Social Indicators Research
https://doi.org/10.1007/s11205-020-02358-z

ORIGINAL RESEARCH

Understanding the (non-)Use of Societal Wellbeing Indicators in National Policy Development: What Can We Learn from Civil Servants? A UK Case Study

Christine Corlet Walker[1,2] · Angela Druckman[2] · Claudio Cattaneo[3]

Accepted: 16 March 2020

Abstract

Gross Domestic Product is often used as a proxy for societal well-being in the context of policy development. Its shortcomings in this context are, however, well documented, and numerous alternative indicator sets have been developed. Despite this, there is limited evidence of widespread use of these alternative indicator sets by people working in policy areas relevant to societal wellbeing. Civil servants are an important group of indicator end-users. Better understanding their views concerning measuring societal wellbeing can support wider discussions about what factors determine indicator use and influence in policy decision-making. Taking the UK as a case study, we ask what views exist among civil servants in the UK about measuring societal well-being? To answer this question, we used a bootstrapped Q methodology, interviewing 20 civil servants to elicit their views about measuring societal well-being. Three distinct discourses emerged from our analysis: one that was concerned about the consequences of ignoring natural, social and human capital in decision making; one that emphasised opportunity and autonomy as key determinants of well-being; and one that focused on the technical aspects of measuring societal well-being. Each of these discourses has direct implications for the way that we integrate societal well-being into policy making and highlights the potential benefits of including end-users in indicator development and strategy.

Keywords Societal well-being · Beyond GDP · Social indicators · Ecological indicators · Q methodology

✉ Christine Corlet Walker
c.corlet@surrey.ac.uk

[1] School of Geosciences, University of Edinburgh, Edinburgh, UK

[2] Centre for the Understanding of Sustainable Prosperity, University of Surrey, Guildford, UK

[3] Department of Environmental Studies, Faculty of Social Studies, Masaryk University, Brno, Czechia

Published online: 24 May 2020

1 Introduction

Gross Domestic Product (GDP) has been adopted by many national and international bodies over the last century as a proxy for the health and progress of a society (Kubiszewski et al. 2013; Van den Bergh 2009). However, it is widely acknowledged that this was never the intended purpose of the GDP indicator (Kubiszewski et al. 2013) and there have been countless efforts to devise better suited measures, which capture not only the economic, but the social and environmental components of our well-being too. Notably, initiatives like the OECD's 'Better life' initiative (OECD 2018), and the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz et al. 2009) have made large strides towards identifying, articulating and measuring what makes society prosperous, equitable and sustainable (Jackson 2010). Although there is no universally agreed definition of societal wellbeing, we situate our understanding of 'indicators of societal wellbeing' in the context of these initiatives. This, therefore, captures both objective and subjective notions of wellbeing and encompasses all those indicators that attempt to measure the progress of our societies and the health of our ecosystems. These indicators take a wide range of forms and foci, with ongoing debates in the literature focusing on the monetisation of nature and wellbeing, the use of objective versus subjective measures of wellbeing, and whether and how to aggregate fundamentally incommensurable measures (Barrington-Leigh and Escande 2018; Yang 2014). However, there is often an over focus in the literature on the technical characteristics of these new indicators, without due attention being paid to the ecosystem surrounding the indicator, including who the end-users are, how they interpret the indicators, and the role that the indicator and its end-user ultimately play in the policy-making process.

1.1 Use of Indicators in Policy Development

Much has been written about the policy process in different countries and policy domains. Authors have variously scrutinised the actors involved, the influence of power and politics in agenda setting (Gerston 2014; Birkland 2015), the (mis- or non-)use of different types of information and tools for designing and appraising policy (Marmot 2004), and what constitutes a policy cycle (Howlett et al. 2009), among other areas. Of particular interest to the scientific community has been the ways in which policy makers interact with and use different forms of evidence and information in policy making. From experiential or expert-based knowledge, to public surveys, ad-hoc scientific studies, assessments and indicators, there is a rich literature devoted to this issue (Bauler 2012; Weible 2008). Here we consider specifically the use of indicators of societal wellbeing by civil servants.

Civil servants fill a range of key roles in policy development, appraisal and implementation. Their ability to positively affect societal well-being through these roles is, in part, dependent on their ability to effectively absorb, translate and apply relevant evidence and information to the policy problems they face. Indicators in particular act as a succinct and accessible form of information with the ability to track trends across time and compare different sub-groups within the population. Given these characteristics, and the continued dominance of economic indicators such as GDP (Bell and Morse 2011), indicators of societal wellbeing may have an important part to play in centralising wellbeing and the environment in policy decision-making (Allin and Hand 2017, pp. 17). The analyses and inputs of civil servants are among many factors (e.g. public opinion, political agendas, financial constraints) considered by high level civil servants and government ministers, and indicators themselves are only one form of evidence that civil servants may choose to use. Nonetheless, understanding how and why civil servants use indicators in their work is one crucial facet of the policy process.

The literature on indicator use among policy makers has largely focused on issues of policy relevance or indicator content (Hezri and Dovers 2006), and on assessing the technical characteristics of the indicator, such as statistical robustness and accuracy (Lehtonen et al. 2016; Bauler 2012). Much of this research rides on the underlying assumption that indicators inform decisions in a direct and linear way; otherwise known as *instrumental use* of information (Lehtonen et al. 2016; Weible 2008; Hezri and Dovers 2006). However, where there is a high level of complexity and conflicting opinions—as there is in national-level policy making—such instrumental use of information is often impractical (Rinne et al. 2012). Instead, the information delivered through these indicators may lend itself more readily to *conceptual or political use* (Lehtonen et al. 2016; Bauler 2012; Hezri and Dovers 2006). In *conceptual use* of information, indicators operate as message carriers, shaping decision-makers' "frameworks of thought", rather than as direct tools for decision making (Lehtonen et al. 2016, p. 2). *Political use*, by contrast, describes the use of indicators in contributing to complex types of learning; for example, being used as ammunition to influence political agendas and to redefine problems (Lehtonen et al. 2016).

This distinction in types of 'use' is important because it shapes what we see as relevant in determining who uses indicators and how. In particular, conceptual and political use of indicators brings into focus the importance of the characteristics of end-users and the political conditions in which the indicator is deployed. For example, Sébastien and Bauler (2013, p. 3) note that user-factors such as the "expectations, belief systems [and] mental models" of policy actors may be more significant in determining the use and influence of sustainable development indicators at the EU level than their technical characteristics (Sébastien et al. 2014). Crucially, they (Sébastien and Bauler 2013, p. 5) also suggest that the degree of resonance between the mental models of the end-users and the way in which the indicator "frames the reality and the problems in question" may be a key determinant of the likelihood the indicator will be used and embedded at the collective level. Of course, this is only one part of the complexity that forms end-user characteristics, with indicator literacy, organisational information cultures, and other factors also forming important parts of the puzzle.

The concept of bounded rationality helps us to understand how mental models, or 'worldviews' may play a role in determining the use/non-use of information by policy-makers (Turnhout et al. 2007). We briefly define worldviews as "general social, cultural and political attitudes toward the world and 'orienting dispositions' that guide individual responses in complex situations" (Leiserowitz 2006). Individual actors, including civil servants, often fail to make *rational* decisions in complex decision environments because of cognitive artefacts or limitations (e.g. the inability to calculate complex trade-offs accurately, attentional deficits, the influence of emotion, habit and unreliable memory), which interfere with their decision processes (Jones 2002). This results in the use of cognitive shortcuts which aid decision making (Jones 2002). In particular, individuals may disfavour certain types of information over others. For example, information from sources external to their network (Rich 1991), or information which contrasts with their worldview (Zagorin 1998), may be more readily rejected. Bell and Morse (2011) find that practitioners and policy-makers themselves recognise the importance of these factors, with many noting that the success of indicators is partially determined by "who has developed the [indicator] and who is championing it" (Bell and Morse 2011, p. 292).

The use of short-cuts for deciding which information is more or less trustworthy, combined with the high error costs associated with making the 'wrong' decision at the national policy level, may lead policy-makers to be heavily critical of new information which does not resonate with their existing worldview (Turnhout et al. 2007; Collingridge and Reeve 1986). One result of this is the pursuit of "endless technical debates" between scientists and policy-makers, as neither party fully recognises the role that these end-user characteristics play in determining whether an indicator will prove acceptable to its intended users (Turnhout et al. 2007, pp. 223). Understanding the plurality of views that exist among civil servants may therefore be important in breaking this deadlock and designing indicators that are likely to have wider uptake.

1.2 Case Study

We take the UK as our case study for better understanding the (non-)use of indicators of societal wellbeing. The UK's Measuring National Well-being (MNW) programme was launched in 2010 by the Office for National Statistics (ONS) in order to "start measuring our progress as a country, not just by how our economy is growing, but by how our lives are improving" (Cameron 2010). The MNW programme collects and reports on a dashboard of 41 measures of well-being, covering personal well-being, relationships, health, what we do, where we live, personal finance, economy, education and skills, governance and environment (Office for National Statistics 2018). This work has been complemented by a number of companion programmes including the 'National Performance Framework' in Scotland (Scottish Government 2018) and the 'National Indicators for Wales' (National Assembly for Wales 2015).

While it certainly sits 'beyond GDP', the MNW framework still faces some major limitations as a way of measuring societal well-being. Of particular interest for this study, there is limited evidence of the widespread uptake and use of the indicators produced by the MNW programme in driving UK policy. The intentions of government in creating the MNW programme were explicitly focused on *measuring* well-being, with no clear commitments made about how the new measures would be used, and by whom[1] (Cameron 2010). Since its launch there have been only a handful of concrete examples of use of the MNW indicators to assess a specific policy problem (e.g. for the assessment of a series of airport schemes, Pwc 2014). In 2013, the UK government stated that "it should be emphasised that this is a long-term programme… and as such we should not expect to have examples of major decisions that have been heavily influenced by wellbeing at this stage" (GOV.UK 2013). Nevertheless, accounting for policy effects on wellbeing has certainly been encouraged more generally in recent years, both within government (e.g. HM Treasury's 'The Green Book' 2018) and by intermediaries (e.g. What Works Centre for Wellbeing's 'Wellbeing in Policy Analysis' 2018). This may have impacted attitudes towards, and use of, national indicators of wellbeing by civil servants. However, the lack of publicly available evidence and guidelines for indicator use in policy making means that it is still unclear whether and how things have progressed in the 7 years since that statement.

[1] The UK Government's Green Book discussion paper released in 2011 encourages the use of subjective-measures of well-being, specifically in policy cost–benefit analyses (Everett 2015; Fujiwara and Campbell 2011) but fails to go further in its commitments.

Despite the likely importance of understanding the views and underlying mental models of indicator end-users, there appears to be little research looking at the views of civil servants about measuring societal well-being. The ONS talk about "engagement with policy departments" during the development of the MNW programme (Matheson 2011, p. 20). However, the contents of this engagement appear not to be documented in the ONS archives, meaning that indicator developers and the broader scientific community are not able to utilise its insights. There is, therefore, a space for transparent analysis of what views civil servants hold about measuring societal wellbeing, and how these might be affecting their use (or not) of indicators.

Our study begins to address this gap in the literature by asking: what views exist among civil servants in the UK about measuring societal well-being? From this point, we then aim to reflect on whether these views are adequately catered for by the MNW programme or other indicators of societal wellbeing. For this task we used Q methodology; an interview-based methodology, lauded for its ability to explore and capture the diversity of views that exist among a group of stakeholders about a particular topic, in a formal way (Gall and Rodwell 2016; Steelman and Maguire 1999). Because of their position as a central group of indicator end-users, better understanding the views of civil servants about measuring societal wellbeing may also contribute more broadly to understandings of how we can improve the efficacy and universality of indicator use in policy making.

In the remaining sections of this paper we give a background to the methodology, including its benefits in the context of our study (Sect. 2), followed by a detailed account of our methods (Sect. 3). In Sect. 4 we present the results of our study. The significance of these results and their implications for measuring societal well-being in the UK and beyond are then discussed in Sect. 5, alongside some recommendations for future research.

2 Review of the Methodology

2.1 The Process

Q methodology is a quali-quantitative technique for eliciting the subjective views of participants about a topic, which are not ordinarily observable (Cross 2004), in a structured way (Gall and Rodwell 2016). It achieves this by presenting participants with a set of carefully constructed, opinion-based statements, known as the 'Q-set', which in theory represent the full array of views held about the topic (Watts and Stenner 2005). Participants are then asked to sort these statements into a grid which consists of a series of numbered columns labelled from 'least agree' to 'most agree' (or some variant thereof), according to how they feel about the statement (Watts and Stenner 2005).

The grid shape, or distribution, is selected by the researcher and often takes a quasi-normal shape, with columns at the extremes of the grid holding fewer statements than those in the middle (see Fig. 1).

Participants' 'sorted' grids (i.e. those for which one statement has been assigned to each grid cell) are analysed using Principle Component Analysis (PCA). PCA identifies similarities in the way that participants have sorted the statements, resulting in a set of participant groupings, or 'factors' (Watts and Stenner 2005). Information about each factor is then brought together with any qualitative data collected from interviews with participants to develop a 'discourse' (i.e. text that describes the views held by the participants associated with that factor). This process is detailed in Fig. 2.

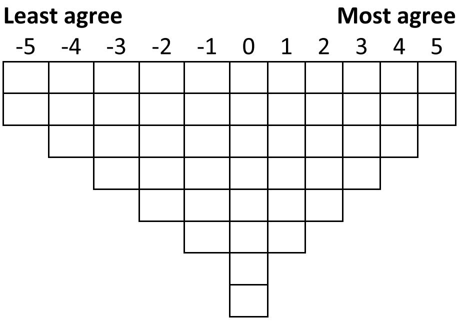

Fig. 1 This figure shows an example quasi-normal grid distribution for Q statements to be sorted into. Each statement is given a number, and one number is allocated to each cell. For example, in this grid, only 2 statements can be sorted into the least agree (– 5), and most agree (+ 5) columns

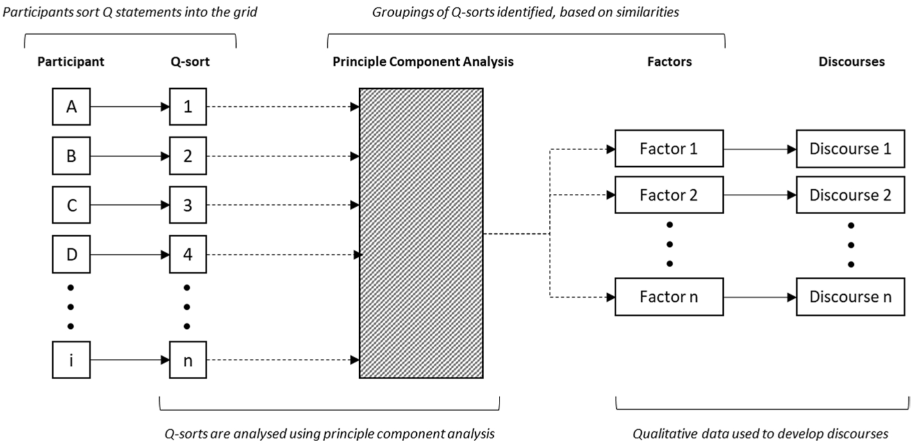
Participants sort Q statements into the grid
Groupings of Q-sorts identified, based on similarities
Participant
Q-sort
Principle Component Analysis
Factors
Discourses
A
B
C
D
i
1
2
3
4
n
Factor 1
Factor 2
Factor n
Discourse 1
Discourse 2
Discourse n
Q-sorts are analysed using principle component analysis
Qualitative data used to develop discourses

Fig. 2 Q-study procedure, from statement sorting to discourse analysis

2.2 Can it Really Work?

Q methodology assumes a finite diversity in the ways that people express their views (Cross 2004), meaning that there are a limited number of discourses in circulation about a topic at any one time. This leads Q researchers to claim that the methodology can identify the full range of existing views held by a population about a specific topic, using a relatively small sample size (Brown et al. 1999; Stainton Rogers et al. 1995; Brown 1980). This idea is reflected in the literature, with more than a third of Q studies published in the last 10 years using fewer than 30 participants ("Appendix 1"). Central to this point is the argument that "Q methodology has no interest in estimating population statistics" and so has no need for a large or representative sample of participants (Cross 2004, p. 210). Instead it is more important to prioritise a diverse sample of participants likely to hold differing views (Zabala and Pascual 2016; Cuppen et al. 2010). Further, Q is considered to be structurally different to traditional R methodology, with the Q-set forming the equivalent of the 'sample', and the participants instead representing something akin to the 'experimental condition' (Cross 2004). In this way, criticisms based on sample size are often considered misguided (Brown et al. 2015).

The methodology has also been criticised as "impotent" to find all existing opinions within a population, owing to the limited nature of the Q-statements as compared to the potentially infinite nature of the opinion domain (Kampen and Tamás 2014, pp. 3113; Cross 2004). However, Q methodology is a scientific tool, and as such there are, of course, limitations to both its accuracy and precision (Brown et al. 2015). This cannot fairly be levelled as a criticism against it. The more important question is whether the outputs of the study can be considered useful and reliable. More statements could be added to the Q-set to increase the 'precision' with which participant views are characterised. However, large numbers of statements can result in participant fatigue, risking the reliability of the study. In any case, the purpose of most Q studies is to identify broad commonalities in the viewpoints held by individuals in a population (Brown et al. 2015), with qualitative interviews providing more detailed information, where needed. Even those studies with large numbers of Q-statements rarely identify more than 3−6 distinct views (Brown et al. 2015), validating the position that current 'best practice' applications of Q methodology are perfectly adequate to meet their aims.

It is also important to note that the ability of a Q-study to capture the full range of opinions that exist within a population will *in practice* depend on a number of other factors, in addition to the number of statements presented to participants. For example, the construction of the statement set by the researcher (i.e. is it thorough and does it represent the diversity of discourses currently in use?), the size of the population, and the degree of heterogeneity of opinions within it, will all affect the efficacy of a Q-study. Variability in these factors is a limitation of Q methodology, not because a single study may not capture the full diversity of opinions within a population, but rather because it is difficult to confirm the validity of the results through further research. That is to say, we could only attempt to confirm that we have captured the full diversity of views within a population by conducting a Q-study with a very large number of statements, involving the whole population. Importantly, this limitation does not undermine the views that *are* revealed through the study, which still themselves represent valid expressions of opinion that exist within the population, given the set of statements presented to the participants. Rather it is a limitation that should be considered from the outset when deciding on the desired outcomes of a Q-study. In particular, if the study is exploratory in nature, there is no reason this limitation should present a barrier to such research, although it should be taken into consideration when drawing conclusions.

2.3 Example Applications

Q methodology has been used widely to inform policy development, most commonly in relation to specific environmental management issues (Ockwell 2008; Ellis et al. 2007; Steelman and Maguire 1999). However, it has only rarely been used in the development or appraisal of social, environmental and economic policy indicators, as we do here. Of particular relevance to our study, Doody et al. (2009) sought to identify publicly acceptable sustainable development indicators in the UK. Using Q methodology, the authors were able to identify key areas of concern for the public, and areas that appeared to be irrelevant or of little interest. This ultimately enabled them to develop indicators which better reflected the views of the public (Doody et al. 2009).

Doody et al. (2009)'s study highlights two significant benefits of using Q methodology for investigating a complex and multi-faceted issue, such as measuring societal well-being. First, participants can make clear and nuanced prioritisations by integrating complex trade-offs implicitly into their internal decision-making process (Zabala and Pascual 2016). Second, by presenting all participants with the same set of opinion statements, analysts can directly compare the views of participants on *all* of the issues covered. This allows for the identification of specific areas of consensus and conflict (Steelman and Maguire 1999), which can guide future research and indicator development. Specifically, by moving beyond 'practiced' rhetoric on a topic, which is often elicited using more traditional interview techniques, it becomes more straightforward to identify areas of common ground to bridge between differing views. This characteristic of Q has proven to be particularly useful in assessing environmental policy where there is pre-existing conflict (Barry and Proops 1999; Van Eeten 2000). These strengths make Q methodology a strong candidate for investigating the range of views that exist about measuring societal well-being within the UK civil service.

3 Methods

3.1 Study Design and Data Collection

Q is a flexible method that can be implemented in many different ways, from the types of items being sorted (e.g. O'Neill et al. 2013 used images instead of statements) to the interview technique and selected grid shape. For this reason, transparency is a key element of Q studies. We have therefore included a table below detailing each design component of this study and a justification for our selected approach (Table 1). In brief, participants were given a set of statements (the Q-set) which reflected the central debates in the literature, the media and among civil servants themselves around measuring societal wellbeing. We asked participants to sort these statements into an 11-column grid, ranging from −5 (least agree) to +5 (most agree) (as per Fig. 1 above). This process results in one completed grid, or 'Q-sort', per participant (see Fig. 2 for diagram detailing the process). After the sorting exercise, each participant was interviewed to provide context to the quantitative results.

3.2 Data Analysis

We conducted a Principle Component Analysis of the completed grids, or Q-sorts (see "Appendix 3" for full R code). The PCA identified clusters in the way that participants sorted their statements into the grid. Each of the identified clusters, or 'factors', represents a distinct group of Q-sorts, reflecting participants with similar views on the study topic (Zabala and Pascual 2016). In order for each of the factors in the PCA to be considered distinct from one another (i.e. that they each represent a genuinely unique view point), they must all meet the set of criteria laid out in Table 2.

Once the final number of factors was decided on, a representative Q-sort was constructed for each factor. This reflects the *mean* view of the participants associated with the factor (Zabala and Pascual 2016). 'Distinguishing' and 'consensus' statements were also identified for each factor at this stage. Distinguishing statements are those statements (from the Q-set) for which one factor's mean positioning of that statement in the sorting grid is significantly different from the other factors' positioning of the same statement, at the 5% level. Consensus statements, by contrast, are those statements for which each factor's positioning of the statement was not significantly different from one another; in other words, their views on the statement were not distinguishable. The idealised Q-sorts, distinguishing and consensus statements, along with the qualitative interview data provided by

Table 1 Q methodology study design decisions

Design component	Description and justification
The concourse	The concourse represents, in theory, the full suite of opinions and arguments communicated between individuals in a population about a particular topic (van Exel and de Graaf 2005). A range of sources were used in its development, including (1) academic literature[a], (2) social media, (3) newspaper articles, (4) parliamentary debates, (5) pre-existing interviews with civil servants and politicians, and (6) pilot study participants. A full list of sources can be found in "Appendix 2". The breadth of sources was used to bridge the gap between what civil servants are talking about, and what solutions are being discussed in the literature. This resulted in 401 relevant statements being collected from the sources analysed, which ultimately formed the unabridged concourse
The Q-set	The Q-set is the short-list of statements selected from the concourse, which are considered to be representative of the full opinion domain. To create the final Q-set, our research team filtered the concourse of 401 statements, applying the following criteria for what makes a good Q statement: the statement should represent a single, targeted opinion; it should be stand alone; it should be easy to understand; and, it should have some multiplicity in possible interpretations (Watts and Stenner 2005). We aimed to represent the full range of opinions encountered about societal well-being, but without repetition, so as to avoid participant fatigue during the sorting process. This process ultimately resulted in a Q-set comprised of 48 statements
Grid distribution	Brown (1980) found that the distribution of the grid has "virtually nil" effect on the factor analysis outcomes (Watts and Stenner 2005, pp. 77). The 'forced' distribution serves instead to make the sorting process easier for participants to interact with. In particular, participants with more strongly-formed, well-articulated opinions may benefit from a shallower distribution to enable greater differentiation between statements (Van Exel and De Graaf 2005). In light of this, and the high level of expected knowledge of our participants, we opted to use a quasi-normal shape, 11-column grid, as per Fig. 1
Pilot study	A short pilot study was conducted with six well-informed students to ensure that the selected opinion statements (or Q-set) were comprehensible and well-balanced. Pilot participants went through the full Q interview procedure (detailed below) and provided feedback about the statement sorting process, the clarity of the statements, and any subject areas they felt were not adequately covered
The P-set	The P-set is the participant sample (Baker et al. 2006). In our case, the P-set consisted of twenty UK civil servants involved in the policy design, implementation or appraisal process. Certain departments were targeted because their work was most directly relevant to the study subject, and they were considered most likely to have reason to use alternative indicators of societal well-being in their decision-making. Our participants were elite, and as such had a set of challenging characteristics. For example, they were difficult to contact and unlikely to respond if contact information was available (Lancaster 2017). For this reason, we used snowball sampling to identify further participants. Although this was the most appropriate method, it also acted as a limitation because we were dependent on our participants' contacts. This meant that we were unable to sample from all relevant departments

Table 1 (continued)

Design component	Description and justification
The statement sorting procedure	One-on-one interviews were conducted in person or over the phone, during which participants sorted the Q statements into the grid provided. Interviews were conducted between July and October 2017. For the telephone interviews, the standard interview methodology was adapted: participants conducted the statement sorting process using a specially-developed Excel tool, whilst being guided over the phone by the researcher
Qualitative interviews	Qualitative data can also be collected from participants as part of a Q-study, to provide context to the sorted statements. We asked why participants sorted the statements as they did, focusing on the two extremes of the distribution grid, as these have the most impact on the outcome of the PCA. Participants were also asked whether they thought there were any opinions about measuring societal well-being not included in the statements they sorted, and whether they had cause to use indicators as part of their job. All interviews were recorded, and data were transcribed for analysis by the lead analyst

[a] A literature review was conducted in Web of Science on 24/05/2017, using the search terms: **TITLE:** ([measure* OR indicator*]) AND **TITLE:** ([social OR societal OR society OR human OR economic]) AND **TITLE:** ([welfare OR well-being OR progress])

Table 2 Criteria used for factor extraction (Zabala and Pascual 2016; Davies and Hodge 2007)

Criteria	Threshold
Number of significantly loading sorts	Each factor must have two or more Q-sorts which significantly load onto it, after confounded Q-sorts have been removed[a]. In order for a Q-sort to 'significantly load' onto a factor, it must have a loading score greater than a particular threshold, called the 'significance level'[b]. Additionally, the square of the Q-sort's loading score for the factor in question must be greater than the sum of its loading values for all other factors
Eigenvalues	The Eigenvalues for each factor must be greater than 1
Explanatory variance	The sum of the explanatory variances for all extracted factors must be greater than 35%
Humphrey's rule	The cross product of the two highest loading Q-sorts must be greater than two times the standard error
Correlation between factors	Correlation between factors should ideally not be greater than the significance level

[a] Confounded Q-sorts are those Q-sorts which significantly load onto more than one factor, and hence cannot be considered uniquely associated with one particular factor. These do not contribute to the final discourses as they are not exemplary of a single factor

[b] The significance level is calculated as $2.58 \times \frac{1}{\sqrt{n}}$, where n = the number of Q statements (Watts and Stenner 2005)

participants, formed the basis for discourse construction. One discourse was developed per factor extracted from the PCA.

Table 3 Breakdown of participants, by government department

Department	
Department for Environment, Food, and Rural Affairs (DEFRA)	5
Department for International Development (DfID)	5
Her Majesty's Treasury (HM Treasury)	3
Home Office	3
Local Government	2
Business, Energy and Industrial Strategy (BEIS)	1
Scottish Government	1
Total	20

3.3 Reliability Testing

We used reliability testing here to better understand how stable our Q-study results were; i.e. how consistent the PCA outputs were under repeated samples. We chose a bootstrapping methodology which allowed us to calculate distributions and new standard errors for various key statistics, such as factor loadings and z-scores (Zabala and Pascual 2016) (See "Appendix 4" for a detailed explanation of the bootstrapping methodology). This enabled us to calculate more accurate measures of reliability through repeated re-sampling and replacement of Q-sorts (Zabala and Pascual 2016).

We opted for 1000 bootstrap repetitions, in line with recommendations of "at least 40 times the size of the sample" (Zabala and Pascual 2016, pp. 8). Because this Q-bootstrapping methodology is relatively new, and because our sample size is less than the 45 Q-sorts recommended to achieve *highly* accurate results (Zabala and Pascual 2016), we used the bootstrapping results primarily as a guide for interpretation. Hence, although we used the bootstrapping results to inform discourse development, we reported both the standard and bootstrapped PCA results, and supported the discourse development with the qualitative interview data. Further, we relaxed the range for Q-sort instability, such that a Q-sort must be flagged in between 20 and 75% of repetitions to be considered unstable.[2] This reflects our cautious approach to using this new methodology.

4 Results

4.1 Factor Scores and Distinguishing Statements

Forty-eight statements were selected from the concourse to form the final Q-set to be sorted by participants (see Table 1 for Q-statement selection criteria; see "Appendix 5" for list of statements and breakdown by topic area). Thirty-five UK civil servants were contacted for participation in the study. We obtained a 59% response rate, with 20 civil servants ultimately taking part from a range of departments (see Table 3). Participants had a variety of job roles, largely focused on policy design, implementation and appraisal in

[2] Zabala and Pascual recommend that a Q-sort be considered unstable if it is flagged for a factor in between 20 and 80% of bootstrap repetitions (Zabala and Pascual 2016).

Table 4 Q statements and their standard (std.) and bootstrapped (bts.) factor scores, for each of the three factors (f1, f2 and f3)

Statements		Factor scores					
		f1		f2		f3	
		Std.	Bts.	Std.	Bts.	Std.	Bts.
1	To track societal welfare, we should be measuring growth in total wealth, crucially including natural (e.g. stock of forests), social (e.g. interpersonal relationships) and human (e.g. literacy skills) capital	4**	5*	1		1	2
2	It is crucial to acknowledge the importance of natural capital (e.g. green space, proportion of forest cover, water quality) for societal well-being in any composite measure	5**	4*	1		1	
3	There are certain goods and services that the environment provides, which are important for societal well-being, that cannot be replaced by man-made goods and services	3^{Δ}	$^{\Delta}$	3^{Δ}	$^{\Delta}$	3^{Δ}	$^{\Delta}$
4	We should include the value of natural capital in a composite indicator so that decision makers can better include the environment as they allocate resources to promote the growth of the economy	4**	*	1		1	2
5	The scope of societal welfare is too broad and subtle for a single measure to evaluate satisfactorily	2	1	1		5**	*
6	Using a composite indicator to create a bottom-line is more useful in gathering the interest of media and policy-makers than the use of sets of indicators	0	$^{\Delta}$	0	$^{\Delta}$	2	1^{Δ}
7	Societal welfare cannot be considered simply the sum of individual welfare, as this ignores certain important synergies and thresholds	1**	*	− 1**	*	5**	*
8	The UK government should focus primarily on devising policies that minimise unemployment, with concerns about maximising GDP coming later	-1^{Δ}	$^{\Delta}$	-1^{Δ}	$^{\Delta}$	0^{Δ}	$^{\Delta}$
9	The quality and security of available jobs is a key determinant of societal well-being	1**		4		4	
10	Ecosystem interactions with human well-being (e.g. recreation time in nature, provision of food, cultural significance) have no place in a quantitative measure of societal well-being	-3^{Δ}	-4^{Δ}	-2^{Δ}	$^{\Delta}$	-3^{Δ}	$^{\Delta}$
11	When measuring environmental well-being, we should focus on the measurement of final ecosystem goods and services (e.g. quantity of fish provided), rather than the ecosystem functions that underpin them (e.g. quality of stream habitat)	-2^{Δ}	$^{\Delta}$	-2^{Δ}	$^{\Delta}$	-3^{Δ}	-2^{Δ}
12	GDP per capita alone does not capture how successfully most individuals can access the resources required for a decent standard of living, for this we need some measure of economic security	3^{Δ}	$^{\Delta}$	3^{Δ}	$^{\Delta}$	3^{Δ}	$^{\Delta}$
13	We should be assessing quality of life in terms of the opportunities people have to achieve well-being (e.g. access to education, access to healthcare), rather than whether or not they actually achieve it (e.g. literacy rates, life expectancy)	− 1		4**	*	− 3	− 4

Table 4 (continued)

	Statements	Factor scores					
		f1		f2		f3	
		Std.	Bts.	Std.	Bts.	Std.	Bts.
14	Indicators should be focused on well-being outcomes (e.g. health status), as opposed to inputs (e.g. health-care expenditure)	0**	*	5		4	
15	An absolute increase in GDP can be considered broadly synonymous with enhanced societal well-being, even in wealthy nations	−2		0*		−2	
16	We should focus on making technical changes to the way GDP is calculated, so that it more accurately reflects welfare, rather than developing new indicators	0		0		−2	−3
17	GDP is the most appropriate and reliable measure we have to capture societal well-being	-4		0**	*	−5	
18	GDP should be adjusted to include income disparities	1^{Δ}	Δ	0^{Δ}	Δ	1^{Δ}	Δ
19	Adjusting GDP to account for government borrowing would provide a better idea of how financially sustainable our social progress is	1	0	0		−5**	*
20	We need to better capture the contribution of non-traditional economic activity (e.g. open source information sharing, technological innovation, the gig economy) to societal well-being	2	Δ	2	Δ	4*	Δ
21	Economic growth is the essential foundation of all our well-being aspirations	-3^{Δ}	-2^{Δ}	-3^{Δ}	Δ	-4^{Δ}	-3^{Δ}
22	Government policy should prioritise economic growth over ill-defined concepts of sustainability and societal well-being	−5**	*	−2		0	
23	Quality of life is strongly associated with material wealth, and the well-being you derive from having material possessions	-1^{Δ}	Δ	-1^{Δ}	Δ	-1^{Δ}	Δ
24	People’s wealth in a given period is a better measure of their well-being than yearly income flows	0^{Δ}	Δ	-1^{Δ}	Δ	0^{Δ}	Δ
25	Inequality is currently too poorly defined and too difficult to measure to be included in national accounts	−1	-2^{Δ}	−4	Δ	−4	Δ
26	If well-being indices are meant to capture non-economic well-being and determine the real inclusive and equitable growth of society, they should also move beyond the measurement of purely economic inequality	3^{Δ}	Δ	4^{Δ}	Δ	3^{Δ}	Δ
27	Inequality has less of an impact on well-being in wealthier countries, therefore we do not need to include it in a UK measure of societal well-being	-4^{Δ}	Δ	-5^{Δ}	-4^{Δ}	-4^{Δ}	Δ
28	Comparisons of living standards over time need to take into account the amount of leisure that people enjoy	0	1^{Δ}	2	Δ	−1*	0^{Δ}

Table 4 (continued)

Statements		Factor scores					
		f1		f2		f3	
		Std.	Bts.	Std.	Bts.	Std.	Bts.
29	All aspects of well-being can be fairly expressed in monetary terms	− 4	− 3$^{\Delta}$	− 4	Δ	0**	− 1$^{\Delta}$
30	Focusing governmental interventions on enhancing GDP, as a way to improve welfare, is problematic because it can have unintended, negative consequences	2	Δ	3	Δ	1	Δ
31	In any measure of societal welfare, we must account for the social and environmental damage caused by economic activity (e.g. deforestation from crop production, pollution from energy generation, lung cancer from tobacco sales)	5*	*	3		2	
32	It is not possible to accurately value non-market goods (e.g. volunteering, unpaid care, national forests), and to do so in a measure of societal welfare would be misleading	− 2$^{\Delta}$		− 1$^{\Delta}$	*	− 1$^{\Delta}$	
33	Non-market goods and services (e.g. volunteer work and unpaid care) form an invisible pillar of our economy, and should therefore be included in a measure of societal well-being	4		2		3	
34	The role of strong interpersonal relationships and social networks (e.g. local communities) in contributing to well-being is over-played	− 2	− 3$^{\Delta}$	− 3	Δ	− 1*	Δ
35	People derive more well-being from their existence within a community than from personal consumption	2		2		0	
36	The government has a critical role to play in promoting stable relationships and good parenting	1$^{\Delta}$	2	1$^{\Delta}$		0$^{\Delta}$	*
37	Social contacts and trust play an important role in expanding markets and increasing income	1$^{\Delta}$	Δ	1$^{\Delta}$	Δ	2$^{\Delta}$	Δ
38	Although human rights (e.g. freedom of speech, the right to vote) are fundamental to well-being, they do not need to be measured in the UK	− 3$^{\Delta}$	Δ	− 3$^{\Delta}$	Δ	− 2$^{\Delta}$	Δ
39	People who feel empowered and in control of their own destiny feel more fulfilled, so we should be measuring freedom and ability to choose across multiple domains (e.g. healthcare, education, housing, faith)	3**	2	5**		0**	
40	We should measure the government's success in abating deficiencies (e.g. unemployment), rather than fulfilling desires (e.g. achieving a degree)	− 1$^{\Delta}$	Δ	− 2$^{\Delta}$	Δ	− 1$^{\Delta}$	Δ
41	You can't legislate for fulfilment or satisfaction, and therefore we should not include it in a government measure of societal well-being	− 2	− 1$^{\Delta}$	− 4	-5$^{\Delta}$	− 3	Δ
42	Government should seek to describe societal well-being using subjective measures rather than prescribe it by measuring things we presume to be good for well-being (e.g. education level or health status)	0$^{\Delta}$	Δ	0$^{\Delta}$	Δ	1$^{\Delta}$	Δ

Table 4 (continued)

Statements		Factor scores					
		f1		f2		f3	
		Std.	Bts.	Std.	Bts.	Std.	Bts.
43	Subjective assessment of well-being can be affected by many factors (e.g. cultural background, socio-economic context and even participant mood), therefore it is not suitable to be used within a national measure of societal well-being	− 1		− 3		− 1	
44	The state of housing in the UK should primarily be captured through objective measures (e.g. proportion of houses in each council tax band), not through subjective measures such as the Gallup question, which asks the public whether they feel there is enough, good quality housing available	0		0		− 2**	
45	Using subjective well-being measures as an indicator of societal welfare could lead society to over-value the ability to be happy or contented with one's "lot in life", no matter how limiting or inequitable that lot is	0		− 1*	*	2	1
46	Any measure of sustainability needs to keep track of environmental processes that may be irreversible or non-linear over certain time-lines	2^{Δ}	3^{Δ}	2^{Δ}	$^{\Delta}$	2^{Δ}	$^{\Delta}$
47	We do not need to be concerned about sustainable well-being for future generations whilst we still have the well-being of the current generation to worry about	− 5		− 5		− 2**	*
48	We do not have an agreed definition for sustainability, so we shouldn't try to include it in our national accounts yet	− 3		− 2		0**	*

Scores range from − 5 (strongly disagree with the statement) to + 5 (strongly agree with the statement). Bootstrapped factor scores are only shown if they are different from standard factor scores. Distinguishing statements are signified with asterisks: * denotes $p < 0.05$ and ** denotes $p < 0.01$. Consensus statements are signified with a triangle. Nb. bootstrapped statements were only analysed at the 5% level societal wellbeing relevant domains. Thirteen respondents were classified as mid-level civil servants, and seven as senior-level.[3] Twenty Q-sorts (or sorted grids) of 48 statements each were therefore analysed using a standard *and* bootstrapped PCA. From the standard PCA (i.e. without bootstrapping repetitions) we found that a three-factor solution met all the relevant criteria for extraction (see Table 2 for extraction criteria; see "Appendix 6" for full factor results against each extraction criteria). Eighteen out of the 20 Q-sorts loaded significantly onto one of the three factors, and two Q-sorts were confounded. Together the three factors accounted for 72% of the study variance, well above the threshold of 35% set out in Table 2 (see "Appendix 7" for full bootstrapping results, including bootstrapped factor scores and standard errors).

The factor scores calculated for each Q statement against each factor, in both the standard and bootstrapped PCAs, can be found in Table 4. Distinguishing and consensus statements from the standard and bootstrapped PCAs are also shown here. After applying the bootstrapping procedure, we found a number of unstable statements associated with each factor, whose factor score or status as a distinguishing or consensus statement changed (Table 4). Importantly, our analytical choice to use these bootstrapped factor scores in place of the standard factor scores when developing the discourses (see Sect. 3.3) did not dramatically change the interpretation of the factors. In particular, factors 1 and 2 were largely unaffected. However, it did lead to a slightly different emphasis for factor 3, with six distinguishing statements becoming no longer distinguishing. The bootstrapping analysis also highlighted a number of unstable Q-sorts with large standard errors or ambiguous flagging frequencies. The significance of these unstable Q-sorts is discussed in Sect. 4.2.

4.2 Discourses

Qualitative data was collected from 18 of our Q-sort participants[4] and used to aid construction of the final three discourses. Below we give brief summaries of each of the discourses; full discourses can be found in "Appendix 1", alongside discussions of the implications of any unstable statements and Q-Sorts.

4.2.1 #1 The Socio-Environmental Discourse

This discourse is defined largely by a concern that measurement of, and decision-making about, societal well-being should include the full range of natural, human and social capital; taking proper account of the potentially damaging effects of economic activity on each of them, in both the short and long term. Factor 1 formed the basis for this discourse, for which summary information can be found in Table 5.

Participants who loaded onto this factor were concerned that GDP does not capture a holistic view of the world around us (S1: +5*).[5] In particular, they showed concern that certain elements of value generated by the environment are overlooked (S2: +4*), and

[3] As per the Institute for Government's classifications, we defined mid-level civil servants as grades EO, HEO and SEO, and senior-level civil servants as G7, G6 and the Senior Civil Service (Institute for Government 2018).

[4] Two participants declined to provide additional qualitative data.

[5] S_ gives the statement number. The notes in brackets, therefore, indicate that statement S1 has a factor score of +5 for factor 1. The asterisk shows that it is a distinguishing statement for this factor. See Table 4 for a full list of statements and associated factor scores, against each factor.

Table 5 Summary information for factor 1

Characteristic	Description
Number of significantly loading Q-sorts	Ten Q-sorts loaded significantly onto this factor
Study variance accounted for by factor	33% of study variance is accounted for by this factor
Participant characteristics	Five participants were from DfID, three from DEFRA, one from the Home Office and one from Local Government
Unstable Q-sorts	Bootstrapping analysis revealed that six of the ten significantly loading Q-sorts could be considered unstable, including P3[a], P4, P9, P11, P13 and P14. Five of these Q-sorts flagged onto the factor in more than 60% of repetitions. P13, however, had a flagging frequency of just 52%, which is very far from our 75% threshold for instability, indicating that this Q-sort is not at all exemplary of the factor
Distinguishing statements	This factor was characterised by 9 distinguishing statements before bootstrapping, and 7 after
Unstable statements	Statements 9 and 39 were no longer distinguishing after bootstrapping

[a] P_ gives the participant number

strongly supported better integration of the value of natural capital into decision making (S4: +4*). In support of these ideas, participants commented that:

> We should be measuring economic growth, but also natural capital, social capital, human capital. That just gives you a much more well-rounded view of society as a whole (Participant 14)

> When things like health and education are clearly so important and so immediate, I think there's a danger of some environmental things getting left out of the assessment of how we're doing as a society (Participant 11)

> While GDP remains (wrongly in my view) the indicator of choice of wellbeing it should at least include a value for the resources used so that sustainability is more central to policy making (Participant 12)

In this vein, the participants who loaded onto this factor strongly believed that building sustainable well-being is not only important but in fact necessary, both for future generations and for current generations too (S47: −5). Even when challenged with the idea that the concept of sustainability may be poorly defined (S22: −5*), these participants felt that:

> Current sustainable well-being and future sustainable well-being are inextricably linked and if we make bad decisions… now, the impact for current and future generations is significant (Participant 6)

> Although the concept of sustainability [is] ill-defined, [it is] crucial to understanding the state of our population, and we should make work to define [it] further rather than disregard [it] (Participant 3)

Additionally, those who were associated with this factor drew attention to the need to take proper account of the damage caused by economic activity (S31: +5*), such as the negative health effects of the tobacco industry.

> It seems to me that GDP and some other indicators or measures of progress completely neglect the damage that we cause in the process (Participant 11)

> The tobacco industry is doing absolutely no good whatsoever for society, and yet being propped up and… allowed to function… Even doctors argue for it at times, reasoning that the taxes people pay on cigarettes funds the NHS. I think this is fundamentally twisted and flawed logic and we need to seriously re-think our society (Participant 6)

4.2.2 #2 The Self-determination Discourse

This discourse is defined by the strong belief that access to opportunity and the ability to define one's own destiny are key determinants of wellbeing. Factor 2 formed the basis for this discourse, for which summary information can be found in Table 6.

Participants who loaded onto this factor felt strongly that being empowered to make choices about your own destiny was central to well-being (S39: +5). This was exemplified by the quote:

> I think the key to happiness and 'well-being' is being in control of your own life and feeling as though you have the freedom to influence its direction and outcomes (Participant 18)

This concept was also reflected in their opinion that quality of life should be assessed in terms of the opportunities people have to achieve well-being, rather than whether or not they actually achieve it (S13: +4*). This came from two distinct perspectives, one reacting against the idea of a 'nanny state'—"I think it's patronising to kind of prescribe 'this is what makes people happy" (Participant 2)—and another advocating the idea that a "[level] playing field" in terms of access to opportunity is key for societal well-being (Participants 7 and 8).

Table 6 Summary information for factor 2

Characteristic	Description
Number of significantly loading Q-sorts	Five Q-sorts significantly loaded onto this factor
Study variance accounted for by factor	23% of study variance was accounted for by this factor
Participant characteristics	Two participants associated with this factor were from the Home Office, one was from the HM Treasury, one from local government and one from BEIS
Unstable Q-sorts	No bootstrapped factor scores had ambiguous flagging frequencies, indicating that all exemplary Q-sorts identified through the standard PCA were strongly representative of the factor
Distinguishing statements	This factor was characterised by 6 distinguishing statements before bootstrapping, and 5 after
Unstable statements	Statements 15 and 39 were no longer distinguishing after the bootstrapping procedure, but statement 32 became distinguishing

These participants also placed an emphasis on economic and job security (S9: +4, S12: +3), which is consistent with the ideas above about the importance of autonomous decision-making, commenting that:

> The availability of a job lets you access all the other [elements of well-being] that might be measured. [For example], without a job you might not have the social life that you want… or [be able to] raise your children how you want (Participant 5)

This discourse was further distinguished by a more favourable view on subjective measures of well-being than the other two factors, which again supports the ideas expressed above that people generally know what is best for them. This was manifest in participants disagreeing that subjective measures were unreliable and might lead people to be contented with their ‘lot in life’, no matter how bad (S45: −1*, S43: −3). One participant stated:

> I objected to the ones that suggested you shouldn’t trust people to know what they’re talking about when they give subjective opinions… Particularly when you aggregate them all, despite variations, they probably know what they’re saying (Participant 8)

Participants associated with this factor were also distinct from those associated with other factors in their consistent indifference towards GDP as a measure of societal well-being, and any adjustments to it (S15: 0; S16: 0; S17: 0*; S18: 0; S19: 0).

4.2.3 #3 The Technocratic Discourse

Participants associated with this factor give close attention to the technical difficulties of measuring societal well-being and the potential pitfalls of trying to alter GDP. Factor 3 formed the basis of this discourse, for which summary information is included in Table 7. Of note, only one of the three Q-sorts associated with this factor using the standard PCA procedure was found to still be exemplary after bootstrapping. This calls into question the status of this factor as representing a unique view point (as per the extraction criteria in Table 2). However, closer inspection of the qualitative data justifies maintaining three

Table 7 Summary information for factor 3

Characteristic	Description
Number of significantly loading Q-sorts	Three Q-sorts significantly loaded onto this factor
Study variance accounted for by factor	17% of study variance was accounted for by this factor
Participant characteristics	Two participants were from HM Treasury and one was from DEFRA
Unstable Q-sorts	The bootstrapping analysis revealed that two out of three Q-sorts which significantly loaded onto this factor were unstable. The loading scores for Q-sorts P1 and P10 had standard errors of 0.37, the largest across all Q-sorts. They also had ambiguous flagging frequencies of 0.65 and 0.48, respectively, indicating that they are not strong representatives of the factor. This leaves just one Q-sort as a clear exemplar of factor 3
Distinguishing statements	This factor was characterised by 11 distinguishing statements before bootstrapping, and 6 after
Unstable statements	Statements 20, 28, 29, 34, 39, and 45 were no longer distinguishing after bootstrapping, but statement 36 was

factors (instead of dropping to two). The interview data clearly supports the idea that factor 3 brings a unique perspective when compared to the other two factors but leads us to cautious interpretation of the factor outputs for discourse development.

Participants associated with this discourse acknowledged the complexity of the concept of societal well-being and the difficulty of capturing it adequately in a single measure (S5: +5*, S7: +5*). They further expressed that they felt GDP was not the best way to capture this complexity (S15: −2, S17: −5). However, participants from this discourse did not think that altering the way in which GDP is calculated would be the solution to this problem (S19: −5*, S16: −3). These sentiments were supported by interview quotes:

> There's always more than one number (Participant 10)

> I don't think [GDP] is enough to say [whether] someone has societal welfare or not. There are loads of other factors (Participant 1)

> GDP is primarily an economic indicator and it is useful for that… It would be [better] to have multiple indices that look at different things, rather than changing something that essentially was never intended to be a measure of societal welfare (Participant 1)

Those who associated with this factor were also distinguished by a belief that we need to better capture the contribution of non-traditional sectors of the economy, such as the gig economy, to societal well-being (S20: +4). This is again more of a technical issue than a value-based issue about what we should measure as part of societal well-being.

Finally, participants showed general indifference or indecision (particularly when compared to other factors) towards more moralistic issues such as: whether we should be concerned about sustainable well-being for future generations (S47: −2*); whether empowerment is a key part of well-being (S39: 0*); the relative importance of community and interpersonal relationships (S35: 0, S34: −1); and whether well-being can be expressed in monetary terms (S29: −1). They were, further, reluctant to show strong views on the role of government in promoting stable relationships and parenting (S36: 0*), and whether government should prioritise economic growth over other (perhaps less well defined) factors, such as sustainability and wellbeing (S22: 0).

4.2.4 Areas of Consensus Between Discourses

There was a broad base of consensus among all factors, with 24 consensus statements identified after bootstrapping (Table 8). This means that there were 24 statements for which the mean positioning of the statement was indistinguishable between all three factors.

The need to measure inequality and basic human rights in the UK was a stance that was shared across all factors (S25: −2, −4, −4; S26: +3, +4, +3; S27: −4, −4, −4, S38: −3, −3, −2). In particular, one participant felt that:

> We might be better than many other countries on some of these measures, but we are a very long way from perfect. And actually, if we assess these [things] we might not find we are quite as good as we like to think (Participant 11).

Table 8 Summary information for consensus statements

Characteristic	Description
Consensus statements	The PCA revealed 18 consensus statements before bootstrapping, and 24 after
Unstable statements	Statements 32 and 36 were no longer consensus after the bootstrapping procedure, but statements 6, 20, 25, 28, 29, 30, 34 and 41 were

All discourses also shared the stand point that economic growth is not the foundation of well-being (S21: −2, −3, −3).[6] Qualitative data from participants suggest that the primary reason for disagreeing with statement 21 was the importance of other factors in determining well-being too. In particular, they indicate an aversion to the *centrality* of economic growth and GDP in measuring societal well-being, rather than a disagreement with it having any role at all. This is demonstrated by the following quotes:

> [I don't believe that] economic growth is the essential foundation of everything, of all our wellbeing. I think there's lots of other things that are important as well. (Participant 11)

> There are things that don't necessarily correlate with GDP like people's mental health or people's relationships, so I just wouldn't call it a reliable measure at all (Participant 14)

In line with this, all discourses also agreed that GDP per capita is not a good measure of standard of living (S12: +3, +3, +3), in particular that it does not give a fair reflection for *most* people in the UK.

> Economic wealth is largely in the hands of a few individuals – so GDP doesn't tell you much about the quality of life for the citizens of that country (Participant 12)

They also felt that all aspects of well-being cannot be fairly expressed in monetary terms (S29: −3, −4, −1) and that focusing on enhancing GDP as a way to improve well-being might, therefore, lead to unintended, negative consequences (S30: +2, +3, +1). One participant highlighted some of the potential negative consequences of this 'over-focus' on monetary values and GDP, such as increasing inequality and environmental decline (Participant 19). This stance was exemplified by the following quote:

> I think there are elements or aspects of wellbeing where it's so difficult to put a monetary value on them that we don't, and because there's so much emphasis on the monetary value, those factors just get left out altogether (Participant 11)

5 Discussion and Conclusion

5.1 Recap of the Discourses

Using Q methodology, we have investigated the views that exist among civil servants about how we should measure societal well-being in the UK. The three discourses identified

[6] Although statement 21 had a large standard error for factor 3 (SE = 1.01), the upper error bound still placed the statement in the 'disagree' part of the spectrum. This indicates that although there is uncertainty in the *degree* of disagreement with the statement, all factors did disagree with it to some extent.

accounted for 72% of the study variance, with each representing a distinct perspective on measuring societal well-being. In brief, those participants who aligned with the socio-environmental discourse (#1) were concerned about the potential consequences of ignoring natural, social and human capital in decision making. Those associated with the self-determination discourse (#2) held the strong belief that access to opportunity and the ability to define one's own destiny were key determinants of well-being, with an emphasis on economic security as a way to facilitate individual autonomy. Lastly, those participants associated with the technocratic discourse (#3) were reluctant to express strong views on moralistic issues or on statements about the role of government; instead they tended to focus on the merits and disadvantages of specific ways of measuring societal well-being.

5.2 Implications for Measuring Well-Being in the UK

There were very few statements where discourses were in direct contradiction with one another, with most distinguishing statements differentiating between strong feelings towards a statement and less strong, or neutral, feelings. The three discourses therefore represent different focuses on what is considered by civil servants to be most central to well-being in the UK, rather than direct disagreements about whether or not certain elements contribute to well-being at all. In many ways, this makes the differences between the discourses easier to resolve and highlights the role of Q methodology in allowing differences of opinion to be highlighted in a nuanced and transparent way.

Three recommendations can be drawn for indicator development from this work: first, to increase the use of a capitals-based approach; second, to use both outcome and opportunity-based metrics; and third, to include more disaggregated measures of inequality. We discuss each briefly below.

First, some civil servants clearly favour a more extensive use of the capitals model of national wellbeing, particularly with respect to natural capital. All discourses agreed with this sentiment to some degree, with discourse 1 showing a particularly strong preference for a capitals-based approach. Indicators of capital currently appear in the MNW programme in a very limited way (e.g. only one measure of natural capital is used). More comprehensive capital accounts already exist for the UK in other places (Office for National Statistics 2017, 2019a, b), and integrating them more fully into a centralised indicator would offer decision-makers in the civil service a more complete picture of the 'stock' of wellbeing in the UK today.[7] New Zealand, for example, has integrated a capitals-based dashboard into their national 'Living Standards Framework', and are using it to help identify budget priorities and distinguish between department funding bids (The Treasury 2019). In fact, there are many capitals-based indices from which the MNW programme could draw (e.g. Index of Sustainable Economic Welfare (Cobb and Daly 1989), Inclusive Wealth Index (Thiry and Roman 2014), etc.).

Second, the most contentious statement in our Q-study was that "we should be assessing quality of life in terms of the opportunities people have to achieve well-being… rather than whether or not they actually achieve it…" (S13). Here, the point of contention between discourses centred around whether it would be sufficient to assess opportunity, or whether this would be a meaningless measure in the face of a complex

[7] The ONS undertook a consultation in 2019 focused on improving its human capital measures, making this an opportune moment to include such new measures into the MNW programme (Office for National Statistics 2019c).

society, where other factors could hinder someone's ability to fulfil that opportunity. In this instance, the MNW dashboard of indicators captures almost exclusively measures of outcome, with no significant inclusion of measures of opportunity (Office for National Statistics 2018). It is worth noting that attempting to aggregate outcome measures with measures of input or opportunity can lead to 'double counting' wellbeing, introducing sources of uncertainty into an index (Fu et al. 2011). However, given the dashboard structure of the MNW programme (i.e. measures are not aggregated), there is no theoretical reason not to add a sub-section to the dashboard that reflects citizens' opportunities to flourish.

Third, there was a strong emphasis in all three discourses on the continuing need to measure inequality and human rights in the UK. For example, most indicators in the MNW dashboard are broken down by age and gender. However, the dashboard currently only reports headline figures for these subgroups, and not spread; it gives no indication of the statistical significance of any differences between subgroups; and there is no break down by other important subgroups, such as ethnicity or socio-economic status (Office for National Statistics 2018). Given the apparent importance of disaggregated information for civil servants—a finding that is supported elsewhere in the literature (Sébastien and Bauler 2013)—this may be hindering the use and usefulness of such indicators. Other indicator frameworks do address this issue to some degree (e.g. Global Gender Gap Index, and Inequality-Adjusted Human Development Index, Index of Sustainable Economic Welfare, Genuine Progress Indicator) (Yang 2014). However, their limited focus on a single axis of equality, such as income or gender, leaves room for further development.

In addition to these three concrete recommendations, our study also appeared to reveal a view about economic growth that was common to all three discourses. Specifically, that economic growth is not the foundation of societal well-being, and that monetary expressions of that well-being, such as GDP, do not adequately reflect the standard of living of most people in the UK. These results indicate support for the existence of a view among civil servants that economic growth is not the *central and sole* driver of our well-being. However, discourse three in particular has a large number of statements with large standard errors (i.e. there was a low level of agreement between participants within the discourse) and, as a whole, is highly focused on technical issues. The combination of these two facts causes us to question the simple narrative of a shared sentiment about economic growth, and gives rise to two possible interpretations. The first possible interpretation is that the premise of our study—the need to measure societal well-being, beyond GDP—does not fit the worldview of the participants associated with discourse three. This explanation draws from the "overcritical model" of the use of science in policy making (Turnhout et al. 2007, pp. 223), where actors will try to "deconstruct, discredit and reject scientific knowledge that does not fit with already existing opinions, fixed interests or established consensus" (Turnhout et al. 2007, pp. 223). This could hint at a partial explanation for why uptake of indicators of societal wellbeing is still low within government. If key actors within government are highly critical of the producers of, or the conceptual framework underpinning, the societal-wellbeing indicators, then no matter which specific indicators are chosen it is unlikely that the indicators will have any influence on policy, even if they become embedded in the policy process. The second possible interpretation is that these actors may simply have a clear understanding of the complexity of societal well-being and the limitations of GDP as a measure of it, reflected in their strong opinions about adjustments to GDP and their uncertainty about whether and how to include moralistic aspects of well-being. Although the unstable Q-sorts and statements caution us to tread lightly with any concrete conclusions from discourse three, this result certainly points towards an interesting area for future research.

These insights offer some practical examples of how understanding the views of end-users can help with indicator development, and may support wider use, echoing what civil servants and practitioners have expressed in other studies: that "policy-makers need to become far more engaged in the [indicators] discourse if these tools are to succeed" (Bell and Morse 2011, p. 298).

5.3 Limitations and Future Research

This research was designed as an exploratory study, offering a practical example of the way that better understanding the views of indicator end-users could support improved indicator development. We hypothesise that this might then support more widespread use by civil servants. There is now need for research which takes a less exploratory and more systematic approach, (1) to reveal the actual extent of use of indicators of societal wellbeing within government, and (2) to capture the prevalence of certain views among civil servants about those indicators. Further research that seeks to better understand how a range of factors—including worldviews, organisational culture, data literacy, seniority, supporting legislation, public opinion, and political agendas—affect the use and influence of indicators of societal wellbeing by civil servants in practice is also needed. Recent developments in New Zealand might offer a rich potential case study, as the government released their first "wellbeing budget" in 2019. This provides arguably the most advanced example of a national government integrating indicators of national wellbeing into policy decision-making. Of particular interest, the national indicator set—the 'Living Standards Framework'—was developed by Treasury itself (The Treasury 2018) and is now, in theory, being used to direct policy proposals and to inform budgetary decisions (The Treasury 2019). Further, the Scottish 'National Performance Framework' and the 'National Indicators for Wales' are both supported by legislation mandating that ministers set targets and monitor progress against a set of national wellbeing indicators (Community Empowerment (Scotland) Act 2015; Well-being of Future Generations (Wales) Act 2015). This offers the opportunity, for example, to derive insights about the effectiveness of specific tools to create an environment that encourages indicator use.

A central limitation of this study is that it is not possible to verify how successful we have been in identifying the full diversity of opinions across the civil service about measuring societal wellbeing. In particular, we might have reason to examine the validity of our results due to the very high level of consensus across the statements and factors. As stated in Sect. 4.2, there were 24 consensus statements identified through the bootstrapping analysis, and there was also a high level of correlation between factors 1 and 2 (see Table 11 in Appendix 6). This might suggest that there is a broad base of agreement which underpins all three of the discourses, and particularly discourses 1 and 2. However, it might also speak to the precision of our Q-study, indicating that the selected Q statements were too general to detect nuanced differences in viewpoint. Alternatively, it might even be a result of sample bias, introduced by the snowball sampling technique, where the inclusion of individuals with inter-relationships may "over-emphasise [the] cohesiveness in [the] social network" (Atkinson and Flint 2001, pp. 2). As we argued in Sect. 2, this does not invalidate the opinions expressed by participants, particularly given the exploratory nature of the study. However, providing a clearer answer as to why there was such a high level of consensus would be a useful next step for further research, perhaps by extending this research to a more diverse set of civil servants and adding more (and more specific) statements to improve precision (Brown et al. 2015), or augmenting it with in-depth interviews (as per Steelman and Maguire 1999; Valenta and Wigger 1997; Brown 1993). Alternatively, as suggested above, a more systematic approach to identifying civil servants' views (e.g. through survey-based methods) might now be appropriate. Nonetheless, the views identified here can provide first valuable insights for indicator development.

5.4 Conclusion

Through this study we have explored the views of civil servants in the UK towards measuring societal well-being. Three distinct discourses emerged from our analysis: one concerned about the consequences of ignoring natural, social and human capital in decision making; one that emphasised opportunity and autonomy as key determinants of well-being; and one that focused on the technical aspects of measuring societal well-being. These discourses hold insights that have particular relevance for the further development of the Measuring National Wellbeing programme, as the primary indicator framework used the UK. The data gathering, valuation and aggregation methodologies are generally already advanced enough to implement these kinds of changes to an indicator framework like the MNW. This again draws attention to the need to bring the focus of the indicator literature away from issues of technical development, and towards questions about how end-users' worldviews, organisational culture, data literacy, supporting legislation, and political agendas affect indicator use and influence in policy making. This is not to negate the importance of improvements to data availability and valuation methodologies, but rather to acknowledge that they are one element in a complex indicator ecosystem, of which many parts have so far been understudied. We therefore hope that this paper has effectively highlighted the potential benefits that considering the views of end-users might bring to indicator development initiatives.

Acknowledgements Many thanks go to the twenty UK civil servants who gave up their time to allow this work to happen, and to the pilot study participants who helped to shape the final study design. This research was part-funded by the UK's Economic and Social Research Council (ESRC) South East Network for Social Sciences.

Appendix 1: Review of Participant Numbers Used in Q Studies 2008–2018

A literature search was conducted on 13/11/2018 in Web of Science using the search terms: *TITLE:* ("Q methodology"). This search returned 268 results, of which 251 were journal articles, and 205 were empirical studies. We extracted the number of participants used in each study; the results are detailed in Fig. 3, below.

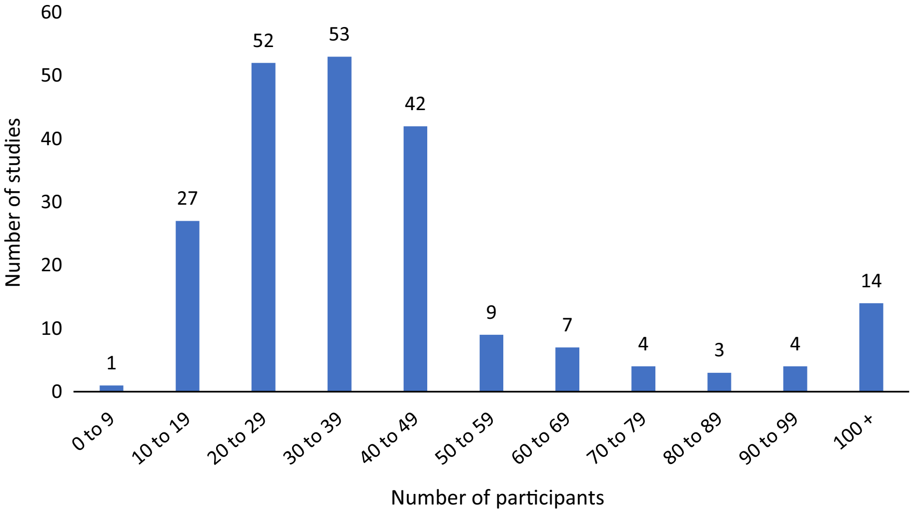

Fig. 3 Graph showing participant numbers for 205 empirical Q studies, 2008–2018

Appendix 2: Sources for Concourse Development

Table 9 details each of the sources used in the development of the concourse, and how many relevant items were identified from each source. In total, 401 relevant statements were identified across all sources.

Table 9 Sources used in development of the concourse, and number of individual items used, per source

Source	Number of relevant items identified
Web of Science	42 peer reviewed papers
Twitter	18 tweets
The Economist	5 articles
The Office for National Statistics	2 publications
TheyWorkForYou	2 debates
UK Government	1 speech
Vimeo	1 video

Appendix 3: Full R Code

Below we detail the R code, including both standard and bootstrapped PCA, developed in R version 3.3.2 (R Core Team 2016).

```
# Open data and check correctly loaded
library(qmethod)
setwd("file path")
mydata<-read.csv("Raw data_v1.csv")
dim(mydata)
mydata
View(mydata)

# Explore the correlations between Q-sorts
cor(mydata)

# Extract factors and examine Q-sort loadings and factor characteristics
results <-qmethod(mydata,nfactors=3)
round(results$loa,digits=2)
results$flag
loa.and.flags(results)
results$f_char$characteristics

# Plot screeplot of eigenvalues
screeplot(prcomp(mydata),main="Screeplot of unrotated factors",type="l")

# View results
summary(results)
results
results$qdc
plot(results)
scores <-cbind(round(results$zsc,digits=2),results$zsc_n)
nfactors <- ncol(results$zsc)
col.order <- as.vector(rbind(1:nfactors, (1:nfactors)+nfactors))
scores <- scores[col.order]
scores
scores[order(scores$zsc_f1, decreasing = T), ]
scores[order(scores$zsc_f2, decreasing = T), ]
scores[order(scores$zsc_f3, decreasing = T), ]
```

```
# Explore distinguishing and consensus statements
results$qdc
results$qdc[which(results$qdc$dist.and.cons=="Consensus"), ]
results$qdc[which(results$qdc$dist.and.cons == "Distinguishes all"), ]
results$qdc[which(results$qdc$dist.and.cons == "Distinguishes f1 only"), ]
results$qdc[which(results$qdc$dist.and.cons == "Distinguishes f2 only"), ]
results$qdc[which(results$qdc$dist.and.cons == "Distinguishes f3 only"), ]

# Save and export results
save(results, file = "myresults.Rdata")
load("myresults.Rdata")

# Table of z-scores:
write.csv(results$zsc, file = "zscores.csv")
# Table of factor scores:
write.csv(results$zsc_n, file = "factorscores.csv")
# Table of Q-sort factor loadings:
write.csv(results$loa, file = "loadings.csv")

# Report all results as a text file
export.qm(results, file = "myreport.txt", style = "R")
export.qm(results, file = "myreport-pqm.txt", style = "PQMethod")

# Run bootstrap analysis and report results to text file
options(max.print=999999)
bootresults <-qmboots(mydata,nfactors=3,nsteps=1000,load = "auto",rotation
= "varimax")
bts <- qmb.summary(bootresults)
bts
export.qm(bts, file = "mybtsreport.txt", style = "R")
export.qm(bts, file = "mybtsreport-pqm.txt")
export.qm(bootresults, file = "mybootresultsreport.txt", style = "R")
export.qm(bootresults, file = "mybootresultsreport-pqm.txt")
```

Appendix 4: Bootstrapping Methodology

To understand why it is necessary to calculate *new* measures of standard error we can look at how the usual Q methodology calculates the standard errors for statement z-scores: a simplified standard error calculation is used, which is particularly sensitive to the number of significantly loading Q-sorts (Eq. 1).

$$SE_f = s_f\sqrt{1 - \frac{0.8p}{1 + (p - 1)0.8}} \tag{1}$$

where *SE* is the standard error for factor *f; s* is the standard deviation of the distribution; and *p* is the number of Q-sorts that load significantly onto that factor (Zabala and Pascual 2016). This means that for small sample sizes like those commonly seen in Q studies, Eq. 1 is highly sensitive to changes in the number of participants and, therefore, is not a particularly reliable measure (Zabala and Pascual 2016). Since these standard errors are used to identify distinguishing and consensus statements—which play an important part in the development of the final discourses—it is crucial to ensure that they are reliable.

Bootstrapping analysis can help us to calculate more accurate measures of reliability through repeated re-sampling and replacement of Q-sorts (Zabala and Pascual 2016). In practice, this means that the PCA is run multiple times, but on each new run a random sample of *n* Q-sorts (or completed grids) is selected from the set of *n* Q-sorts collected in the study. This sample may include repeats of some Q-sorts and may be missing others (Zabala and Pascual 2016).

There are two key measures of uncertainty that emerge from the bootstrapping analysis, which can inform the interpretation of factors and subsequent development of discourses. First, how strongly and stably a Q-sort defines a particular factor. A Q-sort is considered unstable if its loading score has a standard error greater than 0.2, or if it does not load *consistently* onto one particular factor (i.e. it is flagged as significant for a factor in between 20% and 80% of bootstrap repetitions) (Zabala and Pascual 2016). Unstable Q-sorts may change the interpretation of a factor because if certain Q-sorts, which previously contributed to the factor discourse, are no longer considered exemplary of that factor, they will no longer be relevant to the discourse development.

Second, the stability and salience of statements within the factor is another key measure of uncertainty (Zabala and Pascual 2016). A statement is unstable for a factor if it has a large bootstrap estimate of bias, or a large z-score standard error. It is important to note that there may not be any correlation between the standard error of a statement and a change in its factor score, or a change in its status as a distinguishing statement. This is because a factor's relative position also relies on the standard errors and z-scores of adjacent statements (Zabala and Pascual 2016). Statement instability is important if the statement changes position within the idealised Q-sort (e.g. from a factor score of 3, to a score of 5). Alternatively, unstable statements might change status from being distinguishing to being consensus, or vice versa. Both issues may alter the interpretation of the factor and therefore the resulting discourse development.

Table 10 Matrix of statements, by topic area, gov. = government

Statements	Environmental indicators	Social indicators	Economic indicators	Composite indicators	Subjective indicators	Role of gov.	Type of statement
1. To track societal welfare, we should be measuring growth in total wealth, crucially including natural (e.g. stock of forests), social (e.g. interpersonal relationships) and human (e.g. literacy skills) capital	x	x					Object of measurement
2. It is crucial to acknowledge the importance of natural capital (e.g. green space, proportion of forest cover, water quality) for societal well-being in any composite measure	x						Object of measurement
3. There are certain goods and services that the environment provides, which are important for societal well-being, that cannot be replaced by man-made goods and services	x						Relevance to wellbeing
4. We should include the value of natural capital in a composite indicator so that decision makers can better include the environment as they allocate resources to promote the growth of the economy	x						Object of measurement
5. The scope of societal welfare is too broad and subtle for a single measure to evaluate satisfactorily				x			Quality of measure
6. Using a composite indicator to create a bottom-line is more useful in gathering the interest of media and policy-makers than the use of sets of indicators				x			Quality of measure
7. Societal welfare cannot be considered simply the sum of individual welfare, as this ignores certain important synergies and thresholds		x					Quality of measure
8. The UK government should focus primarily on devising policies that minimise unemployment, with concerns about maximising GDP coming later			x				Role of government
9. The quality and security of available jobs is a key determinant of societal well-being			x				Relevance to wellbeing

Table 10 (continued)

Statements	Environ-mental indicators	Social indicators	Economic indicators	Compos-ite indica-tors	Subjective indicators	Role of gov.	Type of statement
10. Ecosystem interactions with human well-being (e.g. recreation time in nature, provision of food, cultural significance) have no place in a quantitative measure of societal well-being	x						Object of measurement
11. When measuring environmental well-being, we should focus on the measurement of final ecosystem goods and services (e.g. quantity of fish provided), rather than the ecosystem functions that underpin them (e.g. quality of stream habitat)	x						Object of measurement
12. GDP per capita alone does not capture how successfully most individuals can access the resources required for a decent standard of living, for this we need some measure of economic security			x				Quality of measure
13. We should be assessing quality of life in terms of the opportunities people have to achieve well-being (e.g. access to education, access to healthcare), rather than whether or not they actually achieve it (e.g. literacy rates, life expectancy)		x					Quality of measure
14. Indicators should be focused on well-being outcomes (e.g. health status), as opposed to inputs (e.g. healthcare expenditure)		x					Quality of measure
15. An absolute increase in GDP can be considered broadly synonymous with enhanced societal well-being, even in wealthy nations			x				Relevance to wellbeing
16. We should focus on making technical changes to the way GDP is calculated, so that it more accurately reflects welfare, rather than developing new indicators			x				Quality of measure
17. GDP is the most appropriate and reliable measure we have to capture societal well-being			x				Quality of measure
18. GDP should be adjusted to include income disparities			x				Object of measurement

Table 10 (continued)

Statements	Environmental indicators	Social indicators	Economic indicators	Composite indicators	Subjective indicators	Role of gov.	Type of statement
19. Adjusting GDP to account for government borrowing would provide a better idea of how financially sustainable our social progress is			x				Quality of measure
20. We need to better capture the contribution of non-traditional economic activity (e.g. open source information sharing, technological innovation, the gig economy) to societal well-being			x				Object of measurement
21. Economic growth is the essential foundation of all our well-being aspirations			x				Relevance to wellbeing
22. Government policy should prioritise economic growth over ill-defined concepts of sustainability and societal well-being			x				Role of government
23. Quality of life is strongly associated with material wealth, and the well-being you derive from having material possessions			x				Relevance to wellbeing
24. People's wealth in a given period is a better measure of their well-being than yearly income flows			x				Quality of measure
25. Inequality is currently too poorly defined and too difficult to measure to be included in national accounts		x					Quality of measure
26. If well-being indices are meant to capture non-economic well-being and determine the real inclusive and equitable growth of society, they should also move beyond the measurement of purely economic inequality		x					Quality of measure
27. Inequality has less of an impact on well-being in wealthier countries, therefore we do not need to include it in a UK measure of societal well-being		x					Object of measurement
28. Comparisons of living standards over time need to take into account the amount of leisure that people enjoy		x					Object of measurement

Table 10 (continued)

Statements	Environmental indicators	Social indicators	Economic indicators	Composite indicators	Subjective indicators	Role of gov.	Type of statement
29. All aspects of well-being can be fairly expressed in monetary terms			x				Quality of measure
30. Focusing governmental interventions on enhancing GDP, as a way to improve welfare, is problematic because it can have unintended, negative consequences			x				Role of government
31. In any measure of societal welfare, we must account for the social and environmental damage caused by economic activity (e.g. deforestation from crop production, pollution from energy generation, lung cancer from tobacco sales)	x	x					Object of measurement
32. It is not possible to accurately value non-market goods (e.g. volunteering, unpaid care, national forests), and to do so in a measure of societal welfare would be misleading			x				Quality of measure
33. Non-market goods and services (e.g. volunteer work and unpaid care) form an invisible pillar of our economy, and should therefore be included in a measure of societal well-being			x				Object of measurement
34. The role of strong interpersonal relationships and social networks (e.g. local communities) in contributing to well-being is over-played		x					Relevance to wellbeing
35. People derive more well-being from their existence within a community than from personal consumption		x					Relevance to wellbeing
36. The government has a critical role to play in promoting stable relationships and good parenting						x	Role of government
37. Social contacts and trust play an important role in expanding markets and increasing income			x				Relevance to wellbeing
38. Although human rights (e.g. freedom of speech, the right to vote) are fundamental to well-being, they do not need to be measured in the UK		x					Object of measurement

Table 10 (continued)

Statements	Environmental indicators	Social indicators	Economic indicators	Composite indicators	Subjective indicators	Role of gov.	Type of statement
39. People who feel empowered and in control of their own destiny feel more fulfilled, so we should be measuring freedom and ability to choose across multiple domains (e.g. healthcare, education, housing, faith)		x					Object of measurement
40. We should measure the government's success in abating deficiencies (e.g. unemployment), rather than fulfilling desires (e.g. achieving a degree)						x	Quality of measure
41. You can't legislate for fulfilment or satisfaction, and therefore we should not include it in a government measure of societal well-being						x	Role of government
42. Government should seek to describe societal well-being using subjective measures rather than prescribe it by measuring things we presume to be good for well-being (e.g. education level or health status)					x		Quality of measure
43. Subjective assessment of well-being can be affected by many factors (e.g. cultural background, socio-economic context and even participant mood), therefore it is not suitable to be used within a national measure of societal well-being					x		Quality of measure
44. The state of housing in the UK should primarily be captured through objective measures (e.g. proportion of houses in each council tax band), not through subjective measures such as the Gallup question, which asks the public whether they feel there is enough, good quality housing available					x		Quality of measure
45. Using subjective well-being measures as an indicator of societal welfare could lead society to over-value the ability to be happy or contented with one's "lot in life", no matter how limiting or inequitable that lot is					x		Quality of measure

Table 10 (continued)

Statements	Environmental indicators	Social indicators	Economic indicators	Composite indicators	Subjective indicators	Role of gov.	Type of statement
46. Any measure of sustainability needs to keep track of environmental processes that may be irreversible or non-linear over certain time-lines	x						Object of measurement
47. We do not need to be concerned about sustainable well-being for future generations whilst we still have the well-being of the current generation to worry about	x						Relevance to wellbeing
48. We do not have an agreed definition for sustainability, so we shouldn't try to include it in our national accounts yet	x						Quality of measure
Total number of statements	10	13	18	2	4	3	

Table 11 Extraction criteria results for each factor

Factor	Number of loading sorts	Eigenvalues	Explanatory variance (%)	Humphrey's rule	Correlation between factors
f1	10	6.53	32.66	0.63	0.77 (f1:f2), 0.68 (f1:f3)
f2	5	4.52	22.60	0.61	0.77 (f2:f1), 0.57 (f2:f3)
f3	3	3.37	16.83	0.52	0.68 (f3:f1), 0.57 (f3:f2)

For this study the significance threshold for sorts to be counted as 'loading' was 0.37. Where a Q-sort met the significance threshold for more than one factor, the squared loading value had to be greater than the sum of the squared loadings for the other factors for the sort to be considered exemplary of that factor. Humphrey's threshold was 0.29

Appendix 5: Statement Matrix

In Table 10 we include a breakdown of Q statements, by topic area. These areas reflect the most significant discussion areas that emerged in the process of concourse development. The statements are split approximately evenly between the three major categories (environmental, social, and economic indicators), with a handful of additional statements covering other notable topics, such as how informative composite indicators are, the relative merits of subjective and objective indicators, and the role of government in measuring and affecting societal well-being.

Appendix 6: Full Standard PCA Outputs

A three-factor solution was found to meet all relevant criteria for extraction from the PCA for further analysis. Table 11 contains details of each factor's results against each criterion.

Although the correlation between factors one and two was particularly high, inspection of the qualitative data justified keeping a three-factor solution, rather than a two-factor solution, which had an even higher correlation between the two remaining factors.

Appendix 7: Full Bootstrapped PCA Outputs

Here we include a series of tables, summarising the bootstrap outputs. Table 12 shows the standard and bootstrapped factor loadings for each Q-sort, against each factor. All standard errors for the bootstrapped results sit within a relatively narrow range, from 0.2 to 0.4. However, they are larger than the standard 0.2 threshold for reliable results. This may be a result of the small sample size.

In Table 13 we have calculated the z-score estimate of bias for each statement against each factor. The z-score estimate of bias is the difference between the standard PCA z-scores and the bootstrapped z-scores (Zabala and Pascual 2016). This can give some indication of the stability of a statement, although it is less informative than the z-score standard error.

Table 12 Standard (std.) and bootstrapped (bts.) Q-sort factor loadings, and standard errors (SE), and bootstrapped flagging frequencies

Q-Sort	Department	Std. factor loading			Bts. factor loading (&SE)						Flagging frequency		
		f1	f2	f3	f1		f2		f3		f1	f2	f3
Factor 1 (f1)													
P3	Home Office	**0.66**	0.34	0.44	**0.57**	*(0.26)*	0.28	*(0.32)*	0.32	*(0.28)*	0.62$^{\Delta}$	0.09	0.09
P4	DfID	**0.60**	0.21	0.34	**0.54**	*(0.26)*	0.18	*(0.25)*	0.24	*(0.34)*	0.62$^{\Delta}$	0.09	0.18
P6	DfID	**0.80**	0.37	0.14	**0.69**	*(0.24)*	0.29	*(0.30)*	0.11	*(0.27)*	0.81*	0.08	0.04
P9	DEFRA	**0.66**	0.49	0.22	**0.57**	*(0.29)*	0.38	*(0.31)*	0.14	*(0.27)*	0.69$^{\Delta}$	0.13	0.04
P11	DEFRA	**0.69**	0.51	0.20	**0.60**	*(0.27)*	0.42	*(0.29)*	0.13	*(0.25)*	0.68$^{\Delta}$	0.16	0.02
P12	Local gov.	**0.72**	0.35	0.42	**0.63**	*(0.28)*	0.29	*(0.30)*	0.32	*(0.26)*	0.75*	0.09	0.04
P13	DfID	**0.61**	0.31	0.47	**0.53**	*(0.25)*	0.26	*(0.30)*	0.37	*(0.28)*	0.52$^{\Delta}$	0.10	0.16
P14	DfID	**0.70**	0.41	0.33	**0.60**	*(0.26)*	0.33	*(0.32)*	0.23	*(0.29)*	0.66$^{\Delta}$	0.13	0.02
P15	DfID	**0.75**	0.27	0.41	**0.65**	*(0.27)*	0.23	*(0.29)*	0.32	*(0.27)*	0.77*	0.08	0.05
P19	DEFRA	**0.79**	0.19	0.22	**0.68**	*(0.25)*	0.15	*(0.28)*	0.19	*(0.24)*	0.82*	0.10	0.04
Factor 2 (f2)													
P2	Home Office	0.06	**0.72**	0.46	0.05	*(0.33)*	**0.62**	*(0.32)*	0.23	*(0.24)*	0.11	0.83*	0.04
P5	Home Office	0.36	**0.81**	0.21	0.29	*(0.26)*	**0.69**	*(0.31)*	0.11	*(0.27)*	0.01	0.90*	0.06
P7	HM Treasury	0.47	**0.74**	0.04	0.36	*(0.25)*	**0.61**	*(0.33)*	−0.01	*(0.31)*	0.10	0.77*	0.11
P8	Local gov.	0.33	**0.70**	0.39	0.29	*(0.21)*	**0.64**	*(0.28)*	0.28	*(0.29)*	0.04	0.80*	0.05
P18	BEIS	0.39	**0.75**	0.16	0.31	*(0.22)*	**0.66**	*(0.29)*	0.11	*(0.26)*	0.05	0.85*	0.05
Factor 3 (f3)													
P1	HM Treasury	0.29	0.22	**0.63**	0.25	*(0.20)*	0.19	*(0.21)*	**0.55**	*(0.37)*	0.13	0.07	0.65$^{\Delta}$
P10	DEFRA	0.44	0.14	**0.50**	0.33	*(0.25)*	0.13	*(0.26)*	**0.45**	*(0.37)*	0.35$^{\Delta}$	0.11	0.48$^{\Delta}$
P17	HM Treasury	0.22	0.17	**0.83**	0.23	*(0.21)*	0.14	*(0.24)*	**0.68**	*(0.31)*	0.06	0.07	0.77*
Confounded sorts													
P16	Scottish Gov.	0.58	0.43	0.53	0.49	*(0.30)*	0.35	*(0.35)*	0.39	*(0.30)*	0.33$^{\Delta}$	0.14	0.14
P20	DEFRA	0.59	0.43	0.44	**0.49**	*(0.28)*	0.36	*(0.32)*	0.32	*(0.29)*	0.37$^{\Delta}$	0.14	0.10

Bold indicates participants that significantly loaded onto the factor. * denotes a flagging frequency ≥ 0.75. $^{\Delta}$denotes a flagging frequency > 0.2 and < 0.75

Table 13 Z-score bias estimates and standard (std.) and bootstrapped (bts.) factor scores for each statement (#) against each factor

#	z-score bias estimate			Factor scores					
	f1	f2	f3	f1		f2		f3	
				Std.	Bts.	Std.	Bts.	Std.	Bts.
1	0.21	0.14	−0.15	4**	5*	1		1	2
2	0.24	0.15	−0.04	5**	4*	1		1	
3	0.24	0.16	0.42	3$^{\Delta}$	$^{\Delta}$	3$^{\Delta}$	$^{\Delta}$	3$^{\Delta}$	$^{\Delta}$
4	0.22	0.09	0.07	4**	*	1		1	2
5	0.27	0.09	0.52	2	1	1		5**	*
6	0.16	−0.10	0.37	0	$^{\Delta}$	0	$^{\Delta}$	2	1$^{\Delta}$
7	0.13	−0.04	0.67	1**	*	−1**	*	5**	*
8	−0.08	0.00	0.11	−1$^{\Delta}$	$^{\Delta}$	−1$^{\Delta}$	$^{\Delta}$	0$^{\Delta}$	$^{\Delta}$
9	0.07	0.39	0.50	1**		4		4	
10	−0.14	−0.26	−0.40	−3$^{\Delta}$	−4$^{\Delta}$	−2$^{\Delta}$	$^{\Delta}$	−3$^{\Delta}$	$^{\Delta}$
11	−0.18	−0.17	−0.30	−2$^{\Delta}$	$^{\Delta}$	−2$^{\Delta}$	$^{\Delta}$	−3$^{\Delta}$	−2$^{\Delta}$
12	0.21	0.15	0.28	3$^{\Delta}$	$^{\Delta}$	3$^{\Delta}$	$^{\Delta}$	3$^{\Delta}$	$^{\Delta}$
13	0.06	0.32	−0.19	−1		4**	*	−3	−4
14	0.20	0.33	0.65	0**	*	5		4	
15	−0.26	0.09	−0.38	−2		0*		−2	
16	0.05	−0.04	−0.24	0		0		−2	−3
17	−0.24	−0.01	−0.57	−4		0**	*	−5	
18	0.08	0.08	−0.05	1$^{\Delta}$	$^{\Delta}$	0$^{\Delta}$	$^{\Delta}$	1$^{\Delta}$	$^{\Delta}$
19	0.25	0.05	−0.64	1	0	0		−5**	*
20	0.13	0.18	0.69	2	$^{\Delta}$	2	$^{\Delta}$	4*	$^{\Delta}$
21	−0.41	−0.23	−0.58	−3$^{\Delta}$	−2$^{\Delta}$	−3$^{\Delta}$	$^{\Delta}$	−4$^{\Delta}$	−3$^{\Delta}$
22	−0.13	−0.23	0.14	−5**	*	−2		0	
23	0.02	−0.05	−0.07	−1$^{\Delta}$	$^{\Delta}$	−1$^{\Delta}$	$^{\Delta}$	−1$^{\Delta}$	$^{\Delta}$
24	−0.06	−0.09	−0.20	0$^{\Delta}$	$^{\Delta}$	−1$^{\Delta}$	$^{\Delta}$	0$^{\Delta}$	$^{\Delta}$
25	−0.11	−0.34	−0.49	−1	−2$^{\Delta}$	−4	$^{\Delta}$	−4	$^{\Delta}$
26	0.26	0.24	0.44	3$^{\Delta}$	$^{\Delta}$	4$^{\Delta}$	$^{\Delta}$	3$^{\Delta}$	$^{\Delta}$
27	−0.28	−0.37	−0.57	−4$^{\Delta}$	$^{\Delta}$	−5$^{\Delta}$	−4$^{\Delta}$	−4$^{\Delta}$	$^{\Delta}$
28	−0.03	0.19	−0.37	0	1$^{\Delta}$	2	$^{\Delta}$	−1*	0$^{\Delta}$
29	−0.36	−0.24	0.16	−4	−3$^{\Delta}$	−4	$^{\Delta}$	0**	−1$^{\Delta}$
30	0.18	0.18	0.02	2	$^{\Delta}$	3	$^{\Delta}$	1	$^{\Delta}$
31	0.22	0.27	0.32	5*	*	3		2	
32	−0.10	−0.19	−0.11	−2$^{\Delta}$		−1$^{\Delta}$	*	−1$^{\Delta}$	
33	0.21	0.26	0.58	4		2		3	
34	−0.17	−0.35	−0.12	−2	−3$^{\Delta}$	−3	$^{\Delta}$	−1*	$^{\Delta}$
35	0.20	0.23	0.03	2		2		0	
36	−0.01	0.19	0.10	1$^{\Delta}$	2	1$^{\Delta}$		0$^{\Delta}$	*
37	0.08	0.06	0.17	1$^{\Delta}$	$^{\Delta}$	1$^{\Delta}$	$^{\Delta}$	2$^{\Delta}$	$^{\Delta}$
38	−0.26	−0.26	−0.31	−3$^{\Delta}$	$^{\Delta}$	−3$^{\Delta}$	$^{\Delta}$	−2$^{\Delta}$	$^{\Delta}$
39	0.28	0.45	0.12	3**	2	5**		0**	
40	−0.12	−0.15	−0.17	−1$^{\Delta}$	$^{\Delta}$	−2$^{\Delta}$	$^{\Delta}$	−1$^{\Delta}$	$^{\Delta}$
41	−0.34	−0.30	−0.30	−2	−1$^{\Delta}$	−4	−5$^{\Delta}$	−3	$^{\Delta}$
42	0.07	−0.09	0.21	0$^{\Delta}$	$^{\Delta}$	0$^{\Delta}$	$^{\Delta}$	1$^{\Delta}$	$^{\Delta}$

Table 13 (continued)

#	z-score bias estimate			Factor scores					
	f1	f2	f3	f1		f2		f3	
				Std.	Bts.	Std.	Bts.	Std.	Bts.
43	−0.22	−0.28	−0.29	−1		−3		−1	
44	0.06	0.01	−0.39	0		0		−2**	
45	−0.03	−0.16	0.29	0		−1*	*	2	1
46	0.12	0.23	0.32	2^{Δ}	3^{Δ}	2^{Δ}	$^{\Delta}$	2^{Δ}	$^{\Delta}$
47	−0.40	−0.46	−0.30	−5		−5		−2**	*
48	−0.31	−0.18	0.07	−3		−2		0**	*

Distinguishing statements are denoted with asterisks; * is a significance of $p < 0.05$, and ** is a significance of $p < 0.01$; consensus statement

Figure 4 shows the z-score estimate of bias for each statement, against each factor. This graph enables us to easily identify distinguishing and consensus statements, from the positioning of the standard error bars.

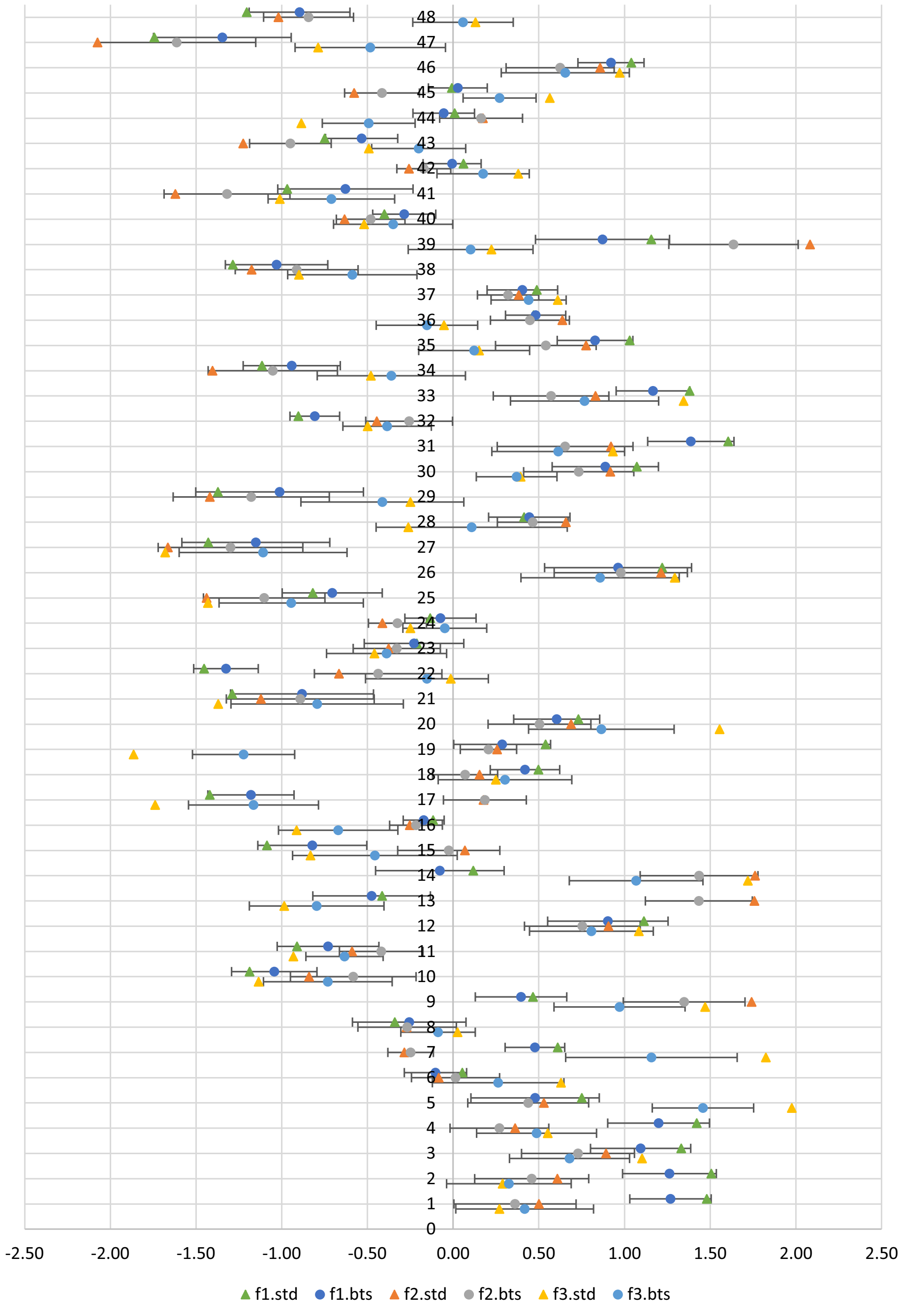

Fig. 4 Bootstrapped and standard z-scores with bootstrapped standard errors, for each Q statement. Triangles = standard z-scores. Circles = bootstrapped z-scores with standard errors

References

Allin, P., & Hand, D. J. (2017). New statistics for old? Measuring the wellbeing of the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 180*(1), 3–43.

Atkinson, R., & Flint, J. (2001). Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social Research Update, 33*(1), 1–4.

Baker, R., Thompson, C., & Mannion, R. (2006). Q methodology in health economics. *Journal of Health Services Research & Policy, 11*(1), 38–45.

Barrington-Leigh, C., & Escande, A. (2018). Measuring progress and well-being: A comparative review of indicators. *Social Indicators Research, 135*(3), 893–925.

Barry, J., & Proops, J. (1999). Seeking sustainability discourses with Q methodology. *Ecological Economics, 28*(3), 337–345.

Bauler, T. (2012). An analytical framework to discuss the usability of (environmental) indicators for policy. *Ecological Indicators, 17,* 38–45.

Bell, S., & Morse, S. (2011). An analysis of the factors influencing the use of indicators in the European Union. *Local Environment, 16*(3), 281–302.

Birkland, T. A. (2015). *An introduction to the policy process: Theories, concepts, and models of public policy making* (3rd ed.). New York, NY: Routledge.

Brown, S. R. (1980). *Political subjectivity: Applications of Q methodology in political science*. New Haven, CT: Yale University Press.

Brown, S. R. (1993). A primer on Q methodology. *Operant Subjectivity, 16*(3/4), 91–138.

Brown, S. R., Danielson, S., & van Exel, J. (2015). Overly ambitious critics and the Medici effect: A reply to Kampen and Tamás. *Quality & Quantity, 49*(2), 523–537.

Brown, S. R., Durning, D. W., & Selden, S. (1999). Q methodology. *Public Administration and Public Policy, 71,* 599–638.

Cameron, D. (2010). PM speech on wellbeing. GOV.UK. Retrieved June 15, 2018, from https://www.gov.uk/government/speeches/pm-speech-on-wellbeing.

Cobb, C. W., & Daly, H. (1989). The index for sustainable economic welfare. In H. E. Daly & J. B. Cobb (Eds.), *For the common good* (pp. 401–455). Boston: Beacon Press.

Collingridge, D., & Reeve, C. (1986). Science and Policy-Why the Marriage Is So Unhappy. *Bulletin of Science, Technology & Society, 6,* 356–372.

Community Empowerment (Scotland) Act. (2015). Retrieved January 21, 2020, from http://www.legislation.gov.uk/asp/2015/6.

Cross, R. M. (2004). Exploring attitudes: The case for Q methodology. *Health Education Research, 20*(2), 206–213.

Cuppen, E., Breukers, S., Hisschemöller, M., & Bergsma, E. (2010). Q methodology to select participants for a stakeholder dialogue on energy options from biomass in the Netherlands. *Ecological Economics, 69*(3), 579–591.

Davies, B. B., & Hodge, I. D. (2007). Exploring environmental perspectives in lowland agriculture: A Q methodology study in East Anglia, UK. *Ecological Economics, 61*(2–3), 323–333.

Doody, D. G., Kearney, P., Barry, J., Moles, R., & O'Regan, B. (2009). Evaluation of the Q-method as a method of public participation in the selection of sustainable development indicators. *Ecological Indicators, 9*(6), 1129–1137.

Ellis, G., Barry, J., & Robinson, C. (2007). Many ways to say 'no', different ways to say 'yes': Applying Q-methodology to understand public acceptance of wind farm proposals. *Journal of Environmental Planning and Management, 50*(4), 517–551.

Everett, G. (2015). Measuring national well-being: A UK perspective. *Review of Income and Wealth, 61*(1), 34–42.

Fu, B. J., Su, C. H., Wei, Y. P., Willett, I. R., Lü, Y. H., & Liu, G. H. (2011). Double counting in ecosystem services valuation: Causes and countermeasures. *Ecological Research, 26*(1), 1–14.

Fujiwara, D., & Campbell, R. (2011). *Valuation techniques for social cost-benefit analysis: Stated preference, revealed preference and subjective well-being approaches: a discussion of the current issues*. London: HM Treasury, Department for Work and Pensions. Retrieved June 15, 2018, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/209107/greenbook_valuationtechniques.pdf.

Gall, S. C., & Rodwell, L. D. (2016). Evaluating the social acceptability of Marine protected areas. *Marine Policy, 65,* 30–38.

Gerston, L. N. (2014). *Public policy making: Process and principles* (3rd ed.). New York, NY: Routledge.

GOV.UK. (2013). Wellbeing policy and analysis. An update of wellbeing work across Whitehall. GOV.UK. Retrieved June 15, 2018, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/224910/Wellbeing_Policy_and_Analysis_FINAL.PDF.

Hezri, A. A., & Dovers, S. R. (2006). Sustainability indicators, policy, governance: Issues for ecological economics. *Ecological Economics, 60,* 86–99.

HM Treasury. (2018). The Green Book: Central government guidance on appraisal and evaluation. London: HM Treasury. Retrieved January 21, 2020, from https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent.

Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying public policy: Policy cycles and policy subsystems* (3rd ed.). Oxford: Oxford University Press.

Institute for Government. (2018). Grade structures of the civil service. The Institute for Government. Retrieved June 15, 2018, from https://www.instituteforgovernment.org.uk/explainers/grade-structures-civil-service.

Jackson, T. (2010). Reviewing the research: Report of the commission on the measurement of economic performance and social progress. *Environment, 53*(1), 38–40.

Jones, B. D. (2002). Bounded rationality and public policy: Herbert A. Simon and the decisional foundation of collective choice. *Policy Sciences, 35*(3), 269–284.

Kampen, J. K., & Tamás, P. (2014). Overly ambitious: contributions and current status of Q methodology. *Quality & Quantity, 48*(6), 3109–3126.

Kubiszewski, I., Costanza, R., Franco, C., Lawn, P., Talberth, J., Jackson, T., et al. (2013). Beyond GDP: Measuring and achieving global genuine progress. *Ecological Economics, 93,* 57–68.

Lancaster, K. (2017). Confidentiality, anonymity and power relations in elite interviewing: Conducting qualitative policy research in a politicised domain. *International Journal of Social Research Methodology, 20*(1), 93–103.

Lehtonen, M., Sébastien, L., & Bauler, T. (2016). The multiple roles of sustainability indicators in informational governance: Between intended use and unanticipated influence. *Current Opinion in Environmental Sustainability, 18,* 1–9.

Leiserowitz, A. (2006). Climate change risk perception and policy preferences: The role of affect, imagery, and values. *Climatic Change, 77*(1–2), 45–72.

Marmot, M. G. (2004). Evidence based policy or policy based evidence? *BMJ, 328,* 906–907.

Matheson, J. (2011). Measuring what matters: National statistician's reflections on the National debate on measuring wellbeing. London: Office for National Statistics. Retrieved June 15, 2018, from https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/uknationalwellbeingindex.

National Assembly for Wales. (2015). Well-being of future generations (Wales) Act 2015. The Stationery Office Limited. Retrieved June 15, 2018, from https://futuregenerations.wales/wp-content/uploads/2017/01/WFGAct-English.pdf.

O'Neill, S. J., Boykoff, M., Niemeyer, S., & Day, S. A. (2013). On the use of imagery for climate change engagement. *Global Environmental Change, 23*(2), 413–421.

Ockwell, D. G. (2008). 'Opening up' policy to reflexive appraisal: A role for Q methodology? A case study of fire management in Cape York. *Australia. Policy Sciences, 41*(4), 263–292.

OECD. (2018). Better life initiative: Measuring well-being and progress. OECD. Retrieved June 15, 2018, from http://www.oecd.org/statistics/better-life-initiative.htm.

Office for National Statistics. (2017). Social capital in the UK: May 2017. Office for National Statistics. Retrieved January 10, 2020, from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/socialcapitalintheuk/may2017.

Office for National Statistics. (2018). Measuring national well-being: Domains and measures. Office for National Statistics. Retrieved June 15, 2018, from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/datasets/measuringnationalwellbeingdomainsandmeasures.

Office for National Statistics. (2019a). UK natural capital accounts: 2019. Office for National Statistics. Retrieved January 10, 2020, from https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/uknaturalcapitalaccounts/2019.

Office for National Statistics. (2019b). Human capital estimates in the UK: 2004 to 2018. Office for National Statistics. Retrieved January 10, 2020, from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/humancapitalestimates/2004to2018.

Office for National Statistics. (2019c). Human capital indicators consultation. Office for National Statistics. Retrieved January 10, 2020, from https://consultations.ons.gov.uk/well-being-inequalities-sustainability-and-environment/indicator-based-approach-to-measuring-human-capita/supporting_documents/Human_capital_consultation_final.pdf.

PWC. (2014). 11 Quality of life: Assessment. Airports Commission. PricewaterhouseCoopers LLP. Retrieved June 15, 2018, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/372165/11-Quality_of_life–quality-of-life-assessment.pdf.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved June 15, 2018, from https://www.R-project.org/.

Rich, R. F. (1991). Knowledge creation, diffusion, and utilization: Perspectives of the founding editor of Knowledge. *Knowledge, 12*(3), 319–337.

Rinne, J., Lyytimäki, J., & Kautto, P. (2012). Beyond the'indicator industry': Use and potential influences of sustainable development indicators in Finland and the EU. *Progress in Industrial Ecology, an International Journal, 7*(4), 271–284.

Scottish Government. (2018). National Performance Framework. Scottish Government. Retrieved June 15, 2018, from http://www.gov.scot/About/Performance/purposestratobjs.

Sébastien, L., & Bauler, T. (2013). Use and influence of composite indicators for sustainable development at the EU-level. *Ecological Indicators, 35,* 3–12.

Sébastien, L., Bauler, T., & Lehtonen, M. (2014). Can indicators bridge the gap between science and policy? An exploration into the (non) use and (non) influence of indicators in EU and UK policy making. *Nature and Culture, 9*(3), 316–343.

Stainton Rogers, R., Stenner, P., Gleeson, K., & Stainton Rogers, W. (1995). *Social psychology: A critical agenda*. Cambridge: Polity Press.

Steelman, T. A., & Maguire, L. A. (1999). Understanding participant perspectives: Q-methodology in national forest management. *Journal of Policy Analysis and Management, 18,* 361–388.

Stiglitz, J., Sen, A., & Fitoussi, J. P. (2009). *The measurement of economic performance and social progress revisited. Reflections and overview*. Paris: Commission on the Measurement of Economic Performance and Social Progress.

The Treasury. (2018). Living standards framework: Introducing the dashboard. New Zealand Government. Retrieved January 21, 2020, from https://treasury.govt.nz/publications/tp/living-standards-framework-introducing-dashboard.

The Treasury. (2019). Budget policy statement 2019. New Zealand Government. Retrieved January 21, 2020, from https://treasury.govt.nz/publications/budget-policy-statement/budget-policy-statement-2019.

Thiry, G., & Roman, P. (2014). The Inclusive Wealth Index. A sustainability indicator, really? HAL Working Paper Series No. 71. Retrieved June 15, 2018, from https://halshs.archives-ouvertes.fr/halshs-01011250.

Turnhout, E., Hisschemöller, M., & Eijsackers, H. (2007). Ecological indicators: Between the two fires of science and policy. *Ecological Indicators, 7*(2), 215–228.

Valenta, A. L., & Wigger, U. (1997). Q-methodology: Definition and application in health care informatics. *Journal of the American Medical Informatics Association, 4*(6), 501–510.

Van den Bergh, J. C. (2009). The GDP paradox. *Journal of Economic Psychology, 30*(2), 117–135.

Van Eeten, M. J. G. (2000). Recasting environmental controversies: A Q study of the expansion of amsterdam airport. In H. Addams & J. Proops (Eds.), *Social discourse and environmental policy: An application of Q methodology* (pp. 41–70). Cheltenham, UK: Edward Elgard Publishing Limited.

Van Exel, J., & De Graaf, G. (2005). Q methodology: A sneak preview. Retrieved June 15, 2018, from www.jobvanexel.nl.

Watts, S., & Stenner, P. (2005). Doing Q methodology: Theory, method and interpretation. *Qualitative Research in Psychology, 2*(1), 67–91.

Weible, C. M. (2008). Expert-based information and policy subsystems: A review and synthesis. *Policy Studies Journal, 36*(4), 615–635.

Well-being of Future Generations (Wales) Act. (2015). Retrieved January 21, 2020, from http://www.legislation.gov.uk/anaw/2015/2/contents/enacted.

What Works Centre for Wellbeing. (2018). Wellbeing in policy analysis. London: What Works Wellbeing. Retrieved January 21, 2020, from https://www.whatworkswellbeing.org/wp-content/uploads/2018/03/Overview-incorporating-wellbeing-in-policy-analysis-vMarch2018.pdf.

Yang, L. (2014). An inventory of composite measures of human progress. UNDP Human Development Report Office: Occasional Paper on Methodology. Retrieved June 15, 2018, from http://hdr.undp.org/sites/default/files/inventory_report_working_paper.pdf.

Zabala, A., & Pascual, U. (2016). Bootstrapping Q methodology to improve the understanding of human perspectives. *PLoS ONE, 11*(2), e0148087.

Zagorin, P. (1998). *Francis Bacon*. Princeton: Princeton University Press.