve changes over time, including, perhaps most importantly, the creation of new technologies and/or the shifting of technology frontiers. Such change will impact multiple sectors and prices.[18]

- Instead, existing 'mainstream' (orthodox) approaches focus on static efficiency. Analytical frameworks that support the understanding and pursuance of dynamic efficiency have evolved from evolutionary (heterodox) economics[19] and are already in

[14] Mariana Mazzucato (2013), *The Entrepreneurial State: debunking public vs. private sector myths*, Anthem Press.
[15] Richard R Nelson (1959), The simple economics of basic scientific research, *Journal of political economy*, 67(3), pp. 297-306.
[16] Kenneth Arrow (1962), Economic Welfare and the Allocation of Resources for Invention, in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, NBER, Princeton University Press, pp. 609-626.
[17] Mariana Mazzucato (2018), *The Value of Everything*, Penguin; and Mariana Mazzucato and Josh Ryan-Collins (2019), Putting value creation back into 'public value': From market-fixing to market-shaping, UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-05): https://www.ucl.ac.uk/bartlett/public-purpose/wp2019-05.
[18] Kattel et al 2018, op. cit.
[19] Richard Nelson and Sydney Winter (1974), Neoclassical vs Evolutionary Theories of Economic Growth: Critique and Prospectus, *The Economic Journal*, 84 (336), pp. 886-905.

use within some governments and international organisations. An example is 'systems thinking', which explores and develops effective action by looking at connected wholes rather than separate parts, and is focussed on the identification of feedback loops.[20] The interaction of reinforcing and balancing feedbacks is the main determinant of how a complex system behaves over time. In the UK government, the technique has also been used by the Department for Education[21] in developing policy on child protection, by the Government Office for Science in understanding obesity,[22] and by the Department for International Development for analysing the interplay between prevention and cure in health systems.[23]

- Related approaches include system dynamics and complexity economics, both of which can provide a quantitative understanding of how systems will behave over time under general conditions, without being limited to the special case of equilibrium. This enables explanation and modelling of how dynamic phenomena, such as exponential growth, asset price bubbles and crashes, patterns in price volatility, network effects, innovation, system lock-in and system change, arise within the economy.[24] The Bank of England has used complexity economics to deepen its understanding of the UK housing market and the likely effects of different forms of government intervention.[25] The literature on techno-economic paradigms and technology transitions includes frameworks for understanding how societies and markets move from one technological revolution and its related 'best practices' or dominant 'socio-technological regime' (a family of technologies fulfilling a social function) to another.[26]
- Table 1 below summarises market-fixing approaches to policy evaluation, characterised by static efficiency; and market-shaping approaches to policy evaluation, characterised by dynamic efficiency.

[20] Donella Meadows (2008), *Thinking in Systems: A Primer*, Chelsea Green Publishing.
[21] Department for Education (2011), Munro review of child protection: a child-centred system: https://www.gov.uk/government/publications/munro-review-of-child-protection-final-report-a-child-centred-system.
[22] Government Office for Science (2007), Reducing obesity: obesity system map: https://www.gov.uk/government/publications/reducing-obesity-obesity-system-map
[23] Department for International Development (2014), Advancing the application of systems thinking in health: why cure crowds out prevention: https://www.gov.uk/dfid-research-outputs/advancing-the-application-of-systems-thinking-in-health-why-cure-crowds-out-prevention.
[24] W. Brian Arthur (2014), Complexity economics: A different framework for economic thought, in W. B. Arthur, *Complexity and the Economy*. Oxford University Press: http://tuvalu.santafe.edu/~wbarthur/Papers/Comp.Econ.SFI.pdf.
[25] Rafa Baptista, J Doyne Farmer, Marc Hinterschweiger, Katie Low, Daniel Tang and Arzu Uluc (2016), Macroprudential Policy in an Agent-based Model of the UK Housing Market, Bank of England Working Paper, no. 619.
[26] Frank Geels and Johan Schot (2007), Typology of sociotechnical transition pathways, *Research Policy*, 36,3, pp. 399-417; and Carlota Perez (2003), *Technological Revolutions and Financial Capital*, Edward Elgar Publishing.

Table 1: Market-fixing vs market-shaping policy frameworks[27]

	Market-fixing	Market-shaping/mission-oriented
Justification for the role of government	Market or coordination failures: • Public goods • Negative externalities • Imperfect competition/information	All markets and institutions are co-created by public, private and third sectors. Role of government is to ensure markets support public purpose
Business case appraisal	Ex-ante cost-benefit analysis (CBA) – allocative efficiency assuming static general relationships, prices etc	Focused on systemic change to achieve mission – dynamic efficiency (including innovation, spillover effects and systemic change)
Underlying assumptions	Possible to estimate reliable future value using discounting/ monetisation of externalities/risk assessment; system is characterised by equilibrium behaviour	Future is uncertain because of potential for novelty and non-marginal change; system is characterised by complex behaviour
Evaluation	Focus on whether specific policy solves market failure and whether government failure is avoided (Pareto-efficient)	Ongoing and reflexive evaluation of whether system is moving in direction of mission via achievement of intermediate milestones. Focus on portfolio of policies and interventions, and their interaction
Approach to risk	Highly risk-averse; optimism bias assumed	Failure is accepted and encouraged as a learning device

- In innovation policy evaluations the focus tends to be narrowly defined and on easily measured targets rather than wider behavioural, dynamic or systemic effects.[28] After reviewing more than 800 policy reports and academic papers in innovation policy evaluation, Edler et al conclude, “The supply side instruments typically seek to address market failures” while the demand side instruments “seek to address deficiencies in the ability and willingness of potential users to demand and apply innovation.”[29] The analytical tools to understand and capture the effect of policy mixes remain particularly weak, pointing to the need to increase capabilities in public organisations to monitor policies, particularly mission-oriented portfolios, rather than rely on external evaluations. In essence, the monitoring and evaluating of innovation policy mixes is not currently considered to be part of governance and implementation processes, and accordingly the evaluation tends to focus on efficiency gains rather on understanding the effectiveness of

[27] Kattel et al 2018, op. cit.
[28] Jakob Edler, Philip Shapira, Paul Cunningham and Abdullah Gök (2016), Conclusions: Evidence on the effectiveness of innovation policy intervention, in Jakub Edler, Paul Cunningham and Abdullah Gök (eds.), *Handbook of Innovation Policy Impact*, Edward Elgar Publishing, pp. 543-564.
[29] Ibid., p. 544.

policy mixes.[30] Within a market-shaping, dynamically open framework, these activities would come closer to the remit of governance.

- The impact of neoclassical economics on private sector evaluation practices and culture has been enormous. Within New Public Management (NPM) reforms in the 1990s and 2000s, the main thrust was to introduce market-like incentives in the public sector as well. This was captured in ideas such as decentralising agencies, contracting out various functions (from the back office to actual service delivery), and performance pay for managers and agencies.[31] Put simply, NPM reforms were meant to rein in public costs and make public organisations more efficient. As perhaps the most significant result of these reforms – at least for innovation policy and for innovation in government – modern governments tend to rely on unified and holistic budgetary and performance systems in which all activities of government are seen through the same lenses of efficiency and saving – and not in terms of effectiveness or outcomes.[32] Innovation processes assume that not all activities are alike and that these differences matter. For public management, however, the reform waves of the 1980s and 1990s created a particularly strongly unified and specific vision of public sector performance. As argued by Bouckaert and Peters, this meant focusing on the 'visible performance' of lower level activities – 'frog view' – and not on higher level (e.g. cross-organisational) policy fields.[33]

- The 'frog view' in appraisal and evaluation manifests itself today in the concern to precision-target government support on the project level, and the development of ever more complex and mostly market-failure-driven policy evaluation tools to understand the impact of such support, mostly with mixed success.[34] This approach assumes that there are correct answers to any policy question that a specific project or proposal can address and that we need to find the proper measurement frameworks through increasingly sophisticated data analytics. This resembles what happens in many team sports where the so-called 'Moneyball' approach to finding talent and the best way to play relies increasingly on advanced statistical models.[35] In economic policy, this approach moves focus away from political discussions about the broader direction of policies, and towards discussion of single instruments and measurement. However, as important as measurement is, not only is innovation inherently open-ended, non-linear and rife with uncertainty, innovations also inherently challenge existing institutional frameworks and values (e.g. sharing economy giants like Uber challenging public transportation systems), which means they challenge the very idea of value that should be measured.[36]

- The attempt to justify and describe the distinctive value created by government and public sector actors has a long intellectual pedigree. Terms such as 'public interest' and 'common' or 'public good' have been a focus for political philosophers going back to Aristotle. More recently, the concept of 'public value' has been developed by scholars in

[30] See also Peters et al 2018, op. cit.
[31] Wolfgang Drechsler (2005), The rise and demise of the new public management, *Post-autistic Economics Review* 33, pp. 17-28.
[32] Rainer Kattel and Mariana Mazzucato (2018), Mission-oriented innovation policy and dynamic capabilities in the public sector. *Industrial and Corporate Change*, 27, 5, pp. 787-801: https://doi.org/10.1093/icc/dty032. See also Peters et al 2018, op. cit.
[33] Geert Bouckaert and B. Guy Peters (2002), Performance measurement and management, *Public Performance & Management Review*, 25:4, 359-362: https://doi.org/10.1080/15309576.2002.11643672.
[34] Edler et al 2016, op. cit.
[35] Michael Lewis (2004), *Moneyball: the Art of Winning an Unfair Game*, Norton.
[36] For a detailed discussion of various approaches to evaluating innovation in government, see OECD Observatory for Public Sector Innovation (2019), Evaluating public sector innovation: Support or hindrance to innovation?: https://oecd-opsi.org/wp-content/uploads/2019/05/Evaluating-Public-Sector-Innovation-Part-5a-of-Lifecycle-Report.pdf.

the public management and administration field. This concept of public value emphasises how public managers have an important role in mediating between the need for efficient services and the engagement of citizens in developing public services and policy.[37]

- We propose an alternative approach, which begins with the notion of public value as collectively generated by a range of stakeholders, including the market, the state and civil society. Key here is the emphasis on value creation at the core: not 'public' value but value itself, with a clear delineation of the role of the different actors that are central to its formation. We argue that public value is value that is created collectively for a public purpose. This requires understanding of how public institutions can engage citizens in defining purpose (participatory structures); nurture organisational capabilities and capacity to shape new opportunities (organisational competencies); dynamically assess the value created (dynamic evaluation); and ensure that societal value is distributed equitably (inclusive growth).[38]
- We are specifically interested in the study of how various policy appraisal and evaluation frameworks and methodologies reflect the idea of value, and whether they focus on private value and public goods or build on some form of public value.

[37] For further detailed discussion, see Mariana Mazzucato and Josh Ryan-Collins (2019), Putting value creation back into 'public value': From market-fixing to market-shaping, UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-05): https://www.ucl.ac.uk/bartlett/public-purpose/wp2019-05.
[38] Mariana Mazzucato and Rainer Kattel (2019), Getting serious about value, UCL Institute for Innovation and Public Purpose, Policy Brief 07: https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/iipp_policybrief_07_getting_serious_about_value.pdf.

Aims and methodology of the study

- The aim of this project is to produce a review of international examples where alternative appraisal and evaluation methods have been applied to develop, appraise, monitor and evaluate policy. Our specific focus is on appraisal and evaluation practices, and on assessing the impact of regulatory practices on innovation.
- The way we have approached the extensive literature on policy appraisal and evaluation in this research has been to identify key hotspots via a literature review using keywords, statistical analysis and mapping against our public value/private value, market-shaping/market-fixing framework. First, this allows us to offer some overall insights on state-of-the-art practices in the international arena and to comment on trends identified. Second, we take a deep dive into specific examples of appraisal and evaluation methodologies to understand how the Better Regulation Executive can learn from, be inspired by, and become further engaged in, examples from around the world.
- **Literature review:** This study is based on an extensive systematic literature review of state-of-the-art innovation practices over the past five years up to 2019. The review covers academic publications, including both mainstream and alternative economics and policy journals; specialised policy analysis journals; and selected international policy cases, including from the European Commission. See Appendices 1-3 for details of our methodology.
- **Survey of the UCL IIPP Mission Oriented Innovation Network (MOIN):** IIPP hosts the Mission Oriented Innovation Network (MOIN), which is discussed in greater detail below. A workshop on alternative evaluation methods was held at the annual meeting of MOIN in Bellagio, Italy, in March 2019, and this contributed to the initial review of alternative evaluation methods. In addition, a survey of the public sector, innovation-oriented members of the MOIN was also conducted to contribute to the review.
- **Keyword search:** We identified and utilised keywords related to policy appraisal and evaluation to search several academic databases for papers and articles related to these topics. This followed the PRISMA flow approach, as outlined in Appendix 2. We then coded and analysed these papers and articles, as well as the selected international policy cases and the survey results, using the framework map shown below.
- **Policy appraisal/evaluation map:** In order to map the existing policy analysis frameworks and tools, the study divided policy evaluation and appraisal frameworks according to whether they focus on market-fixing or market-shaping (or have no market orientation), and whether the underlying value concept is public or private. The study differentiated between policy sector (e.g. education policy), policy programme (e.g. higher education) and policy instrument (e.g. studentships for STEM students) as the level in which specific methodologies are applied.

Mapping policy evaluation and appraisal frameworks

- To map the literature, we started by codifying our own taxonomy and characteristics table for different policy evaluation frameworks with a matrix mapping market-shaping against market-fixing, and private value against public value.
- Table 2 summarises our innovative exploratory taxonomy, which has been developed specifically for this study. We expect to develop future iterations over time, based on co-creation and further research.

Table 2: Policy frameworks and methods mapping methodology

	Public value	Private value
Market-shaping	Policies and interventions with a strategic view to steer economic development to solve politically recognised problems (and not only provide public goods). Markets are recognised as always shaped and constituted by collective action, thus giving policy more options to shape future market and technological outcomes. Examples of policies: - Industrial policies aiming to increase development along the desired direction of travel, including pushing certain technology frontiers - Policies seeking to detect unrealised macroeconomic multipliers by assessment Typical goals: - Mission-dependent technological innovations - Equity and sustainability considerations during anticipated system evolution - Positive spillovers into non-monetary areas of the system	Policies and interventions with a strategic view to maximise benefits from future market opportunities. The technological direction is taken as exogenous (e.g. provided by market forces or through public good-type investment into basic science), but does not consider who will benefit the most or when gains are realised. Examples of policies: - Industrial policies aiming to increase 'value added' either by: 1. achieving first-mover advantages in sunrise-industries (e.g. race for quantum computing) or 2. assisting companies to climb global value chains (e.g. 'the developmental state') - Diplomatic efforts to obtain better trade relations with future importers Typical goals: - Productivity - Competitiveness - Income growth - Exports - Employment - Financial stability
Market-fixing	Policies designed to achieve Pareto-optimal utility when markets are incapable of delivering socially acceptable outcomes. This involves public provisioning of 'public goods' and regulation, when diffuse externalities	Policies designed to allow the market system to function as perfectly as possible and achieve Pareto-optimality by removing distortions of the price-mechanism or forcing market participants

	render markets unfit to incorporate the entire 'social value' of the good or service. Examples of policies: - Public provisioning of services (public goods) - Taxes and fees to shape behaviour and internalise diffuse externalities Examples of goals: - Satisfying rare needs that cannot be met by markets - Sustainable use of common pool resources	to internalise 'externalities' on recognisable citizens. Examples of policies: - Deregulatory policies - Environmental regulation - Lowering of taxes - Improved incentives - Bureaucratic efficiency measures Examples of goals: - Efficiency - Tax revenue/saved expenditures - Income - Unemployment
Non-market-oriented	Policies targeting value creation that is broadly diffused in realms of the social system that are largely independent of market forces. Examples of policies: Regulations and initiatives aiming to influence cultural values and citizens' mutual behaviour Examples of goals: - Social cohesiveness - Tolerance - Safety	Policies aiming to improve the wellbeing of specific citizens in societal spheres that are distant from market exchange. Examples of policies: - Family support schemes (non-monetary) - User experience enhancement - New rights Examples of goals: - Civil liberties - Mental health

Results

Overall results

- It should be noted that of the literature identified through our keyword search and PRISMA method, only 43% was considered eligible for this study. This was for many reasons: some literature did not mention evaluation methods, some was focused on hypothetical interventions, and some was too vague in its use of the key terms to be properly categorised.
- Of all literature and policy case studies surveyed, 35% referred to market-fixing activities, while 46% referred to market-shaping activities, and 24% referred to non-market-oriented activity.
- 18% of the policy evaluation and appraisal examples were identified as market-fixing, based on private value conception. Close behind, 17% of the market-fixing examples were for public value orientation. Conversely, only 6% of all the policies were considered to be market-shaping and private value, while 36% of all policies – the largest category –

were both market-shaping and public value. This gives a sense of our selection bias (see in Appendix 1 for the discussion) in the policy cases considered, and approaches being used by policy-makers, as well as how different policy areas adopt mainstream or alternative evaluation techniques.

- Across all items listed in our review **we were able to identify 80 different policy evaluation and appraisal methodologies** (that figure doesn't include duplicates and near-duplicates, which were removed). These ranged from very traditional – cost-benefit analysis, cost-efficiency analysis, general equilibrium modelling and semi-structured or elite interviews (in the form, for example, of commissions) to much less orthodox approaches – social fabric matrices, asset mapping and public value mapping.

- Methodologies also ranged from being highly technical in description, outlining specific granularity levels of variables, such as panel vector regression, to being more vague or high-level, as with several descriptions of landscape analysis and bottom-up evidence approaches. This is a reminder that not all policy appraisers and evaluators will employ the same methodologies in the same ways; and that their definitions may have dissimilarities. Creating a common language is key when discussing mainstream vs alternative approaches.

- Both market-fixing and market-shaping approaches use a multitude of analytical tools and methods. Indeed, there is some overlap (e.g. general equilibrium model) which might suggest that even approaches that attempt to study market-shaping aspects of policies rely on concepts of evidence and change that derive from market-fixing/market failure frameworks.

Discussion of academic literature

- Figures 2-8 show the results and range of methods being used in the **academic literature**. The results of this mapping show that around half of the evaluation practices studied are based on market-fixing frameworks, with studies focusing on market-shaping making up a little over one quarter of the sample. Almost two-thirds of the academic studies surveyed here focus on policy instruments or policy programmes (rather than on policy sectors), as outlined in Figures 2 and 3 below.

- As expected, a larger percentage of the market-fixing activity was discernible in more orthodox publications, where 55% of activities were market-fixing – in heterodox journals, the proportion was 44% market-fixing activities and 44% market-shaping. In specialised evaluation literature, 82% of all articles were on public value-led policies, compared to 36% in heterodox and 43% in orthodox journals. This implies that alternative evaluation methods are being taken seriously by the academic and specialist community.

- As anticipated, market-fixing frameworks are heavily reliant on quantitative models; analytical methods for markets shaping are, on the other hand, much more diverse, ranging from quantitative approaches such as agent-based modelling to qualitative approaches such as living labs. Figures 4-6 below present the full array of methods used. Word clouds were selected as a quick and easy way to grasp the more prominent methods utilised. The more often the method was mentioned in the literature, the larger and bolder the word appears. The larger the font of a particular method indicates how often it was used in that body of a particular market orientation. Colours are used simply for visual effect and clarity. Methods in the word cloud appeared at least twice, whereas

the list beneath the word cloud shows other methods that were mentioned only once, but are listed to provide a sense of the breadth of methods.

Figure 2: Breakdown of academic literature based on market orientation

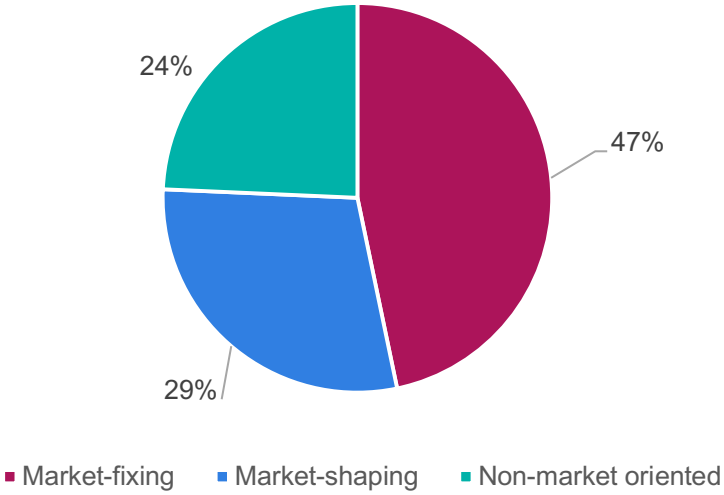


Figure 3: Breakdown of academic literature according to policy level

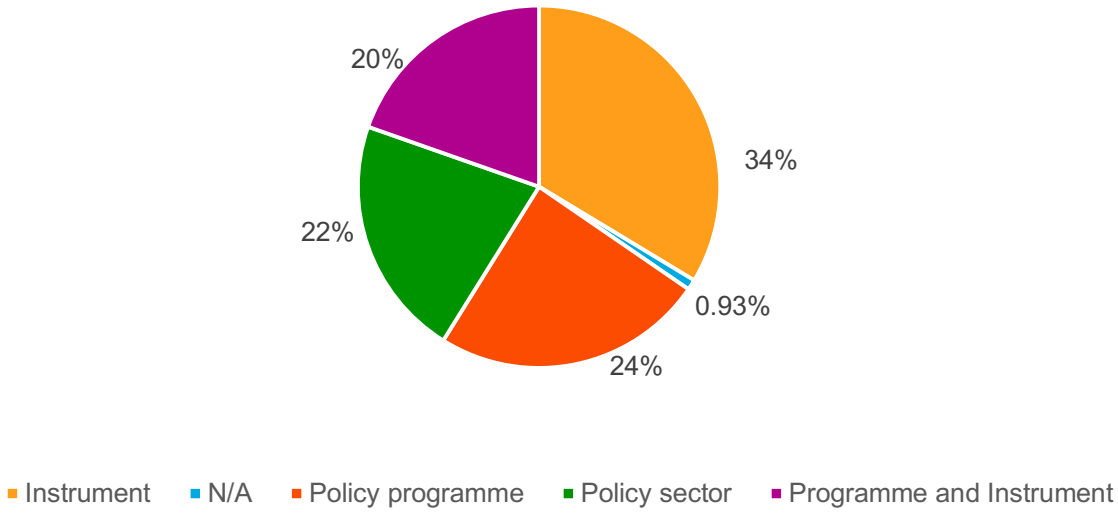
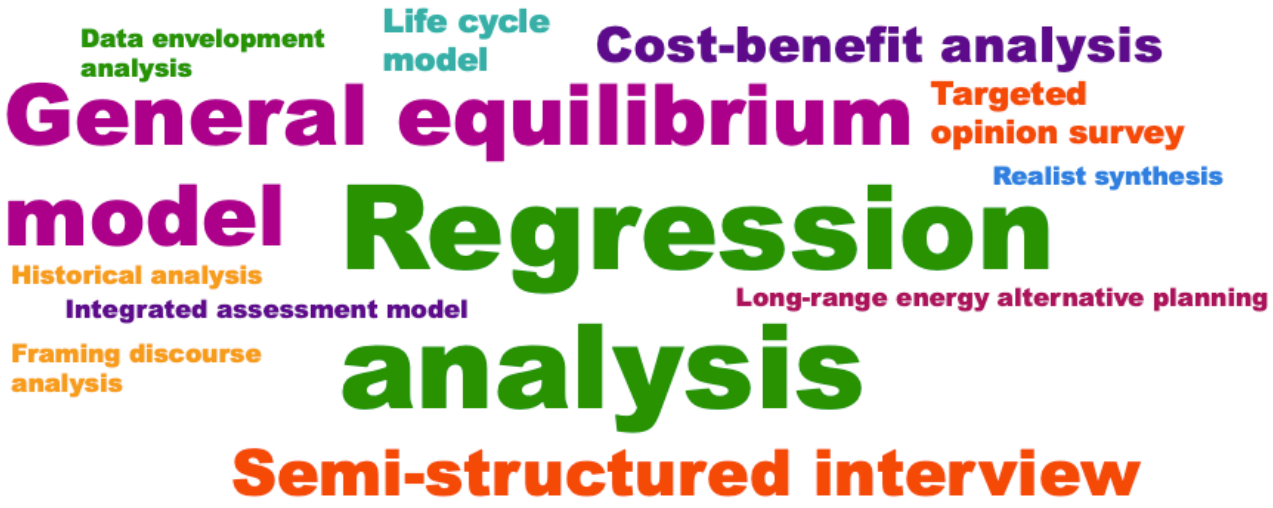


Figure 4: Methodologies deployed in market-fixing frameworks (with weighting)



Methods not included in the word cloud with occurrences of only one: Bibliometrical analysis, Broad opinion survey, Causal mediation analysis, Community-based participatory research, document review, Doubly randomized preference trial, Elite opinion survey, Energy assessment, Game theory, Geographic Information System analysis, Gravity model of trade, Panel vector autoregression (PVAR), Partial equilibrium modelling, Participant observation, Propensity-score matching (PSM), Qualitative comparative analysis (QCA), Quality and Impact of Component Evidence Assessment, Regression discontinuity design (RDD), Social cost accounting, theory-based evaluation

Figure 5: Methodologies deployed in market-shaping frameworks (with weighting)

Methods not included in the word cloud with occurrences of only one: Agent based modelling, Asset mapping, Bivariate analysis, Broad opinion survey, Contribution analysis, Co-production, Cost-efficiency analysis, document review, Full Bayesian before-after-evaluation, Historical analysis, IMPLAN (input-output model), Life-cycle model, Living labs, logic models, network analysis, Normative criteria rating, Public value mapping, Simple statistics, Stock-flow-consistent model (SFC), Surplus value analysis, systematic review

Figure 6: Methodologies deployed in non-market-oriented frameworks (with weighting)

Methods not included in the word cloud with occurrences of only one: Balanced sampling, Balanced scorecard framework, Broad opinion survey, Concept mapping, Document review, Doubly randomized preference trial (DRPT), Framing/discourse analysis, Full Bayesian before-after-evaluation, Historical analysis, Logic models, Meta-analysis, Multiple measures of effectiveness (MoE), Quasi-experiments, Randomized Control trial, Simple statistics

Examples of alternative methodologies in academic literature

The academic literature was very granular and allowed close definition of each methodology. We looked in detail at three methodologies identified in our research: social fabric matrices; living labs; and public value mapping. Here we define how the methodology is different from dominant and mainstream techniques, and the outcomes or implications of this.

Social fabric matrices

The social fabric matrix (SFM) is an analytical framework that accounts for flow relationships between actors, institutions and technologies in a socio-economic setting through time. A variety of flows can be incorporated into the matrix, including concepts that are difficult to measure quantitatively, such as technology, beliefs, environmental sustainability and social influence.

Suitable market orientation: Market-shaping

Suitable value assessment (public / private): Public value

Suitable use in ROAMEF cycle: Objectives, Appraisal, Evaluation

Outcomes: The flow matrix enables a holistic, yet actor-focused, analysis with incorporation of feedbacks and synergies in the system. The flow-accounting framework provides decision-makers with a tool to match the evolutionary development of reality and iteratively develop asset maps, multi-criteria evaluations and strategic planning for the system.[39] In contrast to single-metric cost-benefit analyses that seek to estimate an optimum policy, the iterative nature of SFM is well suited to deal with the reality of evolving technologies and changing socio-economic environments. In the context of socio-economic development policy, Underwood and

[39] F. Gregory Hayden (2006), *Policymaking for a Good Society: The Social Fabric Matrix Approach to Policy Analysis and Program Evaluation,* Springer.

Friesner (2017)[40] deploy the SFM framework to add evolutionary and complex systems thinking to the more widely used asset-mapping and economic impact techniques that are more static in nature.[41] SFM has also been used for more technical economic issues such as public school finance[42] and monetary policy.[43]

Implications of applying this method: Policy analysts obtain a formalised tool to account for the complexities of the governed system in question. Policy goals will more likely be achieved without unanticipated outcomes by embracing and mapping the complexity of the system's interrelations, incentive structures, beliefs and technologies. This is different to the more frequently used systems mapping, as the social flow matrix is first and foremost a matrix-based method, while systems mapping takes a graphical approach.

Living labs

Living labs is a user-centric approach to innovation situated in a real-life environment to enhance the validity of the findings. The philosophy is that if new solutions are co-created between the end users, producers, researchers and utilisers of the knowledge in the user's environment, the innovations are more likely to be successfully introduced to broader use without unforeseen obstacles. Living labs are therefore (temporary) organisational structures where the co-creation takes place. Van Geenheuizen (2018) categorises three types of living labs based on degrees of actor-complexity: person-oriented, organisation-oriented and multi-activity labs (often utilising university campuses as mini cities).[44]

The collaborative learning process depends on the actors sharing attitudes, commitments and a common language. For instance, a person-oriented study of an old person, whose home is turned into a living lab to test digital health care technologies, most likely involves fewer actors and a more predictable environment than on-site innovation in public transport or at care facilities. Likewise, university campuses can emulate the complexity of a city and provide testing grounds for scalable technologies and policies. Labs can be further differentiated based on the types of innovation pursued, the approach to user involvement, the environment chosen, the range of included actors and the scale of the lab.

Suitable market orientation: Market-shaping

Suitable value assessment (public / private): Public and private value

Suitable use in ROAMEF cycle: Monitoring, Evaluation and Feedback

Outcomes: A key difference between living labs and traditional approaches, including cost-benefit analysis, is the attitude of 'curiosity' that characterises the process and the expectations of the solution. The social learning process takes place through improvisations and experimentations, which requires that policy-makers and researchers accept unpredictability.

[40] Scott Fullwiler (2009), The Social Fabric Matrix Approach to Central Bank Operations: An Application to the Federal Reserve and the Recent Financial Crisis, in Tara Natarajan, Wolfram Elsner and Scott Fullwiler 2009, *Institutional Analysis and Praxis: The Social Fabric Matrix Approach*, Springer US, pp. 123-69.
[41] Ibid.
[42] Jerry L Hoffman (2009), Application of the Ithink ® System Dynamics Software Program to the Social Fabric Matrix to Analyze Public School Finance Systems, in Tara Natarajan, Wolfram Elsner and Scott Fullwiler 2009, *Institutional Analysis and Praxis: The Social Fabric Matrix Approach*, Springer US, pp. 237-53.
[43] Scott Fullwiler (2009), The Social Fabric Matrix Approach to Central Bank Operations: An Application to the Federal Reserve and the Recent Financial Crisis, in Tara Natarajan, Wolfram Elsner and Scott Fullwiler 2009, *Institutional Analysis and Praxis: The Social Fabric Matrix Approach*, Springer US, pp. 123-69.
[44] Marina Van Geenhuizen (2018), A Framework for the Evaluation of Living Labs as Boundary Spanners in Innovation, *Environment and Planning C: Politics and Space* 36.7, pp. 1280-298.

Experimentation and trials must be conducted in real-life settings, which increases the validity of the appraisal. Moreover, the living labs approach encourages systems thinking by involving the broadest possiblrme network around the end-users. It is likely that practitioners will encounter unintended outcomes from interactions with the real-life environment.

Implications of applying this method: By welcoming uncertainty and emphasising user experiences, policy analysts increase the potential for reality-proofing their recommendations to policy-makers. The applicability is limited for large-scale capital investments, e.g. in infrastructure. For example, designing a living lab approach to appraise or evaluate a high-speed rail link would require a very different approach to more socially oriented policy.

Public value mapping

Public value mapping is a framework developed to assess the contribution towards public values made by policies and both private and public organisations.[45] In the public value mapping discipline, public value has been defined as "the prerogatives, normative standards, social supports, rights and procedural guarantees that a given society aspires to provide to all citizens"[46] (note, this differs from our discussion of public value above). These more abstract, normative and culturally contingent aspirations can provide an extended to foundation to reflect on prevailing outcomes and lead to recommendations of new ways to regulate economic activity.

Suitable market orientation: Market-shaping

Suitable value assessment (public / private): Public value

Suitable use in ROAMEF cycle: Rationale, Objectives, Appraisal

Outcomes: Public value mapping appraisal and evaluation approaches are distinguished from traditional welfare economics, which are centred on the utility realised by individuals. The strength of the latter approach relates to consumption of concrete tangibles, but leaves much to be desired in relation to more abstract concepts, like social justice and collectivism, which are crucial to public policy-making. Public value mapping allows systematic assessment of these more challenging evaluation dimensions.

Implications of applying this method: Public value mapping leads to a focus on purposefulness and mission in policy design. By explicitly mapping out the contributions of a policy or organisation towards shared societal goals, overarching ambitions are in less risk of being crowded-out by technical and marginal tasks. Moreover, the public value basis of analysis can also be applied to private sector activity, indicating that, with wider take-up, there could be comparative evaluations made between the two. For example, Anderson and Taggart (2016) use public value mapping to argue that public value failure takes place in the US for-profit higher education sector.[47] However, while 'public value' has been defined within the academic literature, there is a risk of policy-makers developing their own idiosyncratic or even politically motivated definitions of what public value means, thus biasing the approach.

[45] Barry Bozeman (2002), Public-Value Failure: When Efficient Markets May Not Do, *Public Administration Review* 62 (2), pp. 145-161; Barry Bozeman (2007), *Public Values and Public Interest: Counterbalancing Economic Individualism*, Georgetown University Press.
[46] Barry Bozeman and Daniel Sarewitz (2005), Public Values and Public Failure in US Science Policy, *Science and Public Policy* 32.2, pp. 119-36.
[47] Derrick M. Anderson and Gabel Taggart (2016), Organizations, Policies, and the Roots of Public Value Failure: The Case of For-Profit Higher Education, *Public Administration Review* 76.5, pp. 779-89.

Discussion of policy case studies

Our policy case studies are taken from two key sources: the **European Commission**'s repository of mission-oriented policies and the **Centre for Policy Impact**'s extensive case study compendium (for further details, see Appendix 1).

- In contrast to academic literature, our sample of policy case studies is dominated by cases focusing on market-shaping, and related evaluations metrics and methodologies. This could be the result of a sample bias as the policy case studies were sourced from two main platforms. The platforms mainly highlighted and included cases focused on policy that would create systemic change. However, focus on policy cases and evaluation practices that evaluate a larger system and are also market-shaping may also reflect a more recent trend towards far more comprehensive and innovative policy initiatives whereby cases have increased in size, scope and scale.
- Figures 7-11 present the range the methods being used in the policy studies. Policy cases show a strong reliance on simple statistics, general forms of monitoring and evaluation (M&E) and usage of key performance indicators (KPIs). The predominant usage of the term 'M&E' demonstrates how precise policy evaluation accountability and transparency is often lacking in the non-academic policy literature.
- Policy-specific literature was far more focused on public value than private value with 95% of examples categorised under public value and 57% of all examples characterised as both market-shaping and linked to public value.
- The policy case studies were mainly oriented towards specific policy programmes. Policy sectors and evaluative instruments were not the main focal points within the policy case studies examined. Most of the policy programmes tended to be focused on benefiting the public domain through efforts to alter markets in healthcare, financing, transportation, labour and several other categories.
- When it came to evaluating the policy programmes, the focus of the policy case studies differed from the academic literature. Instead of trying out and using a vast array of nuanced evaluative techniques suited to a particular policy field or instrument that best measured a desired outcome, the evaluative techniques of policy practitioners were based on a set of relatively crude and general monitoring and assessment instruments. Simple statistics, monitoring and evaluation, key performance indicators (KPIs), regression analysis, cost-benefit analysis and a variety of focus groups or feedbacks and surveys were some of the most frequently recurring methods.

Figure 7: Breakdown of policy case studies based on market orientation.

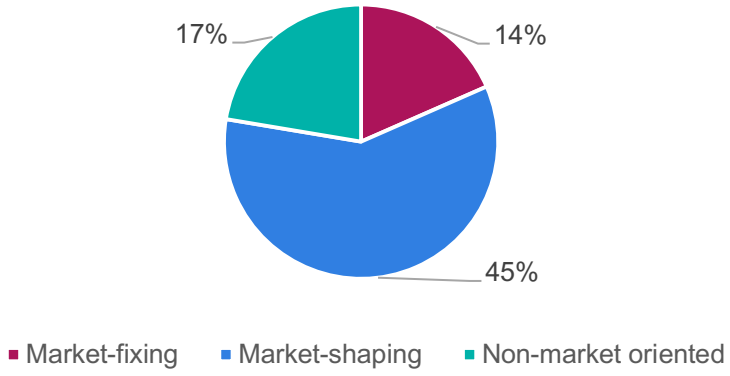


Figure 8: Breakdown of policy case studies based on market orientation.

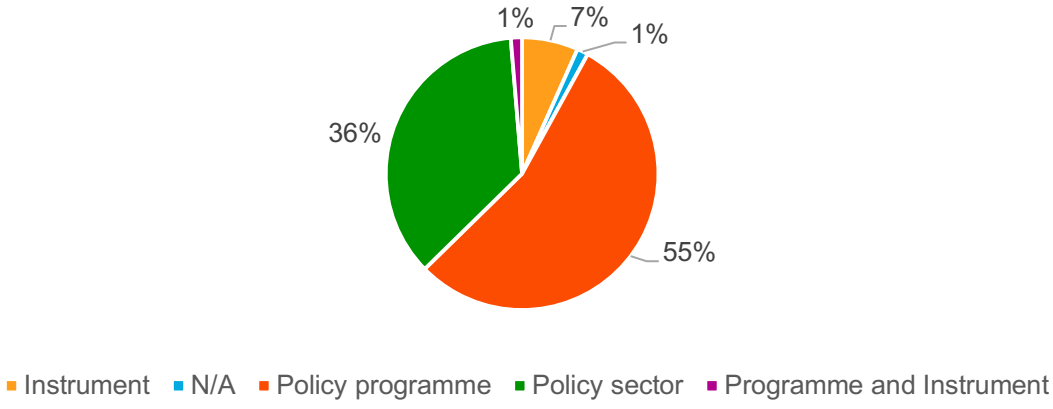


Figure 9: Methodologies deployed in market-fixing frameworks (with weighting)

Methods not included in the word cloud with occurrences of only one: Broad opinion survey, Case studies, Cost-efficiency analysis, donor monitoring, European Commission Better Regulation framework, feasibility study, Geographic Information System analysis (GIS), Logic models, OECD criteria, Process tracing, Programme effectiveness modelling, Realist synthesis, Regression analysis, Technological progress/ innovation analysis, Theory-based evaluation

Figure 10: Methodologies deployed in market-shaping frameworks (with weighting)

Methods not included in the word cloud with occurrences of only one: Econometric analysis, feasibility study, Focus groups, Framing/discourse analysis, Historical analysis, Levelised cost of energy, Life-cycle model, Multiple measures of effectiveness (MoE), Net present value approach, Process tracing, Public value mapping, Quasi-experiments, Social cost accounting, Technological progress/ innovation analysis

Figure 11: Methodologies deployed in non-market-shaping frameworks (with weighting)

Methods not included in the word cloud with occurrences of only one: Broad opinion survey, Cost-benefit analysis, distribution models, Focus groups, IMPLAN or other input-output model, Multiple measures of effectiveness (MoE), Randomized Control trial, Technological progress/ innovation analysis

Examples of alternative methodologies from policy case studies: three pillars of evaluation

When analysing the policy case studies we found both that multiple, simple or high-level methodologies were frequently applied, and that methodologies tended to fall into one of three categories: **stakeholder perspective, standardised policy output or process dimensions** (summarised in Table 3).

Such a combination of methodological tools as an approach to appraise and evaluate policy can be seen as efficiently beneficial, but also restrictive. The methodologies described are highly accessible and link closely to processes applied in the private sector, making them more comprehensive to policy-makers from multiple backgrounds and ensuring interoperability with business stakeholders. However, the use of language is far more imprecise than in the academic literature, leading to some confusion over the individual impact of each methodology.

Stakeholder perspectives were measured through a variety of surveys, interviews and group discussions. Information was ascertained primarily from the stakeholders involved in the dissemination of the particular policy programme, and very rarely were interviews and surveys held which focused on the affected population (which would have helped in providing insight into the policy outcome).

Standardised approaches included measuring the overall success of a programme by focusing on simple statistics, key performance indicators (KPIs) and regression analyses outputs from available data. Use of numeric data occurred only in policy programmes where there was a quantifiable output measurement, such as number of people vaccinated, money invested or change in cyclists in a given area.

Methodologies which were **process-oriented** included monitoring and evaluation, logic models, and forms of historical and document analyses. Monitoring and evaluation frameworks were often cited as an essential policy evaluation methodology. However; the specific methods and evaluation processes used were rarely explicitly documented or discussed.

Table 3: Alternative methodologies from policy case studies

Evaluation category	Methods used	Appraisal and evaluation logic
Stakeholder perspective	• Targeted opinion survey • Group interview • Semi-structured interview • Focus group • Broad opinion survey • Elite opinion survey	Places emphasis on understanding and interpretation of how well policy programming is performing from an individual or group perspective; open to more interpretation, bias and human nuance. Potentially provides greater insight into future areas of improvement as positive and negative policy dynamics are heightened.
Standardised approaches	• Regression analysis • Simple statistics • Cost-benefit analysis • Key performance indicators (KPIs) • Randomised control trials (RCTs)	Helps to standardise policy output measurement, believed to provide ‘clear’ and ‘unbiased’ evidence for budgetary purposes. Easier to compare success of policy programme to others.
Process-oriented	• Monitoring and evaluation (M&E) • Logic models • Historical analysis • Document review • Process tracing	Takes into account the policy programme process. Often constructed in a historical narrative of policy action and policy output or process steps. Attempts to incorporate the policy context and implementation dimensions. Often, though, process forms can be suited to best show off the benefits of the programme, such as is the case with vague M&E frameworks.

Overall discussion

- Most public organisations operate in a context of multi-dimensional appraisal and evaluation in which frameworks and tools originate from various levels of governance (e.g. supranational guidelines created and maintained by the European Commission and the OECD, in addition to national level guidelines). In addition, public organisations do not rely only on external data and evaluations, but are increasingly creating in-house datasets and evaluation tools (e.g. dashboards, surveys, focus groups).

- The emergence and rise of digital and design agencies has led to an influx of diversity of new evaluation methodologies around user experience and service design. These methodologies avoid the idea of aggregated indicators or averages as measurements and focus on mapping a multitude of contextual as well as individualised factors. These agencies are also pioneering real-time monitoring practices such as dashboards[48] and are challenging 'theatre of evaluation' that relies on single indicators or complex modelling.

- Public organisations are increasingly aware of the need to combine retrospective research and evaluation methodologies with prospective methods which emerge from design-oriented research and evaluation activities.[49] Thus, policy evaluation is as much about understanding past actions as it is about shaping future policy frames and expectations.

- Public organisations are increasingly concerned about creating reflexive capabilities,[50] particularly in relation to how these organisations create, maintain and evaluate public value. There is a clear need to significantly deepen our understanding of various public value frames (e.g. economic, legal, organisational, etc) that can be used in evaluation practices.[51] Reflexive capabilities connect evaluation and learning practices within organisations.[52]

- While alternative approaches to appraisal and evaluation are increasingly being developed by innovative agencies, and are more frequently seen in orthodox academic literature, they are far from standard tools for policy-makers. Our policy case studies indicated that, while not averse to new, public-value-oriented evaluation approaches, practitioners often prefer to employ multiple mainstream methodologies. **This offers diversification and triangulation, but is not an ambitious leap in terms of alternative approaches.**

- Current practices could benefit from up-taking alternative policy evaluation and policy appraisal approaches. For instance, one pragmatic change would be for the Green Book and Magenta Book to include in their annexes the identified typologies of appraisal and

[48] See, for instance, Estonian e-residency programme dashboard: https://e-resident.gov.ee/dashboard/
[49] See also A. Georges L. Romme and Albert Meijer (2019), Applying design science in public policy and administration research, *Policy & Politics*: https://research.tue.nl/en/publications/applying-design-science-in-public-policy-and-administration-resea.
[50] See also Arie Rip (2006), A co-evolutionary approach to reflexive governance – and its ironies, in J.-P. Voss, D. Bauknecht and R. Kemp, eds., *Reflexive Governance for Sustainable Development*, Elgar; and Ralf Lindner, Stephanie Daimer, Bernd Beckert, Nils Heyen, Jonathan Koehler, Benjamin Teufel, Philine Warnke and Sven Wydra (2016), Addressing directionality: Orientation failure and the systems of innovation heuristic. Towards reflexive governance, Fraunhofer ISI Discussion Papers, *Innovation Systems and Policy Analysis* No. 52.
[51] Tina Nabatchi (2018), Public Values Frames in Administration and Governance, *Perspectives on Public Management and Governance*, 1, 1, pp. 59–72: https://doi.org/10.1093/ppmgov/gvx009.
[52] OECD Observatory for Public Sector Innovation (2019), op. cit.

evaluation (for instance the categories of economic evaluation, theory evaluation, empirical impact evaluation and process evaluation), and to use a research-informed practice model and incorporate some of the known innovative policy assessment techniques and their uses.

- This document advocates both for a wider user of alternative evaluation and appraisal methodologies, as well as a synthesis of a mixture of different techniques. We recognise that CBA and CEA will not be abandoned, nor should they be eliminated in their entirety, but rather they should be complemented by other measurements which embrace value that is difficult to monetise, such as wellbeing or dynamic processes. For instance, the Green Book claims to focus on improving social welfare or wellbeing, or social value, but since this metric is currently based on standard economic evaluation techniques, it can only capture social value to a limited degree. Complementary tools must be integrated as standard procedure to provide a fuller picture for decision-makers.

- Further, policy appraisal and evaluation should be held to a higher standard than a do-nothing approach or business as usual. This framework inherently sets the standard fairly low in terms of policy outcomes, as they will only focus on a singular parameter changing and will only seek a marginal improvement. Business as usual does not foster an innovative policy landscape when evaluation and appraisal toolkits merely measuring the status quo. Also, business as usual is a false premise as the contextual policy environment is always in flux, and thus the aim should be for appraisals and evaluations to best capture the changing environment and further advocate for active market-shaping as markets are inherently being altered either way.

- Policy evaluation is an iterative process. Thus, when thinking about the ROAMEF cycle, one can cannot only put evaluation at the end of the process. An integrated approach would consider how appraisal and evaluation can occur throughout the ROAMEF cycle, and that some methods are better suited to implementation at different stages. "Evaluations can, in fact, occur at practically any other time. And importantly, decisions affecting and relating to any evaluation will almost always be taken earlier in the policy process."[53] Good practice should ensure that there are forward-thinking, innovative and adaptive ways of considering evaluation methodology type and implementation.

- As part of IIPP's suggestions, not only do the policy appraisal and policy evaluation frameworks have to be altered, but there should also be consideration, evaluation and appraisal of the regulations that are implemented. Policy is only one tool that can create change for the betterment of the economic, public and social sphere. The Better Regulation Framework[54] is specifically focused on regulations that have the greatest potential impact on business and civil society organisations. Evidence, transparency and proportionality are quoted as being essential to the formation of appropriate regulatory mechanisms.

- Regulatory provision in relation to business activity is a statutory provision which imposes or amends requirements, restrictions or conditions; or sets or amends guidance in relation to the activity or the securing of compliance with, or enforcement of, requirements, restrictions, conditions, standards or guidance for the activity. Similar to the general strategies for policy evaluation and policy appraisal, the Better Regulation Framework is also based on the ROAMEF cycle and uses cost-benefit analysis. As

[53] HM Treasury 2011.
[54] The Better Regulation Framework: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/735587/better-regulation-framework-guidance-2018.pdf.

stipulated, the principles of cost-benefit analysis should be used in the development and appraisal of all policies, programmes and projects. Specifically, the indicators are the Net Present Social Value for society and the economy as a whole and the Net Present Value to Business. The business impact target is framed with an austerity angle since regulations are geared towards the overarching goal of saving £9 billion. This is clearly a private value-oriented approach, but it is pivotal that such frameworks to analyse regulatory impact are complemented by public value and market-shaping frameworks and tools to enable better decision-making.

- Overall, though, the existence and uptake of challenge-driven policy frameworks make it clear that the nature of evidence in policy-making is changing. First, rather than focusing on single instruments, there is increasing consensus that we should analyse and evaluate policy mixes, rather than individual initiatives, as the impact of various instruments is rarely linear and can indeed be contradictory, with significant time lags and spillover effects.[55] Second, policy-making takes places in a multiplicity of feedback channels that generate diverse evidence and perspectives on evidence. That is, policy evaluation should take account of the various contested values inherent to policy-making.

Conclusion

In summary, this exploratory study shows that there is a growth in alternative methodologies focusing on market-shaping and that this should help policy-makers in designing and evaluating challenge-driven policies. This report has identified a list of promising methodologies that public bodies could start incorporating into the ROAMEF cycle, as deemed relevant in a particular organisation. A greater understanding of the particular methods and how they are used in conjunction with particular market framing is ongoing. However, based on the initial data there is a clear need for reflexive capabilities, public value frameworks and tools to evaluate complex policy mixes. Further, public organisations need to dedicate significantly more resources to experimenting with, and learning from, new policy evaluation methodologies and tools.

[55] Edler et al (2016), op. cit.

Appendix 1. Literature review methodology in detail: Data selection process

Data sources

Academic sources

As academic publishing is quite strongly sectarian, in particular in economics, the literature review needed to cover diverse sources of academic publishing. Thus, articles were collected from the following sources:

- Mainstream (orthodox) academic journals (Web of Science database)
- Heterodox academic journals (according to Lee et al 2010 ranking[56]
- Specialised academic journals with focus on policy evaluation

Only English material was included in the literature review.

Orthodox articles were sourced from four journal categories: Economics, Political Science, Public Administration, Social Sciences Interdisciplinary.

Heterodox articles were sourced from journals rated as heterodox by Lee et al 2010. Frederic Lee was considered a convening figure in heterodox economics and the list contains a ranking of impact that is used in the screening below.

Specialised journals were drawn from a list of journals designated as evaluation journals by the European Evaluation Society (EES). Only non-field-specific journals from the list were included, while two journals were added based on content judgement (*Evaluation and Programme Planning, New Directions for Evaluation*).

Policy literature

As policy literature does not have a coherent database, we constructed our own literature set. To reflect various viewpoints on evaluation, we included the following literatures:

- The Centre for Public Impact has created a public impact observatory[57] of policy cases with a specific focus on impact and on understanding impact. Their cases (more than 300 altogether) include specific discussion of evidence and measurement, and all cases are graded on whether evidence and other qualifications are well presented in the specific case. We chose only cases with strong evidence and measurement qualifications.[58] This is a highly diverse database that provides for wide coverage.

[56] Frederic S. Lee, Bruce Cronin, Scott McConnell and Erik Dean (2010), Research Quality Rankings of Heterodox Economic Journals in a Contested Discipline. *American Journal of Economics and Sociology*, 69, pp. 1409-1452: https://doi.org/10.1111/j.1536-7150.2010.00751.x.
[57] The database is available here: https://www.centreforpublicimpact.org/observatory/.
[58] See here: https://www.centreforpublicimpact.org/observatory/?searchString=&evidence%5B%5D=4&measurement%5B%5D=4&orderby=dateDesc.

- The European Union has put mission-oriented policies at the heart of its next R&D strategy[59] and as part of the preparations for the new strategy, the European Commission has collated case studies of mission-oriented policies in Europe. Each case includes a specific discussion of impact measurement. There are 17 cases altogether.[60] These cases specifically discuss challenge-driven policies.
- There were eight responses to the MOIN survey. MOIN includes diverse organisations across the world that are all interested in, and engaged with, challenge-driven policies. The survey is in Appendix 3.

Data search

In the Web of Science database, the following search string has been applied to titles and abstracts:

> (methodology* OR framework*) AND Policy AND ((evaluation OR appraisal OR impact) OR (regulat*AND (impact OR analysis)))

The search string was intended to capture articles that focus on policy appraisal and/or evaluation with an explicit focus on methodology or the framework in use. The process we used was the PRISMA process: http://www.prisma-statement.org/.

Data screening

The following steps were taken to narrow down the most useful articles to provide insights to the current state of affairs in policy evaluation.

Time

The academic journals were all limited to the period 2014-2019. This demarcation was set to present the current state of affairs, while including sufficient analytic content.

Top impact journals

Orthodox journals were narrowed down by including only the top five journals in each of the four categories above. The journals were ranked based on their Eigen factor scores, which combines the amount of incoming citations with a weight based on the importance of the citing journals. The ranks for the top journals of each of the three years from 2015-2017 were summed up to provide a more robust selection. In this way, a single year ranked outside the top five would not necessarily leave out the journal if the other two years were better ranked, as such orthodoxy is not defined by inherent content, but by being published in the most cited journals. The resulting 20 orthodox journals were:

[59] For details, see here: https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme/missions-horizon-europe_en.
[60] See here: https://ec.europa.eu/info/publications/mission-oriented-research-and-innovation-policy-depth-case-studies_en.

Table 1: Orthodox journals in literature review

Public Administration	Economics
Journal of European Public Policy	American Economic Review
Environment and Planning C Politics and Space	Energy Policy
Public Administration Review	Journal of Financial Economics
Journal of European Social Policy	Journal of Finance
Journal of Policy Analysis and Management	Quarterly Journal of Economics

Political Science	Interdisciplinary Social Sciences
American Journal of Political Science	Accident Analysis and Prevention
American Political Science Review	Patient Education and Counselling
Journal of Politics	Archives of Sexual Behaviour
Public Opinion Quarterly	Journal of Safety Research
International Organization	Qualitative Health Research

The result was 256 orthodox articles.

Heterodox journals were selected based on the Lee et al (2010) list of heterodox journals. Lee et al ranked the journals by impact indices and the top 25 journals with hits on the search string were included in this review. This included the top 34 journals on the list since nine had no hits. The result was 129 heterodox articles.

Limitation of overrepresentation

For orthodox journals, the journal *Energy Policy* made up 187 of 256 articles, thus signalling a substantial energy policy bias in the resulting data. To avoid this bias, a sample of 50 articles was constructed by including only the top 10 cited articles from 2018-19 and an additional top 40 papers from the entire timespan. The time selection was made not to bias against recent work with naturally fewer citations.

For heterodox journals, *International Journal of Social Economics* included 76 out of 129 heterodox articles. In order to avoid bias, the 20 most cited articles from the entire timespan was included. Inspection showed that time-bias would not be the same issue as for *Energy Policy* articles.

For specialised journals, in order for not to make this journal category too dominant (based on anticipated higher eligibility among final articles), the pool of 112 was limited to 60 articles. All 30 articles published in 2018-19 were included, along with the top 30 most cited articles from 2014-17.

This process provided a total of 192 articles to be screened qualitatively for eligibility and coded.

119 orthodox articles

73 heterodox articles

60 specialised articles

Eligibility

The following two criteria were used to determine whether an article was relevant to the review of the current state of affairs in policy evaluation and appraisal. Eligibility was assessed by first reading the articles' abstracts. If this did not provide conclusive responses to the two criteria, the article would be opened and read until a judgement could be made.

Criteria

Is the subject of the article either:

1. methodology and frameworks for policy impact estimation/appraisal (ex ante) and/or evaluation (ex post)

OR

2. an actual policy impact assessment/appraisal and/or evaluation?

If none of these criteria were met, articles were deemed ineligible to be included in the review. In the beginning of the coding process, 'grey zone' articles were discussed in the research group to secure consistent translation of coding principles into practice.

The result of the eligibility coding was as follows:

Table 4: Eligible academic articles

Journal type	Eligible	Ineligible	Grand total
Heterodox	25	48	73
Orthodox	44	75	119
Specialised journal	38	22	60
Total	**107**	**145**	**252**

Table 5: Eligible policy academic articles/cases

Policy literature type	Eligible	Ineligible	Grand total
Centre for Public Impact Cases	56	0	56

Mission-oriented cases (EU)	12	5	17
MOIN survey	8	0	8
Total	**76**	**5**	**81**

Coding of eligible articles (academic and policy)

We looked at three sets of variables: how the sources define value; mode of policy-making; and at what level of policy the sources focuses on.

In the literature review, there were four main coding variables:

Value targets:

- Public value
- Private value

Mode of policy-making:

- Market-shaping
- Market-fixing
- Non-market-oriented

Policy level

- Policy sector
- Policy programme
- Policy instrument

Methods

The three first variables were categorised with closed coding where the possible values are already defined by the values given above (definitions below). Methods were coded semi-inductively, as newly encountered methods were added to an expanding list of methods, which was used to categorise articles to enable quantitative patterns to emerge.

Definitions

Value targets

This variable distinguishes whether the framework is intended to achieve either private or public value.

1. **Private value** is defined as exchange value or otherwise imputable individual gains. These gains refer to higher income and employment, fewer inefficiencies, higher

productivity, fewer hours wasted etc. Private value can usually be assessed at the output stage of activity and is therefore able to be expressed in monetary values, especially if the good is marketable in principle.

2. **Public value** is defined by the achievement of common goals that are unachievable through individuals' action. Public value gains are diffuse and thus hard to allocate to specific individuals, yet enjoyed by most in the abstract. Public value is usually assessed at the outcome stage and thus harder to quantify in monetary terms.

Mode of policy-making

This variable distinguishes between how the policy-maker, civil servant or academic reasons the public or private value objective shall be achieved.

1. **Market-shaping policies** are defined by their aim for dynamic efficiency, where the policy objective is to make the best use of resources to achieve desired changes over time, including through the creation of new resources and expansion of the technological frontiers. Markets are understood as outcomes of interactions with the prevailing policy and regulatory framework: why policy-makers must consider how they intend markets to be co-created. In the co-creation process, 'systems thinking' is needed to account for emergent properties, feedback loops, lock-in effects and non-linear causalities at the meso-level. Given the system dynamics and the uncertainty of the future, policy-making is conducted with scepticism in terms of realising an enduring state of optimality, though with a constructive attitude to irreducible uncertainty.

2. **Market-fixing policies** are defined by their aim for allocative efficiency. The policy objective is to make best use of the (fixed) resources at a fixed point in time. This requires the identification of 'market failures' where the price mechanism is unable to balance supply and demand of a good or service. Policy-making therefore strives to correct or accommodate these deficiencies and achieve a state of optimality. These policies are usually assessed independently of other moving parts in the system, i.e. in a micro-economic manner.

3. **Non-market-oriented policies** are defined by their aim to change output and outcomes in realms of society largely isolated from the effect of market forces. These relate to outcomes within bureaucracies and in the third sector, including cultural, family and personal goals.

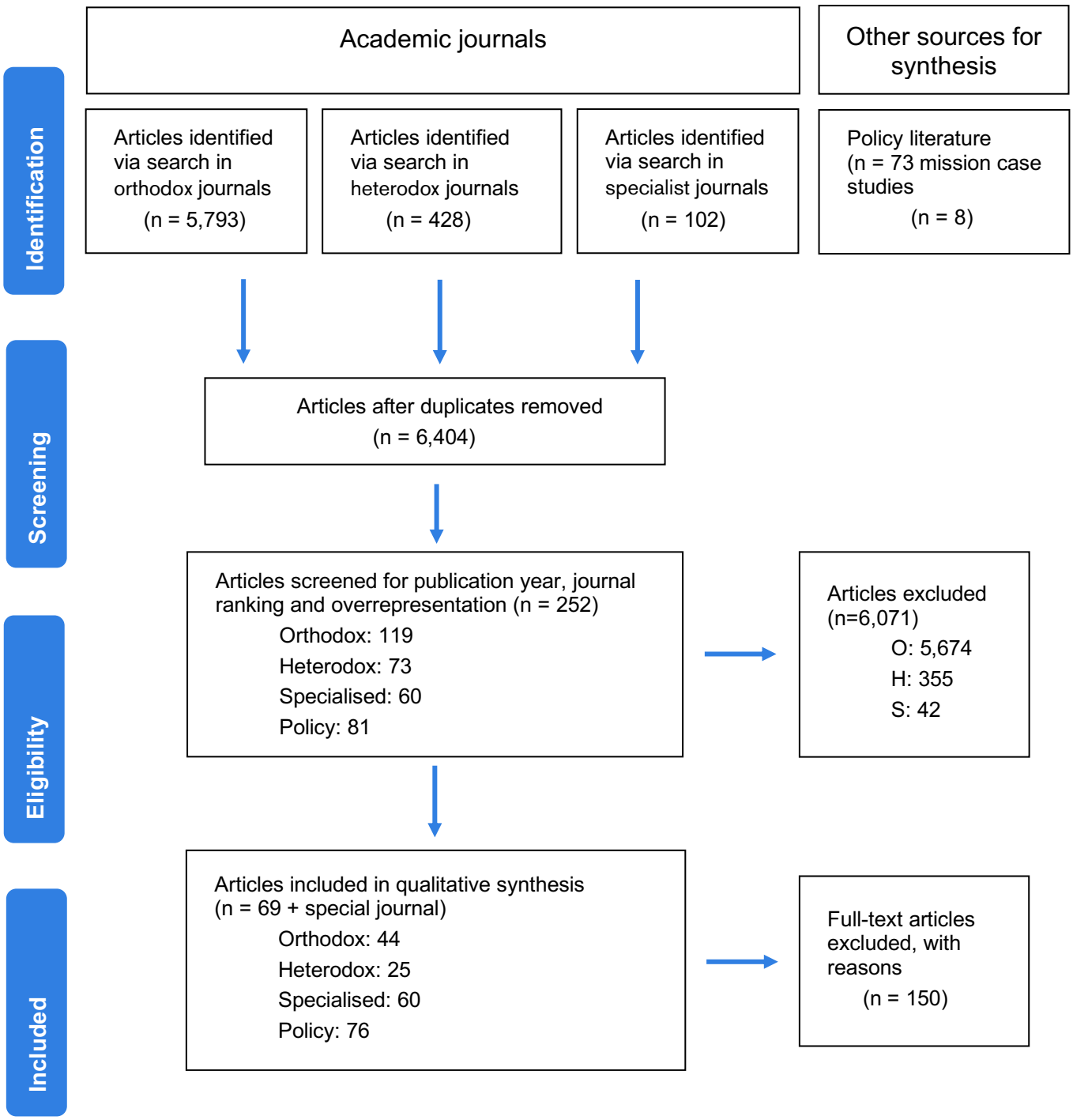
Policy level

Here we follow typology by Peters et al 2018[61] who divide policy into the following levels:

- High-level abstraction (policy-level): policy goals – what general types of ideas govern policy development?
- Operationalisation (programme-level): programme objectives – what does policy formally aim to address?
- On-the-ground specification (measures-level): operational settings – what are the specific on-the-ground requirements of policy?

[61] Peters et al 2018, op. cit.

Appendix 2. Literature review flow diagram



Appendix 3. MOIN survey

The survey was carried out through surveymonkey.com in April-July 2019.

Q. Using examples, what kind of analytical frameworks and policy toolboxes are you using in policy appraisal and evaluation processes?

Q. Considering the examples above, do these frameworks and tools specifically mention or discuss public value or similar concepts?

Q. If you use such analytical frameworks and policy toolboxes, how are they internalised within your organisation? (e.g. written guidelines; from organisations or from external actors such as parent ministry; how often are they updated and discussed?)

Q. Have you attempted to create alternative appraisal and evaluation tools? If yes, which ones and with what degree of success?

Q. Does your organisation deal with investments or grants? If yes, how do you evaluate the impact of these investments or grants?

Q. Do you use multiple alternative metrics, if yes could you specify which ones?

This publication is available from: www.gov.uk/beis

If you need a version of this document in a more accessible format, please email enquiries@beis.gov.uk. Please tell us what format you need. It will help us if you say what assistive technology you use.