Show your workings

Assessing how government uses evidence to make policy

Jill Rutter and Jen Gold

Contents

About the authors

Jill Rutter leads the Institute for Government's work on better policy making. Before joining the Institute, Jill was Director of Strategy and Sustainable Development at the Department for Environment, Food and Rural Affairs. Prior to that she worked for BP for six years following a career in the Treasury, where she was Press Secretary and Private Secretary to the Chief Secretary and Chancellor, as well as working on areas such as tax, local government finance and debt and export finance. She spent two-and-a-half years seconded to the Number 10 Policy Unit.

Jen Gold is a Senior Researcher at the Institute for Government, where her work largely focuses on policy implementation and ministerial effectiveness. Before joining the Institute in July 2014, she was a Senior Policy Associate at the University of Toronto's Mowat Centre for Policy Innovation, where she was practice lead for its government transformation programme. She has also worked at the Foreign & Commonwealth Office and Demos. Jen holds a PhD from the University of Cambridge and was a Visiting Fellow at Harvard University's Kennedy School of Government in 2009.

Acknowledgements

Foreword

There has been a lot of talk about 'evidence-based policy making' over the last 35 years. In the 1980s, academic studies showed that there was an enormous gulf between how policy makers retrospectively described how policy decisions were reached – through careful sifting and weighing of the evidence – and the 'Heath-Robinson' reality.

The 1990s saw explicit moves by government to strengthen the use of evidence. A few of these moves, such as the creation of the National Institute for Health and Care Excellence (NICE), have led to a lasting change in the landscape. But many other such moves stumbled on both the academic and government sides, leading to the sense that policy was more often 'evidence-sprayed' than 'evidence-based'.

Sometimes the simplest questions lead to the most important insights. This report and framework, developed independently of government, started with a discussion about whether government departments and public servants are using evidence well. The small Cabinet Office 'What Works' team had conducted a small number of joint reviews with government departments on how well they generated, synthesised, shared and used evidence – both inside the department and with professionals and commissioners that they related to. But the 'evidence on evidence' was often weak. How many government statements and policy papers are based on, or at least draw on, good evidence? And what would 'good' look like? These were questions beyond the capacity of a small Cabinet Office team, and given the chequered history of this area, perhaps better pursued externally anyway.

I am very grateful to the Institute for Government, and a group of other scholars and institutions, for taking up these challenging questions. I also commend the way they have gone about it. Rather than leaping into a hurried assessment, they have pulled the problem apart and sought to take it one step at a time, starting with thinking about what 'good' might look like, and in particular focusing on the prior question of transparency around the use of evidence. Like good examiners, they want the candidate to 'show their workings', not just give the answer.

Surely they are right. When something is claimed as true, effective or 'what works', we only really have two ways to assess it. First, we can compare it to something else we know to be true. But how do we know that this other belief itself is true, especially as behavioural science has shown that we are all – experts and public alike – prone to overconfidence, and to ignoring that which does not fit with our prior beliefs? Second, we can assess the method or route that gave rise to the claim. If it looks sound – if the 'workings' make sense – then the claim is more likely to be true.

The authors of this report, and the framework for assessing the quality and transparency of evidence that they present, stress that this is only the start of a process. It will need others to use it and populate it to bring it to life, and to move from whether evidence was used, to whether it was used well. I hope they will. I also hope that government departments, and the policy profession in particular, will embrace this activity as a tool to sharpen their own skills and work.

There will always be many factors that go into a policy decision, such as the legitimate political commitments of an elected government. But the wise and appropriate use of evidence should be at the heart of policy too: what is the problem; what are its causes; and what might work to change it? Recent developments in 'experimental government', and the generation and use of evidence, are encouraging. Yet we should restlessly seek to use evidence better, and welcome the efforts of those who challenge and support us in this goal.

David Halpern, What Works National Adviser, Cabinet Office

1 Introduction

The origins of this piece of work lie in a seemingly simple idea from government What Works National Adviser and former Institute for Government Research Director, Dr David Halpern: to see if it was possible to develop a rapid assessment tool to rate government departments on the use of evidence behind policy decisions. The idea was to be able to compare and rank departments – to show who used evidence well and who less well, and in doing so highlight and celebrate good practice while incentivising others to match the standards of the best. In doing so, the Civil Service would also be able to show the extent to which it was making progress on the commitment in the *Civil Service Reform Plan: One year on report* to publish more of the evidence base that supports policy making[1] and meeting its commitments on transparency in the Open Government *Action Plan*.[2]

The Institute had already done work on the use of evidence in policy making – in particular looking at the barriers to the use of evidence. In our 2012 report, *Evidence and Evaluation in Policy Making: A problem of supply or demand?*, we concluded that, while there were issues with the supply of high-quality and timely evidence, the bigger problem lay on the demand side.[3] One reason for this was that there were few obvious political penalties for failing to base decisions on the best available evidence – the Treasury had not, to date, been very demanding and Parliament was also not particularly interested or systematic in its analysis. In other work, produced jointly with the Arts and Humanities Research Council, we also looked at the uses of history and intercultural learning in policy making.[4]

There were important moves in the last Parliament to promote both the better use of evidence and justification of policy decisions. These are summarised in Box 1.

But while all are individually worthwhile, none of these approaches constitutes a transparent, systematic method for assessing the use of evidence across the full range of government policy making. In 2012, the National Audit Office launched an inquiry into the quality, use, cost, and extent of the evaluation activity undertaken or commissioned by Whitehall departments. The final report was published in 2013[5] and concluded that commissioning procedures lacked coherence, evaluation was generally under-resourced, and evaluation evidence was rarely being fed back into policy making. But the focus was on evaluation, not on the way evidence was used to inform policy decisions. There was also a very worthwhile exercise undertaken by the House of Commons Education Select Committee at the end of the last Parliament to check the evidence behind nine Department for Education policy areas, and invite public comment and subject them to parliamentary scrutiny. We recount this innovation in more detail in Annex 1. But their work suffered from the lack of a clear framework to be used to assess evidence use in each area of policy making.

This is the gap we were trying to fill: to develop an approach to assessing the evidence behind departmental decisions. This short report does four things:

1. It sets out in detail our attempts to develop a framework for assessing departments' use of evidence – and why such an apparently easy task proved to be so difficult to put into practice.

2. It explains the approach we ended up testing.

3. It sets out the framework that emerged from that testing.

4. It makes recommendations for how those responsible both for policy making and for holding policy makers to account should respond both to the framework and also to the issues raised in the report – and sets out the next steps we plan to make to use the framework to benchmark government performance.

---

1   *Civil Service Reform Plan: One year on report*, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/211506/CSR_OYO_LOW_RES_PDF.pdf

2   Open Government Partnership, *UK Action Plan 2011–13*, www.opengovernment.org.uk/resource/uk-action-plan-2011-13/. Gavin Freeguard et al, *Whitehall Monitor 2015* (Institute for Government), forthcoming.

3   Jill Rutter, *Evidence and Evaluation in Policy Making: A problem of supply or demand?* (Institute for Government), 2012, www.instituteforgovernment.org.uk/sites/default/files/publications/evidence%20and%20evaluation%20in%20template_final_0.pdf

4   Catherine Haddon, Joe Devanny, Charles Fosdick and Andrew Thompson: *What is the Value of History in Policymaking?* (Institute for Government), January 2015, www.instituteforgovernment.org.uk/sites/default/files/publications/Making%20History%20Work%20Report%20-%20Final_0.pdf

5   National Audit Office, *Evaluation in Government*, December 2013, www.nao.org.uk/wp-content/uploads/2013/12/10331-001-Evaluation-in-government_NEW.pdf

BOX 1: SUMMARY OF KEY INITIATIVES BETWEEN 2010 AND 2015

1. The Office for Budget Responsibility

The Office for Budget Responsibility (OBR) was established immediately after the 2010 general election to take over the Treasury's forecasting responsibilities. As part of its role in preparing fiscal forecasts, the OBR scrutinises the assumptions departments use to calculate the fiscal impact of measures associated with the Budget and Autumn Statement.

2. What Works Centres

A network of independent What Works Centres was launched in March 2013 to gather together, scrutinise, and share evidence on effective interventions. Spanning health, justice, education, local economic growth, early intervention, wellbeing and ageing, the centres are intended to inform "decision-making on some £200 billion of public spending".[6] This built on the earlier establishment of NICE (1999) and the Educational Endowment Foundation (2011).[7]

3. Departmental evidence audits

In 2014, the Cabinet Office's What Works Team began a series of reviews into the generation, dissemination, and use of evidence by individual government departments. The Department for International Development was one of the first departments to participate and the review of this department is the only one to date for which results are publicly available.[8]

4. Independent Commission on Aid Impact

The Independent Commission on Aid Impact (ICAI) was created in May 2011 to scrutinise the effectiveness of the UK's aid spending. Andrew Mitchell, the-then Secretary of State for International Development, described the thinking behind the Commission in October 2010: "The Commission will shine a light on where aid works and where improvements are needed. We will lead other donors in opening our books to independent scrutiny. We have a duty to squeeze 100 pence of value from every pound".[9] The work of the ICAI is in turn scrutinised by a parliamentary sub-committee.

5. Regulatory Policy Committee

The Regulatory Policy Committee (RPC), established in 2010, scrutinises the quality of the evidence base behind departmental regulatory proposals before ministers consider them. The RPC issues a rating based on the evidence supporting the cost, benefit, and risk estimates within each proposal's impact assessment.[10]

6   Cabinet Office and Treasury, Press Release: New world leading evidence centres to drive better decisions across £200 billion of public services, 4 March 2013, https://www.gov.uk/government/news/new-world-leading-evidence-centres-to-drive-better-decisions-across-200bn-of-public-services

7   Education Endowment Foundation, History, https://educationendowmentfoundation.org.uk/about/history/

8   Department for International Development and Cabinet Office, What Works Review of the Use of Evidence in the Department for International Development, 2014, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/340971/What_Works_Review_DFID.pdf

9   Department for International Development, News Story: New independent commission unveiled, 29 October 2010, https://www.gov.uk/government/news/new-independent-commission-unveiled

10 Regulatory Policy Committee, *Securing the Evidence Base for Regulation: Regulatory Policy Committee scrutiny during the 2010 to 2015 Parliament*, March 2015, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/415102/2015_03_03_RPC_Annual_Report_2014_website_copy_revised_2015_03_19.pdf

2 Developing an approach to evidence assessment

Approach

We decided to adopt an iterative and experimental approach to developing our framework – modelling the approach that government itself commends as part of open policy making!

To start we needed some raw material to assess. The Cabinet Office helped us to identify 10 policies which had been recently updated on gov.uk.[11] We then brought together a group of practitioners from Parliament and several research organisations to assess the evidence base behind these policies. We assigned one or two policies to each reviewer, making sure that each policy was assessed at least twice. We issued each reviewer with a template to fill in to capture the results of their review, their methodology, and also their views on the strengths and weaknesses of the approach they adopted. We also asked them how long the process took to complete and how confident they were in the results they generated. We then shared those results with a wider group of practitioners and used their feedback to develop a prototype framework that we then subjected to a second round of reviews.

We agreed at the outset that, in order to achieve our longer-term objectives, we needed an approach that:

- could be applied quite rapidly
- did not require subject matter expertise
- produced meaningful results that gave useful feedback
- produced consistent results when applied by different people
- allowed comparison between different policy areas or departments.

Issues

The initial commission sounded easy: to develop a framework allowing a consistent assessment of departments' use of evidence in policy development. In practice this proved much harder – that in itself is quite instructive. The particular pitfalls we encountered are set out below.

There was a general problem in defining a 'policy'. In our 2011 report, *Policy Making in the Real World,* we struggled to come up with a neat definition of policy in one sentence – instead we described the different meanings of 'policy'. Part of the problem lies in the difficulty of defining policy making. In that report we said:

Our focus is on policy making as an activity intended to achieve the purposes of politicians in government. But the 'policies' that this activity produces can refer to many different things, including:

- *the goals or strategies of [political] leaders*
- *specific acts such as decisions, announcements and statutes*
- *an overriding logic of action (e.g. 'our policy on the environment')*
- *a structure of practice (e.g. 'the school's policy on late essays').*

Despite this variety of uses, 'policy making' is often understood to mean the formal expression of activities undertaken by government to achieve outcomes, often through legislation. Yet policy may have none of these characteristics. Policy may need to be inferred from practice (rather than being formally stated); policy may require government to do nothing (rather than to act); and policy may be created purely for temporary political positioning (with no intention of achieving outcomes).[12]

---

11   This phase took place in early 2015.
12   Michael Hallsworth, with Simon Parker and Jill Rutter, *Policy Making in the Real World* (Institute for Government), April 2011, P19. www.instituteforgovernment.org.uk/sites/default/files/publications/Policy%20making%20in%20the%20real%20world.pdf

For the purposes of this exercise we were principally interested in the evidence base for new policies – where government had made an explicit decision to intervene to improve the status quo. That meant we were focusing on the flow of new policies, rather than on the much bigger stock of existing policies. This focus meant it was realistic to look for a statement from government on the problem it was trying to address, the justification for intervention, and the rationale for its choice of intervention. We also looked to see how the Government planned to assess the outcome of its intervention. However, even with that – very narrow – focus, it proved surprisingly difficult to take the exercise forward. Below we set out some of the problems we encountered.

1. Where do we find policies?

This sounds easy in theory but in practice proved very difficult. The Government's unified information platform, gov.uk, has a section called 'policies' alongside other such generic categories as 'publications and announcements'. In the week after the election was called there were 224 'policies' on gov.uk.[13] On advice from the Cabinet Office, we decided to select these policies as the initial unit of scrutiny.

However, looking within each policy listed on the gov.uk website, it is far from easy to work out what the 'policy' is in any particular area. The discipline of producing a consultative green paper, followed by a more definitive white paper, has largely disappeared. Instead, under a single heading, there is a series of links to early policy documents, some reports on consultations, and some regulations. But there is no single 'policy idea' and no clear line of sight from an original proposal to the final change that is being implemented. Within a policy heading there are often a number of different policies mentioned (for example, on patient choice, the original document was the 2010 NHS White Paper and this featured as one chapter; on dementia, we produced very different results – when one reviewer looked at the overall policy and another focused on the best-documented element of the strategy, dementia-friendly communities). There are also multiple links whose relation to each other (and to previous stages of policy development) is very unclear.

In testing our framework the second time, we decided that we could not leave people simply to work through the set of links presented as a 'policy' on gov.uk – so instead we extracted a document or set of documents which would replicate the material that would have been made public at the time of a significant policy announcement. The need to be clearer about decision points for policies and present them more accessibly is something we look at later.

2. What is the evidence base for policy?

Our first examination suggested it is easiest to establish the evidence base for the initial policy proposition in a published document. If there are any citations or footnotes, this is where they are to be found.

Most academics regard peer-reviewed literature, or, even better, systematic reviews, as the most robust form of evidence. But even here, there may be a temptation to interpret results and over-extrapolate from a single study to a more general case. But government documents cite many other forms of evidence – including:

- previous government documents – in many cases the most frequent links from a document are to earlier policy statements by the same department
- findings from external reviews – which may or may not clearly document the sources they draw on
- evidence from internal management systems
- reports from the National Audit Office and other scrutiny bodies
- evidence from external interest groups – which may or may not look like research evidence
- user experience
- case studies – which often appear selected to make a point – with no obvious attempt to assess how typical they are
- results of consultations.

Box 2 gives an example of an independent review (on school capital) and how it described its use of evidence. It is hard to trace any line of sight between the sources outlined and specific insights or recommendations.

---

13 The revamp of gov.uk since the election means that there is no longer an overall number and whereas previously a policy moved to the top of the list when a new document or link was added, now the list is entirely alphabetical. The list starts with the 2012 Olympic and Paralympic legacy.

> **BOX 2: James *Review of Education Capital* – methodology[14]**
>
> The Capital Review official call for evidence ran from 6 August 2010 to 17 September 2010. In total, 492 responses were received.
>
> Along with site visits and individual meetings, the Capital Review team ran a series of workshops involving a range of key partners (more than 100 people were involved) including Academy [schools] sponsors and local authority representatives, designers, contractors, consultants, and project managers. Their expertise was invaluable in mapping out existing capital processes and developing recommendations for a future approach to capital management. A thorough review of all published research on the Department for Education capital programmes was conducted and references to major reports such as the National Audit Office's report of 2009 can be found throughout the Capital Review report. The Capital Review team also analysed international evidence available from the Organisation for Economic Co-operation and Development (OECD), focusing on how civic building programmes and property management are managed in other developed countries.

3. What evidence has been used – and what is missing?

When we asked reviewers to look at the evidence base behind policies, there was a clear difference in the assessments of people who had some prior knowledge of the policy area and those who came to it cold. The latter tended to be more accepting of the evidence base produced by departments, whereas those with prior knowledge tended to point out issues and areas which were missing, criticise the quality of the evidence cited, and question the framing of the issue.

4. How has evidence been used?

Without in-depth subject matter knowledge it was very difficult to assess how the evidence base cited had been used. Academic articles are often behind paywalls and a citation does not necessarily imply that the right conclusion has been drawn or that less convenient findings have been ignored. Without going back to source documents where cited, it is impossible to tell whether evidence has been cherry-picked.

There is a further problem in tracing the impact of consultation. It is often difficult to see how the Government has responded to consultations. However, an example of best practice is the way in which the Committee of Advertising Practice (CAP) and the Broadcast Committee of Advertising Practice (BCAP) reacted to the results of a consultation on applying advertising standards to e-cigarettes, where each response was cited and the committees gave their reasons for accepting or rejecting it. This process has the added benefit of exposing the evidence base (or lack of it) behind consultation responses, rather than just looking at the numbers of people or organisations responding who may not be basing their responses on strong evidence. We produce an extract from the consultation on e-cigarettes in Table 1, where respondents were being asked to give their views on the adequacy of a rule that "marketing of e-cigarettes should be socially responsible".[15]

---

14   Sebastian James, *Review of Education Capital*, April 2011, https://www.education.gov.uk/consultations/downloadableDocs/James%20Reviewpdf.pdf P77.
15   Committee on Advertising Practice, *The Marketing of Electronic Cigarettes: CAP and BCAP's evaluation of responses,* 2014, http://cap.org.uk/News-reports/Consultations/Closed-consultations/~/media/Files/CAP/Consultations/ecig%20consultation/Evaluation%20of%20responses.ashx, accessed 7 May 2015. Thanks to Clive Bates of Counterfactual for drawing this to the authors' attention.

Table 1: Responses to consultation on e-cigarette marketing

Respondent making comment	Comment	Reaction
ADPH, ASH, ASH Scot, BLF, CRUK, FPH, Healthy Stadia, NHS Grampian, SSA, ASH-W, RTCA, BrBC, CCC, CFNI, SCOT, CIS, LCC, TSI, SPH, PV, PHE,TFF, HAT, HTCSG, SG, SFNA, SFN, PHS, NSTAG, SHWB, SFSW, SFCDTCA, TCRG, WHOT, Mr B., STCPAB, RSPH, CIEH, WSSS, TC[16]	Agree in principle but consider tighter wording is needed. Specifically, explicit reference to the fact that electronic cigarettes and other nicotine containing products are an alternative to tobacco, and that they are therefore not suitable for use by people who do not currently consume tobacco products.	CAP and BCAP disagree. BCAP are mindful that these products have a primary appeal to those looking for an alternative to smoking but it is not BCAP's role to dictate their proper use, nor are they aware of an evidential basis which requires mandating ads to be addressed only to existing smokers / nicotine users in all instances. In the absence of that BCAP consider it disproportionate to mandate this type of presentation in all advertising, but have provided a separate rule which prevents advertising which addresses non-smokers explicitly. Additionally, BCAP have yet to be persuaded of the value of compulsory messages in advertising. This is discussed in more detail in the evaluation of proposed rule 11, below.

SOURCE: Committee of Advertising Practice (2014), *The Marketing of Electronic Cigarettes: CAP and BCAP's evaluation of responses,* p. 9

Conclusion

These issues led us to conclude that there were two fundamentals to consider in assessing the evidence base behind government policy:

- the transparency of the evidence base
- the quality of the evidence base and the way it had been applied.

We concluded that assessing quality of evidence required subject matter expertise and the ability to interrogate the evidence base and in particular to identify instances of cherry-picking misuse or omission. That meant it failed to meet our criterion of being able to be done by people without specific subject knowledge and we decided not to pursue it.

But we also concluded that an essential first step to assessing quality was for the basis for the decision to be transparent – if it is not clear on what basis decisions have been made, it is impossible to judge the robustness of those decisions. We concluded that it was possible for a lay team to assess the transparency of evidence use. So the next stage of our work focused on developing an evidence transparency framework, to 'show the workings' behind government decision-making.

---

16   A full list of respondents and their abbreviations is published on pages 3 to 7 of the CAP report.

3 Testing the evidence transparency framework

We therefore decided to develop and test a simple framework that allowed us to rate the evidence base behind government policy announcements. The focus on transparency came both because it is an essential first step but also because, in our first exercise, people were generally much more confident in their ability to assess how transparent an evidence base was than to assess the quality of the evidence base.

In order to develop the assessment framework we used a 'chain of reasoning' approach.[17] This broke the assessment into five discrete steps:

- **problem** – the evidence for the issue under consideration requiring government action
- **intervention (what)** – the rationale for the choice of intervention (the hypothesis)
- **intervention (how)** – the rationale for the choice of delivery mechanism
- **value for money** – the assumptions on which any costing or cost-benefit analysis has been undertaken
- **testing and evaluation** – the use of testing or evaluation evidence from pilots, and clarity on plans to test and evaluate whether the policy achieves its objectives.

This follows the logic of other work on policy implementation both at the Institute for Government[18] and within government.[19]

As part of the framework we developed a transparency standard for each step and then rated each document on a scale from 0 to 3 for each category. A score of 0 signalled that the standard was not met at all; a score of 3, that it was fully met. With a maximum of 15 points in play, we used a RAG scale of 0–5 (Red), 6–10 (Amber), and 11+ (Green). But both the overall score and the individual elements are relevant so we decided to capture both.

We tested the initial framework on a number of policy documents published under the Coalition Government, using reviewers from the Institute for Government and other interested organisations. The policy documents selected ranged across a number of departments (the Department for Environment, Food & Rural Affairs (Defra), the Department of Energy & Climate Change (DECC), the Department of Health and the Cabinet Office) and from very specific policies to strategies and action plans.

In themselves these were not a representative sample as, in order to find something to analyse, we focused on policies supported by clear documentation – many of the 'policies' listed on gov.uk are not so supported. Doing this retrospectively also led to problems. In some cases links were broken and supporting documentation had disappeared and some documents appeared more like reports on activities undertaken so far than policy proposals: one was described by one reviewer as a 'puff piece'.

However, notwithstanding those reservations, six documents were reviewed by three or more people. Five of those ended up with an amber rating – averaging between 6 and 8.5 points (see Figure 1). One document was rated red by everyone, suggesting that the method was robust enough to identify documents with next to no evidence.

17   Beth Adams, Europe wants scientific advice to become pluralistic, PMLiVE, 15 May 2015, www.pmlive.com/pharma_news/europe_wants_scientific_advice_to_become_pluralistic_737218; Tracey Brown, The power of asking for evidence, TEDx Talks, 5 January 2015, www.youtube.com/watch?v=H_rFVBlbyZ4
18   Emma Norris et al, *Doing Them Justice* (Institute for Government), July 2014, www.instituteforgovernment.org.uk/publications/doing-them-justice
19   Cabinet Office Implementation Unit, *Implementation Insights*, 2015, www.gov.uk/government/uploads/system/uploads/attachment_data/file/428949/Implementation_Insights_.pdf

Figure 1: Summary of reviewers' ratings

Policy	Problem	Intervention – what	Intervention – how	Value for money	Testing and evaluation	Total
Policy A	2.5	0	0	1	1	4.5
	1	3	2	1	0	7
	3	1	0	2	1	7
	1	2	2	3	1	7
						6.375
Policy B	0	2	1	1	2	6
	2	1	1	2	2	8
	3	1	1	0	3	8
	3	2	1	0	2	8
	2	2	3	2	2	11
	2	3	3	0	0	8
						8.25
Policy C	3	1	0	0	1	5
	2	2	1	0	1	6
	3	2	1	1	2.5	9.5
						6.83
Policy D	1	2	1	1	2	7
	3	2	0	0	0	5
	1	2	1.5	0	0	4.5
	2	3	3	1	1	9
						6.38
Policy E	0	0	0	0	0	0
	1	0	0	0	0	1
	0	0	0	0	0	0
	0	1	1	0	0	2
	0	0	0	0	0	0
						0.6
Policy F	2	2	0	0	1	5
	2	2	2	0	1	7
	3	2	2	2	1	10
						7.3

However, although the aggregate results were quite closely clustered, scores under each category varied widely. In particular there were problems with:

- defining what problem or issue the policy was designed to address – reviewers differed on what the starting point for a policy was and they also pointed out that government action was not solely triggered by 'problems'
- distinguishing the "what" of intervention from the "how" of intervention
- assessing "strategy" documents with multiple interventions: Was it enough for some of these to be clearly evidenced – or did they all need to be? And could one score be assigned to a whole document?
- how much credit to give to policy documents that mention plans to generate evidence or select interventions through testing
- assessing costs and benefits, in particular how comprehensive the analysis needed to be – in general, documents were better at describing costs than at setting out the potential benefits
- rating documents where the document made clear there were gaps in the evidence base – but set out plans to fill some of those gaps.

In the light of these comments, we modified the framework to:

- rename 'problem' as 'diagnosis' to make it more neutral
- rename intervention (what) and (how) as 'proposal' and 'implementation' to distinguish these more clearly
- simplify the language to make it more accessible to non-expert users.

But we also developed it further to set out expectations for the presentation of the thinking behind a policy decision – to act as a checklist that policy makers could use in preparing documentation. Broadly speaking, the scale now builds from mentioning evidence with some explanation of how it is used (level 1), to linking the evidence to policy in a clear and traceable way (level 2), to doing that but also assessing the evidence base for uncertainties and contradictions (level 3). The framework remains a work in progress, but the version we ended up with is reproduced in Table 2 and is also available separately at http://www.instituteforgovernment.org.uk/publications/evidence-transparency-framework. We have also added an example of how it might apply to a hypothetical policy to guide users on what we thought needed to be considered under each heading.

But producing a framework is no guarantee of action or change. In the final section we set out our recommendations for how government should respond, and the next steps we propose to take.

Table 2: Evidence transparency framework

POLICY PROPOSAL	LEVEL:	0	1	2	3	WORKED EXAMPLE
DIAGNOSIS This concerns why something is proposed, i.e. what the issue is that will be addressed. **The document should explain:** • what policy makers know about the issue, its causes, effects, and scale • how policy makers have assessed the strengths and weaknesses of that evidence.		Not clearly enough for level 1.	Evidence is mentioned, with some explanation of how it has been used.	As in level 1 but the supporting evidence is linked to the relevant parts of the policy, properly cited and you could find the source.	As in level 2 but the evidence base is also assessed and uncertainties and contradictory information are acknowledged.	The Government has assessed the extent of problem drinking in the UK: the economic and human cost.
PROPOSAL What is the Government's chosen intervention? **The document should explain:** • why the Government has chosen this intervention • what evidence, if any, that choice is based on • how policy makers have assessed the strengths and weaknesses of the evidence base, including what has been tried before and whether that worked or not • whether there are other options and why they have not been chosen • what the Government plans to do about any part of the intervention that has not yet been decided upon.	**SO, CAN YOU SEE WHAT EVIDENCE HAS BEEN USED AND THE ROLE IT HAS PLAYED?**	Not clearly enough for level 1.	Evidence is mentioned, with some explanation of how it has been used.	As in level 1 but the supporting evidence is linked to relevant parts of the policy, properly cited and you could find the source.	As in level 2 but the evidence base is also assessed and uncertainties and contradictory information are acknowledged.	The Government has chosen to implement minimum unit pricing for alcohol, instead of, for example, increasing alcohol taxes or starting a new educational campaign.
IMPLEMENTATION How will the chosen intervention be rolled out? **The document should explain:** • why this method for delivering the intervention has been chosen • what evidence, if any, that decision is based on • whether there are other methods and if so the reasons for not choosing them • if the way to deliver the intervention is still being decided, what the method is for deciding.		Not clearly enough for level 1.	Evidence is mentioned, with some explanation of how it has been used.	As in level 1 but the evidence is linked to relevant parts of the policy, properly cited and you could find the source.	As in level 2 but the evidence base is also assessed and uncertainties and contradictory information are acknowledged.	The Government has decided to implement minimum unit pricing through a voluntary agreement with major retailers rather than through legislation.

VALUE FOR MONEY This considers the costs and benefits of the policy to show why the Government thinks it is worth doing **The document should explain:** • what the costs and benefits are estimated to be • the assumptions behind those calculations • what evidence is being used to make those assumptions • the uncertainties about the costs and benefits and how likely the figures are to change.	**SO, CAN YOU SEE WHAT EVIDENCE HAS BEEN USED AND THE ROLE IT HAS PLAYED?**	Not clearly enough for level 1.	The assumptions (the basis) for conclusions about risks and benefits are described.	As in level 1 but supporting evidence is also properly cited and you could find the source.	As in level 2 but it is also clear how the uncertainties in these assumptions have been considered.	The assessment shows the potential of the proposal to reduce problem drinking, but also the impacts on business (e.g. supermarkets, pubs), the public sector (e.g. police, NHS), and the public of raising prices.
TESTING AND EVALUATION How will we know if the policy has worked? **The document should explain:** • plans to measure the impact of the policy and the outcomes that will be measured • plans to test the policy first, or reasons why not • plans to evaluate the effects of the policy, including a timetable.		Not clearly enough for level 1.	Some indication of success measures but no plans for testing/evaluation (or explanation of why inappropriate).	More comprehensive success measures (or process for developing them outlined). Also provides details about use of testing and plans for evaluation or explains why testing or evaluation would not be appropriate.	As in level 2 but explains the reasons for the use of testing and plans for evaluation. It is also clear what will happen to the results of testing and evaluation, including timing and plans for publication.	The Government sets out how it plans to measure the results of the policy. The Government sets out plans for piloting, initial evaluation of those results and timetable for publication and then describes decision process around roll-out if the evaluation is satisfactory.

4 Conclusions and next steps

Understanding why government is doing what it does is essential for accountability – to citizens and the electorate, to Parliament and indeed to other parts of government. Clear communication to others can also hold a mirror up to internal thinking and ensure that ministers and departments are themselves clear on why they have chosen a particular course of action.

That is not to say that policy choices are only legitimate when rooted in 'evidence' – there are plenty of reasons why government acts in the absence of evidence. Some decisions are simply an expression of values and preferences; in other cases government may be stepping into new territory where it has to act in the absence of prior experience; some changes are not amenable to the sort of methods that produce the most rigorous evidence.

Notwithstanding those caveats, our earlier work has shown that policy is more likely to succeed if governments:

• have properly analysed the issue they are trying to address

• have conducted a wide-ranging search for evidence, both of what has gone before and of other interventions

• have comprehensively considered the form of proposed interventions

• are clear on the assumptions on which they assess the benefits and costs (and possible risks around them)

• put in place plans for feedback, testing, evaluation.

In these cases where a policy change is 'instrumental' rather than the end in itself, it makes sense to ask why the Government believes it may have the desired results (the underlying hypothesis) – and how it is going to tell if it was right.

This will also make it easier to distinguish between good failures (ideas that were worth trying – but didn't work) and bad failures (ideas that were weakly supported, or where evidence suggested there was little chance they would produce the desired effect without unintended consequences) that should never have made it off the drawing board. This in turn should give government greater space to innovate.

Recommendations

It is much harder than it should be to establish not just the reasoning behind a government policy, but also, more prosaically, what a policy is. The presentation of government activity in a policy area on gov.uk is not helpful (although it does contain some useful links to what is going on). In particular, the lack of a controlling mind to show how policies evolve make looking at the existing stock of 'policy', and working out what lies behind it, a near impossible task. This is not a good position for a government committed to transparency. So we recommend that:

• **Heads of the policy profession (in conjunction with directors of communication and the Government Digital Service) should develop a better way to present 'policy' and, in particular, policy changes to the public, in a way that gives a clear line of sight between initial proposal and final decisions.** This may require changes to the practice of documentation and online presentation. We have previously suggested a 'Policy Assessment'[20] to accompany policy announcements, which would cover the evidence base behind a policy. This could be done by extending to all significant policy decisions the model used now for impact assessments for regulatory proposals that affect business and civil society. These could be updated as a policy is modified.

• **Government should also consider adopting the Committee of Advertising Practice approach to responding to consultations**, to make clear on what basis it is accepting or rejecting recommendations – and this should be done in standard form.

We think the transparency standard represents a minimal starting point for departments to tell Parliament and the public the basis on which decisions have been made and we recommend that:

• **Chief scientific advisers/chief analysts should produce a statement on how their department will comply with the transparency standard (and ensure that it is complied with).**

But it is also important that departments are held to account for the evidence standard behind policy, and Parliament has a

---

20   Michael Hallsworth and Jill Rutter, *Making Policy Better* (Institute for Government), April 2011, www.instituteforgovernment.org.uk/sites/default/files/publications/Making%20Policy%20Better.pdf

vital role to play here. One way is to have specific evidence-check exercises – however, asking questions about the rationale behind policy choices should be core work for committees. Our framework offers an approach which committees could use. So in addition, we recommend that:

- **Parliamentary select committees and bill committees should hold departments to account for the transparency of the evidence base behind policies.**

It is not just government departments that need to show their workings. Bodies outside government seek to influence policy. Similar considerations should apply to their use of evidence. If government becomes genuinely transparent about its evidence base (and commits to an engaged approach to consultation responses), it is entitled to hold potential influencers, particularly large non-governmental organisations, to the same evidence standard. This would also help trustees ensure that influencing bodies that have charitable status are acting in accordance with their charitable mission.

Next steps

We have set out recommendations above to the Government and Parliament. But we also want to use the evidence transparency framework to recognise best practice and to point out where there remains room for improvement.

For the next stage of the work, Sense about Science, the Alliance for Useful Evidence, and the Institute for Government will work together to promote the use and discussion of the framework. We will make it publicly available[21] and encourage members of the research community, civil society, students, members of the public, and civil servants to test it, and to give us feedback on its usability and usefulness, but also to rate individual policies.

Next year, we plan to run an initial benchmarking exercise to see how government is faring on evidence transparency. The aim will be to identify top performers and to note those who still have some way to go. This exercise will be led by Sense about Science.

From transparency to quality

As we have pointed out, transparency is just the first step. Without transparency, it is impossible to assess the quality of the evidence base. Grading scales already exist to do this: but they can only be applied if departments are explicit about how they come to decisions. Reproduced in Figure 2 is the quality scale used by the Educational Endowment Foundation (one of the Government's What Works Centres) in its teaching and learning toolkit for headteachers.[22]

Figure 2: Education Endowment Foundation evidence grading scale

Rating	Description
*	*Very limited:* Quantitative evidence of impact from single studies, but with effect size data reported or calculable. No **systematic reviews** with quantitative data or meta-analyses located.
**	*Limited*: At least one **meta-analysis** or systematic review with quantitative evidence of impact on attainment or cognitive or curriculum outcome measures.
***	*Moderate:* Two or more rigorous **meta-analyses** of experimental studies of school age students with cognitive or curriculum outcome measures.
****	*Extensive:* Three or more **meta-analyses** from well-controlled experiments mainly undertaken in schools using pupil attainment data with some exploration of causes of any identified heterogeneity.
*****	*Very Extensive:* Consistent high quality evidence from at least five robust and recent **meta-analyses** where the majority of the included studies have good ecological validity and where the outcome measures include curriculum measures or standardised tests in school subject areas.

There is reason to suppose that more transparency could increase the incentives to improve quality. Our work to date suggests that improving transparency is an essential first step.

---

21   It will be available to the public at www.AskforEvidence.org/index with supporting materials for understanding evidence and on the Institute for Government website.

22   Reproduced from the Educational Endowment Foundation (EEF) Teaching and Learning Toolkit, 'About the Toolkit', https://educationendowmentfoundation.org.uk/toolkit/toolkit-a-z/about-the-toolkit/. Note that the EEF uses padlocks, not stars! It also rates cost-effectiveness of interventions on a separate scale.

Annex 1: The House of Commons Education Select Committee's Evidence Check

The assessment framework presented in this report drew upon lessons from an innovative Evidence Check initiative launched by the House of Commons Education Committee in November 2014.

Why did the Committee run an evidence check exercise?

The Department for Education claimed to be a strong advocate for making teaching an "evidence-based profession" during the last Parliament.[23] One of Michael Gove's first acts as Secretary of State for Education was to set up the Education Endowment Foundation – an independent What Works Centre with an endowment of £125 million and a mandate to:

- gather available evidence on effective interventions and share it with frontline practitioners
- fill gaps in the existing evidence base by running randomised controlled trials in schools.[24]

But whether the Department was also putting evidence "at the heart" of its own decision-making was a question the Committee wanted to investigate.[25]

How did the Evidence Check process work?

The Department was asked to produce two-page summaries of its use of evidence in nine policy areas that had not been the subject of recent Committee inquiries. These policy areas ranged from "big ticket political commitments like free school meals, to narrower topics like music education" (see Table 3).[26] Responses were then posted on a series of online web forums hosted on the Parliament website, with the public invited to comment on "the strength of the evidence provided".[27] A separate web forum was also created to solicit general views on "the Department's use of evidence in its policy making".[28] In total, the committee received almost 500 responses from individuals and organisations.

Table 3: Public response to the Education Committee's Evidence Check exercise

Policy areas	Number of responses
Department's use of evidence	154
Summer-born children	110
Phonics	90
School starting age	64
Universal infant free school meals	19
Music education	18
Teaching assistants	16
Impact of raising the participation age	11
National College for Teaching and Leadership	8
Professional measurement metrics	7
TOTAL	**497**

23   Graham Stuart, "Evidence check – are the Department for Education's policies evidence-based?", Institute for Government guest blog, 21 November 2014, www.instituteforgovernment.org.uk/blog/9478/guest-blog-evidence-check-are-the-department-for-educations-policies-evidence-based/

24   See Education Endowment Foundation, About Us, https://educationendowmentfoundation.org.uk/about/

25   Graham Stuart, 'Evidence check', Institute for Government guest blog.

26   Jonathan Breckon, 'Your chance to vet the Department for Education's evidence', Alliance for Useful Evidence blog, 1 December 2014, www.alliance4usefulevidence.org/your-chance-to-vet-the-department-for-educations-evidence/

27   Education Committee, 'Evidence Check' web forum, www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/dfe-evidence-check-forum/

28   Ibid.

Web forum posts were then reviewed and used to identify policy areas that would be the subject of oral evidence sessions. These were:

- The work of the **National College for Teaching and Leadership**. This policy area was selected as a result of respondents' criticism over the lack of evidence in the Department's two-page submission[29] – as one respondent put it: the submission took "the prize for having the least evidence in an evidence check"![30]
- **School admissions policies** in relation to both summer-born children and the school starting age. The scale of the response (174 comments across the two web forums in question) and level of challenge to the evidence provided by the Department made this an obvious choice for the Committee.[31]

Why was Evidence Check innovative?

The Committee's typical inquiries involve the Department submitting written evidence at the same time as members of the public. As a result, the Department's testimony is not opened up to wider scrutiny and the Committee misses out on the benefits of pooled expertise (e.g. being alerted to contrasting evidence or flaws in the Department's interpretation of existing evidence). As the Committee's Chair, Graham Stuart, explained:

Evidence Check turns this around – the Department's view of the evidence across a range of areas is available, and anyone can contribute to an iterative discussion of the strengths and weaknesses of the Department's statement.[32]

What lessons emerge from this process?

1. Transparency is a first step in assessing quality

Even the most cursory of glances at the web forums reveals the extent to which people's ability to comment on the "strength of evidence provided" depends on the quality of the Department's two-page submission. At one extreme, the Department's submission on the work of the National College for Teaching and Leadership (NCTL) simply offered a description of the creation, purpose, and aims of the College, but no evidence of impact or indication of the evidence base underpinning the College's activities. As a result, most respondents resorted to offering personal accounts of their encounters with NCTL alongside their own views on the organisation's effectiveness.

In contrast, the Department's submissions on policies such as phonics offer a much clearer summary of the key evidence underpinning policy decisions. This allowed respondents to directly engage with the evidence cited and highlight contrasting evidence, selection biases, and gaps in the existing evidence base (see Figure 3). The exercise shows that transparency is a first and necessary step in enabling the quality of a department's evidence-based decision-making to be judged.

29   Graham Stuart, "Evidence Check: National College for Teaching and Leadership", Education Committee video, 6 February 2015, www.youtube.com/watch?v=bBiiiwKP0kw
30   Nansi Ellis (ATL), Education Committee web forum: National College for Teaching and Leadership, 10 December 2014, www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/dfe-evidence-check-forum/national-college-of-teaching-and-leadership/
31   Graham Stuart, 'Evidence Check: Starting School', Education Committee video, 24 February 2015, www.youtube.com/watch?v=V1gPyeRDBhM
32   Graham Stuart, 'Evidence Check', Institute for Government guest blog.

Figure 3: Summary of issues raised by contributors to the 'phonics' web forum
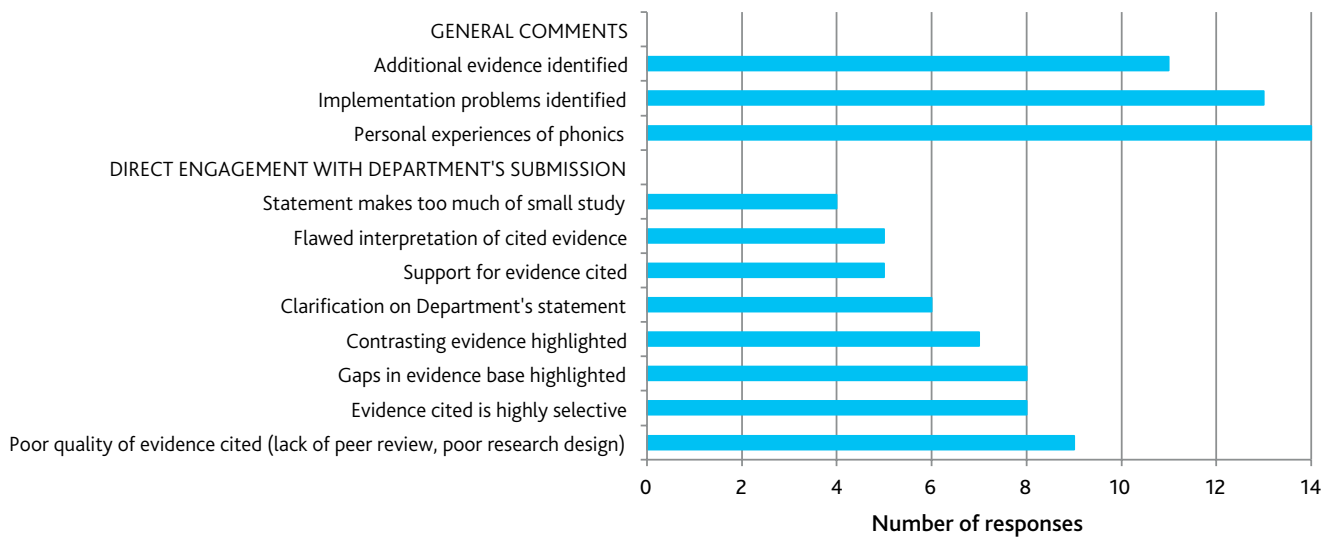
2. Departments need a framework to guide their responses

One way of ensuring that the Department's two-page submissions met a minimum level of disclosure would have been for the Committee to provide the Department with a framework with which to structure its submissions. One contributor to the web forum dedicated to the Department's use of evidence in policy making even offered a sample framework:

1. Full list of different types of relevant research evidence – preferably from systematic reviews with primary research where not covered by such systematic review.

2. The inclusion of both conceptual and empirical research.

3. The strength and relevance of this research for addressing the policy questions (and how these quality and relevance appraisals have been conducted).

4. Evidence gaps to inform further research investment.

5. How the research evidence is combined with the many other factors that policy makers have to take into account in order to make policy.[33]

It is for this reason that we designed a 'transparency standard' as part of our framework – offering clarity on what good practice looks like.

3. Crowdsourcing requires active engagement

Any crowdsourcing exercise requires active outreach. Evidence Check was designed to allow "commentators to get together to build on each other's comments and criticisms".[34] The exercise garnered plenty of input from campaign groups (e.g. Summer Born Campaign) and individual parents. But engagement from academics and professional organisations was more limited in some of the web forums, especially early on. As Jonathan Breckon, Director of the Alliance for Useful Evidence, explains:

The eternal problem with these "calls for evidence" is the biased sample of organisations and people that respond. The research community needs to be better informed and incentivised to take time to give written evidence – not just the usual suspects.[35]

---

33   David Gough, Education Committee web forum: Department's Use of Evidence, 12 December 2014, www.parliament.uk/business/committees/committees-a-z/ commons-select/education-committee/dfe-evidence-check-forum/departments-use-of-evidence

34   Graham Stuart, 'Evidence Check', Institute for Government guest blog.

35   Jonathan Breckon, personal correspondence.

This was a particular problem for the web forum devoted to the Department's use of evidence. Just over two-thirds of the issues raised by contributors (most of whom were parents) were criticisms of specific government policies (see Figure 4). Of these complaints, 91% related to the ban on term-time holidays. And almost none of these comments made any reference to the Department's use of evidence.

What is more, as a response posted by the Association of School and College Leaders (ASCL) indicates, the level of attention the forum attracted from interest groups may well have deterred professional organisations from submitting evidence:

As this forum is clearly aimed at individuals, and as it has very largely been used to disagree with a particular policy, this is not the place for ASCL to comment at length, however it would welcome the opportunity to inform the committee's consideration of the important topic of evidence-informed policy making.[36]

Figure 4: Summary of issues raised by contributors to the "departmental use of evidence" web forum
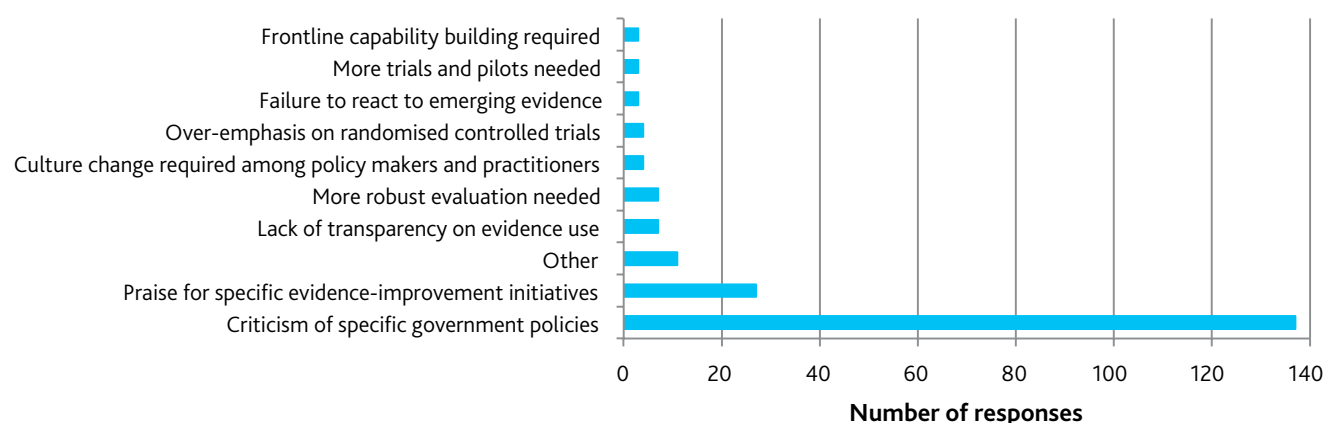
Active outreach proved critical to soliciting input from the research and practitioner community. As the period for submitting evidence wore on, the Committee went to considerable efforts to engage these individuals through a targeted use of social media, writing guest blogs[37] and leveraging the networks of organisations such as Education Media Centre and Royal Society. The Committee also benefited from (unsolicited) advertising from organisations such as the Alliance for Useful Evidence and the Institute of Education.[38]

Concluding thoughts

We need to see many more initiatives like Evidence Check. As the Institute for Government's Director, Peter Riddell, told the BBC last year, this is not about advocating for "rule by experts", but rather forcing policy makers "to explain why they have taken a decision and measure themselves against real evidence".[39] Parliamentary select committees have an important role to play in encouraging departments to disclose their evidence base behind policy decisions.[40] We are pleased that the Science and Technology Committee will undertake an evidence check exercise in this Parliament and hope our framework will act as a useful supporting tool.

36   Martin Ward (ASCL), Education Committee web forum: Department's use of evidence, 10 December 2014, www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/dfe-evidence-check-forum/departments-use-of-evidence/?page=2
37   See Graham Stuart, 'Evidence Check', Institute for Government guest blog.
38   See David Gough, 'Is research evidence informing government policy in education?', IoE London Blog, 10 December 2014, https://ioelondonblog.wordpress.com/2014/12/10/is-research-evidence-informing-government-policy-in-education/ and Jonathan Breckon, 'Your chance to vet the Department for Education's evidence', Alliance for Useful Evidence blog, 1 December 2014, www.alliance4usefulevidence.org/your-chance-to-vet-the-department-for-educations-evidence/
39   Peter Riddell, *World at One,* BBC Radio 4, 25 November 2014.
40   Hannah White, *Parliamentary Scrutiny of Government* (Institute for Government), 2015, www.instituteforgovernment.org.uk/publications/parliamentary-scrutiny-government