



Development
Progress

Project Note

04

Incorporating people's values in development

Weighting alternatives

Laura Rodriguez Takeuchi



Key messages

- In the measurement of multidimensional well-being, weights aim to capture the relative importance of each component to a person's overall well-being. The choice of weights needs to be explicit and could be used to incorporate people's perspectives into a final metric.
- Stated preferences approaches aim to obtain weights from individuals' responses to hypothetical scenarios. We outline six of these approaches. Understanding their design and limitations is vital to make sense of potentially dissimilar results.
- It is important to select and test an appropriate method for specific contexts, considering the challenges of relying on people's answers. Two methodologies, DCE and PTO, are put forward for testing in a pilot project.

This and other Development Progress materials are available at developmentprogress.org

Development Progress is an ODI project that aims to measure, understand and communicate where and how progress has been made in development.

ODI is the UK's leading independent think tank on international development and humanitarian issues. Further ODI materials are available at odi.org.uk

1 Defining Priorities

It is now widely accepted that poverty is multidimensional. Gough and McGregor (2007) propose three universal dimensions of well-being (material, relational and subjective), each of which consists of a cluster of domains, measured by one or more indicators (McGregor and Sumner, 2010). Dimensions are fully universal;

domains are universal but can be expanded or collapsed depending on their purpose; and indicators can be either universal or context specific. Although some would argue that any general framework needs to be adapted to each particular context (e.g. White, 2009), most basic themes that people consider important for well-being are shared across countries and constituencies (Stiglitz et al., 2009: 44).

Buying fresh produce at the market.
Photo: © World Bank Photo Collection

In addition, a growing consensus on the domains of well-being has emerged from the global agreement on the Millennium Development Goals (MDGs), participatory exercises such as the World Bank's 'Voices of the Poor' (Narayan et al. 2000) and initiatives to measure well-being and progress across countries (e.g. the OECD Better Life Index).

Identifying relevant domains is only one part of the exercise of measuring poverty and well-being:

... listing of dimensions of poverty is one thing, but assigning weights to each different aspect of poverty, so policy-makers have some guidance on where to put their (limited) resources, is quite another. (Ravallion, 2010 cited in Melamed, 2011)

A complementary task is to measure the extent to which each part contributes to the whole. Domains can be weighted differently, with weights representing value judgements regarding their relative importance (Alkire, 2008). For example, many people would regard having good health as being more important to their well-being than having voice in community decisions.

Timely information about people's desires could improve policy-makers' ability to allocate resources to maximum effect and to monitor interventions and outcomes, and would also offer greater transparency and accountability (Melamed, 2011). This Methodological Project Note is designed for a pilot study¹ that seeks to build a standard metric to compare the effectiveness of different interventions by accounting for their respective contribution to well-being. By comparing the cost of achieving an equivalent gain in well-being across different types of intervention, such a tool makes the trade-offs between different possible uses of resources more transparent, and could provide a useful means to ensure that resources are allocated in a systematic and rigorous manner. An integral component of such a metric is to obtain a set of weights that people place on different domains of well-being. Such weights would make it possible to aggregate the effect of an intervention on overall well-being.

This Note presents an overview of the main methods that could be used to obtain weights based on people's perceptions. Taking as a point of departure the experience of the health sector in eliciting quality of health weights, the Note highlights the strengths and drawbacks of each method in relation to the pilot case study. It also discusses the

challenges that need to be considered in order to incorporate such methods in the pilot, particularly the possible biases inherent in exercises which rely on people's answers.

The Note is structured as follows. Section 2 presents the concept of weights and their role in a well-being index as a means to measure development interventions. The next section describes in more detail some of the main weighting options and section 4 concentrates on common issues, specifically the issue of aggregation. The note ends by drawing out conclusions and suggesting possible next steps.

2 The role of weights in a well-being index

In measuring well-being, there is growing consensus on the set of domains that are important to people's lives, but less so on the way in which different domains should be prioritised (Coulthard et al., 2014). What matters more to different people should have a greater weight in their overall level of well-being, such that an increase in a more important factor will improve overall well-being more than a similar change in a less important one (Russell and Hubley, 2005). Weights aim to capture the relative importance of the components of well-being as a whole, but can also be used to aggregate sub-components within each domain (see Box 1).

Determining how to weight domains of well-being is not a trivial process. Different methods reflect different underlying assumptions. This makes it imperative to make the choice of method transparent and open to discussion and scrutiny (Anand and Sen, 1997).

Moreover, the use of a particular weighting option, in combination with the way in which domains are measured and aggregated to construct indexes (Decancq and Lugo, 2013), could result in different orderings that would affect comparative resource allocation. This has been found in health interventions² but also in the context of poverty and multidimensional well-being. For example, in different contexts, Decancq et al. (2013) and Santos and Ura (2008) show that different weighting systems, which reflect the different relative importance of certain domains, can have an impact on whether or not the same person is classified as multidimensionally poor.^{3,4} This would clearly have implications for policy targeting on the basis of poverty and for the allocation of budgets and aid. When comparing results and checking the robustness of different methods, the main difficulty is that different methodologies reveal different results, which makes it difficult to determine how to proceed. The valuations

1 See Melamed et al. (2012) for an overview of the motivation and content of the project and Coulthard et al. (2014) for details on the proposed methodology for the pilot.

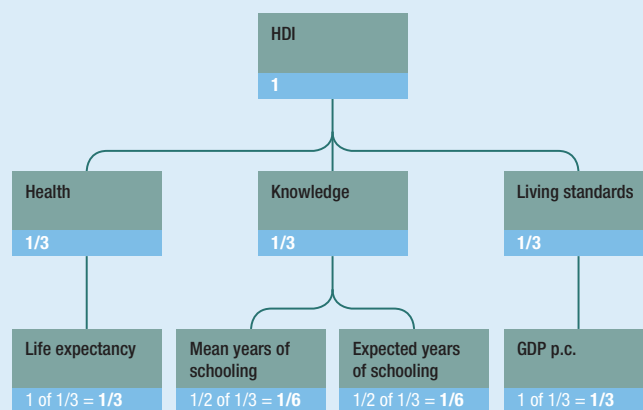
2 For example Marra et al. (2007) found that the use of different instruments to derive weights led to different Quality of Adjusted Life Years (QALY) estimations, which had implications for the statistical and economic evaluation of the effectiveness of treatment for rheumatoid arthritis. They found a difference of over 100% in the incremental cost of QALY between the lowest and the highest estimates.

3 Decancq et al. (2013) compare a hypothetical outcome for a multidimensional poverty index based on a broad range of data-driven, normative and hybrid weights. They found that only 1.2% of the population would be classified as being in the worst-off group regardless of the weighting scheme, compared to 13.7% if any of the methods were used.

4 In Bhutan, Santos and Ura (2008) show that when the equal weighting approach is used, people would be classified as multidimensionally poor when they are deprived in at least one domain, regardless of which one. By contrast, when using weights based on the Gross National Happiness Survey, the same people would not be considered poor if they lacked access to safe water, for instance, because this domain's weight is less than one.

Box 1: Human Development Index (HDI) components and weights

Health, knowledge and living standards contribute equally in the construction of the Human Development Index (HDI) used in the United Nations Development Programme (UNDP) Human Development Report, and therefore are assigned equal weights.ⁱ In the knowledge domain, two indicators (mean and expected years of schooling) are weighted equally to generate the knowledge score, while the health and living standards scores are each measured by one indicator, which carries the full weight of its corresponding domain. This is called an equal weight nested system. In the HDI, the implication is that the two education indicators each contribute to one sixth of the HDI value, while the income and health indicator each contribute one third.



ⁱ Given the transformation of some variables and the way in which variables and domains are aggregated, the final trade-off is not entirely equal, however. Decancq and Lugo (2013) note the difference between measuring the domains in their original or transformed units. If, for example, Gross Domestic Product (GDP) is measured in logarithms, the trade-off between per capita GDP and life expectancy is not constant across all countries' income level. Otherwise stated, the income per capita required to 'compensate' for one year's greater life expectancy is higher in a rich country than in a poorer one.

Source: <http://hdr.undp.org/en/statistics/hdi/>

generated by a given method may reflect methodological design rather than differences in preferences (Parkin and Devlin, 2006; Torrance and Feeny, 1989).

It follows that understanding the implications of the chosen methodological approach is crucial for the successful application of any single metric in decision-making. It is important always to cross-check results, testing and re-testing methods among different groups in order to identify any significant differences in the frames of reference used by specific sub-groups, limited information and difficulties in interpreting the exercise and the questions involved. Because there is no gold standard, awareness of the cause of the differences in methodologies is fundamental to analyse possibly divergent results both across methodologies and for different groups of people. Some of these issues will be examined further in section 4, following a summary of the different weighting methods in section 3.

3 Weighting options

Three methods are discussed in this section: normative, data-driven and elicited weights. The main focus is on the potential of elicited weights as they rely directly on people's perceptions of well-being. This is not an exhaustive list, but reflects a selection of methods most used in health economics and development, and which could potentially be used to construct a well-being index to guide decisions about how to allocate budgets.

3.1 Normative weights

Normative weights in well-being rely on value judgements about what 'ought to be' a good life, although some argue that normative weights inevitably involve some degree of paternalism (see for example Decancq and Lugo, 2013). Because opinions on what constitutes well-being vary greatly, some people's well-being will be assessed on the basis of other people's value judgements.

If it is judged that domains have equal value, a first option is to use equal weights. For example, Atkinson et al. (2002) suggest selecting domains that are proportionally relevant to what is being analysed and thus have broadly similar, if not equal, weight. Equal weights are implicitly applied when 'no weights' are assigned, in order to keep the level of simplicity deemed necessary for advocacy and straightforward comparisons (Chowdhury and Squire, 2006). Equal weights have advantages in enabling comparability across countries and over time. Using 'no weights' is not more impartial, objective or accurate than any other alternative. This option still reflects trade-offs between domains, assuming an equal contribution of all components to overall well-being, which means that the decision to adopt it must be made explicit.

The best-known example of a normative equal weighting system is the Human Development Index (HDI) (Box 1). Equal weights are also used in some applications of the Multidimensional Poverty Index (MPI) methodology developed by Alkire and Foster (2011), as, for instance, in the cross-country MPI used in the Human Development Report (UNDP, 2010) or in Colombia's application of the MPI (Angulo Salazar et al., 2011). In other cases, such as the MPI for Bhutan (Santos and Ura, 2008), the equal weighting system was used along with other sets of weights for comparative purposes. But societies do not always assign equal importance to well-being domains. In the case of Belgium, for instance, Decancq et al. (2013) show that if the Flemish population was to vote on its well-being preferences, from all nine weighting schemes tested (three data-driven, two equal weighting and four hybrid), an equally weighted multidimensional well-being index would have the least support.

Another normative weighting option is to rely on expert opinions. For instance, Chowdhury and Squire (2006) surveyed academic researchers from 125 countries

about their opinions on the HDI weights, who expressed overall support for the equal weighting system. This raises the question of whose value judgements to consider in the weighting system. If the purpose is to have a metric that reflects people's perceptions of a good life, the reliance on experts' opinions is harder to justify. It would nevertheless be possible to compare different sets of weights, for example those obtained from experts, donors and those affected by development interventions.

- There would be practical advantages in using equal weights for the construction of a well-being index. First, the data requirements are low as the only information needed is the total number of domains, each of which would be weighted equally. Moreover, it would make it possible to compare different contexts and be based on normative preferences. It would not, however, be possible to account for the diverse value judgements about well-being made by different people or groups of people.

3.2 Data-driven weights

The idea behind data-driven weights is to 'allow available data to speak for themselves in determining the relevant variables and optimal weights assigned ... rather than making a priori assumptions' (Njong and Ningaye, 2001: 3).

A common data-driven option is to use frequency weights, assigned according to the frequency of achievements (or deprivations) in the population under study. This option would require a preliminary scan to measure deprivations in the relevant area or a secondary database from which to extract this information (e.g. a recent census or household survey). Weights would be assigned on the basis of relative deprivations, for instance giving a higher value to the well-being domain in which there are higher levels of deprivation. This type of weights requires reliable and up-to-date data on the distribution of types of deprivation in the population to avoid allocating resources on the basis of inadequate or poor data. For example, this option was discarded when the MPI was being designed for Colombia because of concerns about using outdated data that would not accurately reflect the country's current situation (Angulo Salazar et al., 2011). It is possible to give a higher weight to better-quality, more recent or more complete data. This might improve the reliability of the measure in question, but could have a negative bias towards domains that are harder to capture⁵ or that are more relevant in areas or for population groups that are harder to reach.

Similarly, it is possible to use statistical techniques such as principal components analysis. This method aims to reduce redundant data, that is, to keep only the data that carry the greater part of the information sought. It enables identifying unique combinations of domains in

overall well-being and using the coefficients obtained as weights. For example, if we were trying to measure material well-being using five indicators – income, employment, consumption, housing quality and access to water and sanitation – only those indicators that contribute most to the overall variance in well-being would be used. For example, income may not add much to the total variance if it has already been captured in the indicator on employment status. This method would require data on total well-being, as well as on each of the constituent domains, in order to identify how much each one contributes to the whole and which domains or indicators are redundant.

- Frequency weights and statistical techniques could be alternatives to normative weights. They rely on existing information about the state and distribution of well-being domains in a given location. A purely statistical methodology can, however, be rigid and not easy to apply in specific policy settings (Njong and Ningaye, 2001). Moreover, statistical methods can lead to normatively contradictory results since some domains may be eliminated because they are statistically redundant even when they are normatively important. For these reasons, a data-driven methodology is perhaps better used to narrow down indicators within domains, rather than to weight between them (Decancq and Lugo, 2013).

3.3 Eliciting weights from preferences

Some methodologies try to obtain a more direct measure of the trade-offs involved in how people reach their preferences. The idea is either to ask people directly by using stated preferences methods, or to use information obtained by observing the choices people make in practice, such as those made in budget decisions, using revealed preferences methods. Economists have traditionally favoured the latter (Kahneman and Krueger, 2006), but in health and in behavioural economics,⁶ it is more common to use stated preferences. Reflecting on the psychology literature, Kahneman and Krueger (2006) argue that it is difficult to deduce preferences from people's actual choices because of limited rationality:

[People] make inconsistent choices, fail to learn from experience, exhibit reluctance to trade, base their own satisfaction on how their situation compares with the satisfaction of others and depart from the standard model of the rational economic agent in other ways. (Kahneman and Krueger 2006: 3)

Rather than using these 'real' choices, stated preferences approaches rely on surveys to obtain weights from individuals' responses to hypothetical scenarios.

⁵ See for example OPHI's Missing Dimensions project, described at: <http://www.ophi.org.uk/research/missing-dimensions/>.

⁶ Experimental games are also commonly used in behavioural economics. They are designed to observe people's choices through experiments (i.e. trying to control the influence of external conditions), and to infer people's motives from these choices. Although the term 'behavioural experiment' often refers to these controlled choice experiments, experimental approaches more broadly (e.g. Randomised Controlled Trials) also rely on respondents' answers (i.e. stated preferences) to obtain information about preferences.

Box 2: An empirical application: The UK's Quality Adjusted Life Years

The UK's National Institute for Health and Clinical Excellence (NICE) bases its recommendations for the allocation of resources largely on a comparison of the unit cost of a given intervention per Quality of Adjusted Life Years (QALY) achieved by each intervention. Because health benefits are measured in different metrics depending on the type of intervention (e.g. children immunised, cholesterol reduced, mobility improved, life years gained, lives saved), the standard cost-benefit analysis has to be adjusted accordingly. QALYs are the 'single currency' designed to compare the well-being benefits of any health intervention, applying a quality weight to the length of life lived in a given state of health. A year lived in perfect health has a quality weight of one, and a year in less than perfect health has a quality weight of less than one. Even negative values are possible in the case of 'worse than death' scenarios (Phillips, 2009).

Quality weights for each health state are derived from a questionnaire called the EQ-5D.ⁱ This examines five dimensions that comprise quality of life in relation to health: **mobility, self-care, usual activities, pain/discomfort and anxiety/depression**. Weights reflect the relative preference of each of these dimensions. To obtain the weights, a sample of the general population is asked about their preferences regarding hypothetical scenarios, which are based on variations of the five health dimensions. The most common methodologies used to elicit weights from the EQ-5D are the Time Trade-off (TTO) and the Visual Analogue Scale (VAS).

i More information is available at: <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d.html/>.

Here, we outline six approaches to eliciting weights using stated preferences. Some illustrative questions for each method are presented in Appendix 1. We start from experiences in the UK health sector where these methods have been used to elicit weights that later informed decisions about resource allocation (Box 2).

The idea behind the **Standard Gamble (SG)**, **Time Trade-Off (TTO)** and **Person Trade-Off (PTO)** is that trade-offs between dimensions are derived from participants' responses to scenarios. The scenarios are combinations of different levels of the health dimensions. In the case of the EQ-5D, for example, there are five levels for each health dimension: no problems, slight problems, moderate problems, severe problems, and extreme problems,⁷ and a scenario could be described as follows:

- *severe* problems in walking about
- *moderate* problems in washing or dressing
- *unable* to do usual activities
- *moderate* pain or discomfort
- *severely* anxious or depressed

Respondents are asked to imagine living in this state and are then given the relevant choice exercise. In SG the choice is between a certain state (i.e. remaining in the scenario described) and a gamble: an intervention that is likely to be successful but that also carries the risk of not working and leaving the person in a worse state than before (Tolley, 2009). In TTO the choice is made in terms of time lived (Tolley, 2009). It asks how much time in a perfectly good scenario (i.e. no problems in walking about, no problems in washing and dressing, being able to do usual activities, no pain or discomfort, not being anxious

or depressed) a respondent would be willing to sacrifice in order to avoid the negative scenario being valued. In PTO the choice is presented in terms of numbers of people.⁸ Respondents are asked to choose between two improvements that affect different numbers of people (e.g. treating/curing, say, 10 people living in scenario A or helping 100 people in scenario B). The number of people in scenario A is altered until the respondent finds the two groups of people equivalent in terms of needing help. In each of these methods, the higher the response – the probability of success, the time willing to be sacrificed and the ratio of people to be helped – the greater relative importance respondents give to that scenario.

The application of any of these three techniques requires first and foremost that the scenarios can be described and understood. In the context of deriving preferences regarding the domains of well-being this involves defining, for each domain, the possible states it might take. In the simpler version of the EQ-5D, there are three levels for each health dimension: extreme problems, some problems, no problems. Using a similar categorisation in the context of well-being, the levels could be expressed in terms of severe, moderate or no deprivations. More levels could be added as necessary in a particular context, although the tasks involved in these exercises are complex, increasing with each additional level of disaggregation. In addition, the health dimensions used in the EQ-5D are relatively closed and restricted, and the three categories (severe, moderate and no problems) may be enough to capture the content of each dimension accurately. In contrast, it may be more difficult to define a more complex concept ('well-being domain') with so few categories.

⁷ A simpler version only includes three levels: no problems, some problems, extreme problems.

⁸ Following the first ODI workshop, Watson suggested investigating the possibility of using PTO (personal communication to Claire Melamed, 27 March 2012).

In addition, in choosing among these three options, further complexities are related to the underlying assumptions, and choosing among them would depend on how realistic and easy to understand they are in the context in which they would be applied. The differences between these three alternatives relate to the unit of account (risk, time and number of people) (Box 3). Moreover, the task involved in all these methods is highly complex, requiring that individuals understand the question and are able to judge risk, time, or proportions, making it difficult for individuals to complete the exercise.

In response to some of those difficulties, **Discrete Choice Experiments (DCE)** and **Rating Scales** have been used as alternatives. They assume ordinal rather than cardinal preferences (Flynn, 2010), which means they emphasise the ranking of preferences rather than the size of the difference between two options. This simplifies the task for the respondents and could be useful if the objective is to derive rankings among dimensions of well-being.

DCE consists of asking respondents to compare pairwise combinations of the same scenarios used in SG, TTO and PTO.⁹ The principle is that the relative difference in preferences between two options can be estimated from the frequency that one is chosen over the other (Flynn et al., 2008). This method was used to determine deprivation weights in the UK 2004 Index of Multiple Deprivation (Dibben et al., 2007). Respondents were asked to choose between pairs of deprivations by stating which they thought was worse and needed greater government support. Each deprivation could take two possible values: deprived or not deprived. The respective levels were described to give detail and context (e.g. being deprived in education was described as having no educational qualifications).

Rating Scales are widely used because they are relatively simple, practical and acceptable (Green et al., 2000). The task consists of presenting respondents with a line or bar with numerically or verbally defined end-points (e.g. from 0 to 10, from very bad to very good), along which they place their response. Rating scales are commonly used

Box 3: Trading-off risk, time and number of people

Standard Gamble (SG) makes key assumptions about individuals' risk behaviour. SG assumes that individuals are always risk-averse, which would be manifested, for instance, by preferring safer options when the gamble is presented. In health, this may hold for common diseases, for example, where the probability of occurrence is high but the damage is not major. However, it is possible that a person's risk behaviour differs across health conditions and over their life-time (Green et al., 2000). Following the health example, in some life-threatening circumstances individuals would act as risk-takers, demanding treatments with a low likelihood of success but that may offer a great benefit to a small proportion of those affected (Loomes and McKenzie, 1989). If different risk behaviours are not taken into account, the results of an SG exercise could suggest that people do not attach much importance to those scenarios that are very unlikely to occur, but would have a clearly negative effect on well-being, when the case is that they find it difficult to infer very low probabilities, and compare the scenarios with more frequent events. Consequently, this method is more appropriate for assessing scenarios with a comparable probability of occurrence.

In contrast, TTO assumes that choices are made in conditions of complete certainty, so that risk-taking (or risk-averse) behaviour is not an issue. Rather, the complexity arises from the use of time as the unit of reference. TTO assumes that time is valued equally at different points in a person's life, or at least that at different points of a person's lifetime there is a constant proportion between the preference for time spent in a good and a poor scenario (Buckingham and Devlin, 2009). Because weights are calculated as a proportion of the remaining lifetime that is forgone, the age of the respondent and the age at which the extra year is to be sacrificed may influence the willingness to trade time (Arnesen and Trommald, 2005). For example, older people may be less willing to trade off one life year since it corresponds to a larger proportion of their remaining life expectancy (Flynn, 2010). Options to correct the timeframe effects include specifying that the trade-off would occur at a specific point (e.g. at the end of the expected life span) or analysing the results taking into account some of the respondent's characteristics (age, sex, etc.).

The attractiveness of PTO is that the values obtained reflect the societal value of interventions because respondents are asked to reflect on the value of aggregate outcomes rather than on their individual experiences. This is felt to be a good representation of the reality of policy-making (Nord et al., 1999 cited in Green, 2001; Damschroder et al., 2005; Arnesen and Nord, 1999). Equity and fairness concerns could be incorporated, for instance, by including information on people's baseline level or by giving greater value to those starting in a worse condition (Weinstein et al., 2009). One disadvantage is that it is often difficult to understand the framing of the question. Also, respondents may feel reluctant to give an answer because they find it unpleasant or unethical to make choices about others' lives (Green, 2001; Damschroder et al., 2007) or because it refers to stigmatised issues (e.g. HIV/AIDS). In addition, PTO has been far less used than other methods, and there is limited empirical evidence regarding its practicability, reliability and empirical validity (Green, 2001). Its main application, in the disability weights of the Global Burden of Disease, was criticised for rendering inconsistent results (Arnesen and Nord, 1999).

⁹ Respondents are usually stacked in groups to answer a few sub-set scenarios because of the numerous possible combinations – a typical DCE of 16 scenarios with a yes/no response would lead to a 65,536 possible response patterns (Louviere, 2006).

to evaluate services by asking respondents to state how far they agree with a statement (e.g. ‘On a scale from x to y , how friendly were the staff?’; ‘How would you rate the waste collection service in your neighbourhood on a scale from very bad to very good?’). One popular rating scale used in preference elicitation is the Visual Analogue Scale (VAS). It has been found useful in contexts where respondents lack the levels of numeracy and literacy required to complete more complicated choice tasks. The approach also avoids the problem of non-constant risk behaviour or time valuing that arises in TTO or SG (Parkin and Devlin, 2006). Despite these advantages, critics argue that the task presented in rating scales does not involve a trade-off choice (Green et al., 2000; Tolley, 2009). The argument is that the questions associated with each option do not take into account its cost or sacrifice, and that as a result respondents would always prefer more rather than less of whatever option they are considering (e.g. more staff, better service quality, more schools built, more vaccines provided). Proponents suggest that the resulting relationships between variable scores can be used to derive viable trade-offs (Green et al., 2000; Parkin and Devlin, 2006). Imposing an overarching restriction could help to solve this limitation. For example, limiting the total number of points to allocate across all domains or categories, could make the choice an explicit trade-off, rather than just a rating exercise (i.e. budget allocation

exercises which ask participants to allocate a given amount of money – or stones, sticks, or other objects representing money – between a fixed number of categories).

The elicitation methods discussed so far are generally applied to a number of respondents on the basis of a questionnaire, the responses to which can then be aggregated. It is also possible to derive weights directly from groups. Although a more detailed review of group-based methodologies is beyond the scope of this Project Note,¹⁰ we describe Swing Weights as an example of a group-based methodology suggested at a workshop discussion for this project. **Swing Weights** have been used mainly in Multi-Criteria Decision Analysis (MCDA). Weights define the relative importance of different decision-making criteria in a way similar to reflecting on the relative importance of well-being domains. The idea is to start from a baseline scenario in which all domains are in their worst possible state (e.g. being in bad health, unemployed, with no voice in community decisions, living in inadequate housing). Respondents choose domains they wish to improve first, second, third, etc., and compare a shift from the top to the bottom of the scale of the first criterion with a similar shift in the remaining domains. For example, in a study to decide on the location of radioactive waste in the UK, participants had to choose between the criteria of radiation, public safety, worker safety, environment and flexibility as the aspects they thought should be improved first in a radioactive location. They were



Farmers in Kenya. Photo: © Annie Bungeroth/CAFOD

¹⁰ A separate review of participatory methods has also been conducted for this project.

also asked about the size of the improvement compared to improvement in each of the other decision-making criteria. Swing weights make it possible to see the value of each domain in a specific context. This is useful when thinking of applying the methodology to the measurement of well-being because the context and its restrictions can change the way in which different options are valued. For example, in a country in which the population has low levels of education, a domain that includes expanding employment in the ICT sector may not be deemed relevant, at least until the population has the education and skills to be able to take advantage of such job opportunities.

Table 1 summarises the advantages and disadvantages of the options described above in the context of eliciting weights that reflect people's perspectives. We turn next to discussing common problems to be addressed, especially in using elicited weights.

The methods are based on different theoretical assumptions and also differ in terms of how realistic they are in the specific context. It is worth examining the practicality, reliability and validity (Box 4) of any valuation technique before using it. For example, the VAS stands out as a practical method for more simple tasks, but its theoretical validity may not be as high as that of the other options. Because there is no standard methodology, it is vital to conduct some sensitivity analysis. This may not lead to a convergence in results across different methods, but is important to understand the specific advantages and limitations of a method and the origin of the potentially dissimilar results. For example, altering the timeframe used in TTO methodologies would help to understand how sensitive the results are to such differences and how different sub-groups of people respond to such changes.

Table 1: Summary of weighting alternatives

Weighting system		Advantages	Disadvantages
Normative	Equal weights	Simplicity. There is no need to rely on elicited values or secondary data (although it is possible that an elicitation technique leads to an equal weighting system).	Choices and trade-offs that need theoretical back-up. Heavily depends on whose value judgements are considered.
	Frequency of deprivations, principal components	Can use existing data from surveys, census or administrative records.	Statistical techniques may be hard to understand and difficult to use. Results driven by the data may not conform with normative value judgements or the preferences expressed by people. Needs up-to-date and reliable data.
Elicited	Revealed preferences	Based on observations of actual choices.	Can be biased if individuals make decisions on the basis of limited rationality.
	Standard gamble and time trade-off	Theoretically strong. They assess the trade-offs between well-being domains.	Can be difficult for individuals to complete the exercise. The definition of the scenarios may be too complex. They depend on assumptions about risk behaviour or timeframe that may not be constant across scenarios or the life-time of a person.
	Person trade-off	Can capture equity and fairness issues. Assesses the social value of different states.	Respondents may be unwilling or find it difficult or unpleasant to answer, and the exercise may be complex for individuals to understand. It has not been extensively validated.
	Discrete choice experiments	Similar to SG and TTO but offers participants a simpler task.	The choice sets can be large. Results represent rankings rather than magnitudes.
	Visual analogue and rating scales	Practical, less costly and quicker.	It may result in people always preferring greater values because the cost (trade-off) of the options is not made explicit.
	Swing weights (group elicitation)	Make it possible to see values in a specific context.	Clearly defining the worst/best scenarios is not simple and needs to be done beforehand (this applies to all methods). Group dynamics can lead to unrepresentative results for certain individuals or sub-groups.

Box 4: Criteria to assess weighting methodologies

- Practicality: Depends on its acceptability to respondents. A function of length, complexity and respondents' interest in the task.
- Reliability: Separate administrations of the method should yield the same results.
- Empirical Validity: The ability to predict preferences on the basis of what actual decisions reveal (measures what it intends to measure).
- Theoretical Validity: Validity of the arguments presented in the literature to support the use of the valuation technique.

Source: Green et al., 2000.

- When trying to obtain a direct measure of the trade-offs involved in people's preferences, normative and data driven methods may not respond to the diverse value judgements about well-being made by different people or groups of people, but rather reflect normative expectations or data patterns. When comparing the stated preference methods reviewed here, and in the context of the pilot project case studies – three development projects in health, environment and livelihood conflicts, and social exclusion and gender (Coulthard et al. 2014) – it is critical to have a practical and understandable set of questions and scenarios. SG and TTO involve complex exercises, which limit their application in contexts characterised by low numeracy and literacy skills. DCE is a more time-consuming method because of the large set of pair scenarios, but it has the theoretical validity of more complicated methodologies while remaining simple enough for people to understand. Moreover, the complete set of pairwise combinations of scenarios (the choice set) can be split and administered to sub-groups, or the choice set can be reduced as appropriate (Mangham et al., 2009).¹¹
- It is also worth testing PTO in a pilot, given that it can approximate to the social value of different interventions. It may work particularly well in the case of environment and livelihoods conflicts, where interventions affect a large number of people and may have important environmental and social externalities. Since there has been considerably less experience in applying this elicitation method, it would be innovative to test it. Given the concerns about the consistency of results, it would be vital to subject the design to sensitivity testing, and to compare different types of PTO question.

4 Limited information, aggregation and further issues

The two main concerns in studies that rely on subjective data for decision-making are the problems of limited information and aggregation.

First, the way people evaluate their life may be affected by their own values, their current state, their knowledge of alternatives, an assessment of how much room there is for improvement, and their interpersonal frames of reference (Alkire, 2008). Sen (1999) highlighted the problem that poor people abandon a desire because of their life-long habituation to their living conditions:

The utilitarian calculus based on, say, happiness can be deeply unfair to those who are persistently deprived, such as the traditional underdogs in stratified societies, oppressed minorities in intolerant communities, precarious sharecroppers living in a world of uncertainty, sweated workers in exploitative industrial arrangements, subdued housewives in deeply sexist cultures. The deprived people tend to come to terms with their deprivation because of the sheer necessity of survival, and they may, as a result, lack the courage to desire any radical change, and may even adjust their desires and expectations to what they unambitiously see as feasible. But the adjustments also have the incidental effect of distorting the scale of utilities. (Sen, 1999: 62-63)

This means that in expressing their preferences, people may respond not with their ideal preferences but rather to preferences based on limited information. For example, when asked about their contraception preferences, adolescent girls may respond on the basis of limited information about sexual rights and the social expectations of family and childbearing. Research has shown that one barrier to expanding adolescent girls' use of contraception is that, having internalised sociocultural and religious expectations, they wish to become pregnant and thus decline contraception (Jones and Presler-Marshall, 2012).

Providing sufficient information is perhaps the only way to address this issue. There are examples that illustrate that if there is self-understanding of citizenship and of having rights, this becomes less problematic (Nussbaum, 2001). A South African study on social perceptions of material needs found that although people who defined an item as essential (such as having a bath or shower in the house) were more likely to have it, those lacking such facilities also thought it was an essential need (Wright, 2008). This works in two ways: those lacking the socially perceived necessities appear to be aware of their deprivation, and those who are not deprived also feel that these constitute a general requirement for the entire population. The study also finds consensus among different groups about what they define as

¹¹ Devlin also suggested this is a viable alternative often used in health economics (personal communication, 18 February 2014).

essential needs and a decent standard of living. This is particularly important because it occurs in the context of an unequal and highly divided society in which it would be conceivable that some groups would retain restricted views about their rights and needs.

A related issue is that people's stated preferences are affected by current 'extraordinary' life events. For example, the reported satisfaction associated with events such as getting married or having children is often temporary and has only a minor overall effect on life satisfaction or long-term happiness (Kahneman and Krueger, 2006). This effect, called the hedonic treadmill, is found when lottery winners report being unhappier than expected a year after winning the prize (Gilbert, 2004) and when there is no substantial increase in people's life satisfaction despite their country's rapid economic growth and improved material well-being (Kahneman and Krueger, 2006; Oswald, 1997). The effect of such events tends to fade away over time. This means it becomes less of a concern when people are asked about what is relevant over the long run for their own and others' well-being, and makes it possible to use stated preference information for the purposes of defining weights in a well-being index. In addition, such results can be validated with results from non-participatory approaches, such as value judgements based on existing literature, experts or external evaluators, or the use of statistical techniques to control for unobserved effects.¹²

An important issue to consider is that people may hold different values in their conceptions of a 'good life'. Different people may attach more or less importance to particular components of well-being, but the objective is to arrive at a common value across a population or any sub-group that is relevant to policy-making. Weights that reflect a 'reasoned consensus' of the community (Alkire, 2008) require public discussion, understanding and acceptance (Sen, 1996 cited in Alkire, 2008). Some argue that the aggregation of preferences is inherently undemocratic or paternalistic because the preferences of certain individuals, however well informed, capable or benevolent they may be, are taken to represent social preferences. This is undesirable because it sacrifices participatory decisions and it is insensitive to the interests of a heterogeneous population (Sen, 1999b).

The diversity in preferences has to be acknowledged and it may be relevant to compare how different groups assess their priorities, as well as to compare them with the decision-makers' priorities. In practical terms, the exercise would need to be designed using words that have a common and clear meaning across different sets of respondents in an attempt to fix survey response

categories to the same frame and for respondents to use the scales in a comparable way. The South African example above (Wright, 2008) shows that it is unnecessary to adjust weights at the community or group level if there were no evidence of intra-group variability. However, this is a case of empirical testing. The consensus on what members of a society should have might break down if people were asked more specifically to trade off domains or to prioritise interventions. It is possible that attempts to develop a well-being measurement tool need to be validated in different country contexts as has been attempted for the EQ-5D.¹³

It is also pertinent to think about group approaches in order to derive weights that have some social significance. If we believe that social well-being is more than just the sum of the well-being of individuals (White, 2009) it would potentially be more relevant to derive weights directly from groups. Group exercises have been applied to the study of well-being (including subjective well-being) in rural communities in India and Zambia by the Well-being and Poverty Pathways Project,¹⁴ with groups of children in studies of the experiences and perceptions of poverty by Young Lives,¹⁵ and in public stakeholder meetings to define criteria weights to use in an MCDA to decide where to place radioactive waste in the UK.¹⁶ Such group exercises face other potential shortcomings. In particular, power and group dynamics could lead to resource allocations that do not respect everyone's preferences, so the group's size and composition need particular attention. For instance, the Young Lives studies report that some carers insisted on being involved in the exercise because they felt that their children were being excluded (Tafere and Camfield, 2009).

This also raises questions about sampling and inclusion. The sample could be drawn from the general population or be limited to people affected by a particular project or policy intervention. There may be a case for having a wide reference group when decisions are to be taken about public spending, which is of public concern.¹⁷ Even those not directly affected may have a say – differences in the preferences of different groups (defined for example by gender, age, location or ethnicity) need to be analysed, and public decisions should be made on the basis of recognising that nobody is exempt from experiencing disadvantage and thus should be concerned with the potential impact of a policy project.

Finally, understanding and answering the questions involved in the elicitation exercises is complex and requires a clear explanation of the tasks, and may also call for considerable verbal and numerical skills. This can pose problems, especially if the intention is to apply them

12 Panel data are helpful but their collection is more intensive. Large, random samples with enough variation could help (Arnesen and Trommald, 2005) because it can be expected that on aggregate, weights will not be affected by this type of shocks and behave in a more or less stable way.

13 Although many countries still use the UK or other country values as a proxy.

14 More information is available at: <http://www.wellbeingpathways.org/home/>.

15 More information is available at: <http://www.younglives.org.uk/>.

16 More information is available at: <http://bit.ly/1qdhc1u>.

17 For example this has been the case in participatory budgeting approaches (started in Porto Alegre-Brazil) and gender-responsive budgeting, both examples of a wide reference group taking an active interest in public spending and accountability.

in groups with limited numeracy and literacy levels. Tasks that do not require these skills could be developed with the help of carefully designed and explained visual props. Researchers at the Young Lives project have relied on drawings (Camfield and Tafere, 2009; Young Lives, 2011) while researchers at the Wellbeing and Poverty Pathways Project have used a mix of quantitative methods, with in-depth interviews and community meetings to try to address difficulties some communities experience in interpreting abstract terms (White et al., 2012).

5 Conclusions

This Note has reviewed alternative methodologies to elicit people's preferences regarding their assessments of well-being. Weighting decisions are often tacit, when national budgets are allocated to particular policies or when equal weights are given to domains of well-being. Despite their challenges, elicitation methodologies are one way to make the weighting decisions more transparent and more responsive to people's perceptions about what is more or less important to their well-being.

The difficulties of relying on people's answers and the possible biases inherent in these exercises make

these approaches more demanding than normative or data-driven exercises. Equal weighting and data-driven approaches are more straightforward to apply if the information is readily available, but the weights obtained with these methodologies do not necessarily reflect people's preferences and could hide intra-group diversity. Elicitation techniques seek to obtain information directly from individual or group responses. Some combination of the two types of approaches could be achieved, for example, selecting domains based on theory, but weighting them based on elicitation methods.

There is a range of possible methodologies and careful thought needs to be given to selecting an appropriate method that is feasible for the context in which it is to be applied. Two methodologies, DCE and PTO, are put forward for testing in the pilot case studies.

The stages in the process of assigning weights to the domains of well-being can be summarised as follows:

- **Determine the domains** using literature reviews, participatory approaches, public debate, consultations with experts, etc.¹⁸ For this pilot project, it is proposed to determine the domains using focus group discussions and individual interviews using the Global Person



Street vendor selling glasses. Ahmedabad, India. Photo: via Flickr User sandeepachetan

¹⁸ This Project Note assumes that these are already defined.

Generated Index (Camfield and Ruta, 2007; Woodcock et al., 2009).¹⁹ These results will be contrasted with theoretical or expert definitions of universally relevant well-being criteria.

- It is useful to **identify marker states**. These describe a typical good, medium and poor outcome for each domain. Although these outcome states can be cardinalised, this is done only once the responses are gathered. It is better not to attach numerical values to verbal responses, especially in the context of working with individuals or groups who would find such versions harder to interpret. Descriptions must be carefully and precisely written, tested, rewritten, and re-tested in order to eliminate all ambiguity.
- **Identify the relevant sampling group**. Keep in mind the possible response biases of particular sub-groups and whether individual or group elicitation is more appropriate.
- **Design, test and apply the chosen methodology**, and if possible, more than one in order to allow for triangulation. Keep in mind the need for carefully designed visual props, precisely written instructions for the interviewer, thorough interviewer training, clear explanation of the purpose of the exercise and the reasons for the chosen method, and continuous monitoring. Recent research on subjective well-being suggests that it is important to focus on questions that refer to the long term rather than current states (Gilbert, 2004; Kahneman and Krueger, 2006; Oswald,

1997) and to test responses in different reference sub-groups to spot possible biases.

- **Conduct sensitivity analysis** and identify the advantages and limitations of the selected method in terms of its practicality, reliability and validity. Triangulation would help in developing a broader approach to perspectives on well-being while strengthening its external validity. It could also help to reconcile quantitative and qualitative research in well-being, recognising that all of the domains may be addressed through both subjective and objective perspectives (White et al., 2012) and different methodological approaches.

The selection of weights is a vital means for incorporating people's values in designing an instrument to evaluate development interventions. Based on the analysis of the methods reviewed, and given that there is no 'gold standard', it is recommended that different methods should be tested and the results compared in the pilot, in particular recognising the existence of heterogeneous responses within groups or sub-groups. The exercise should be seen as a necessary stage in the design of a metric to compare the effectiveness of well-being-enhancing interventions and the trade-offs between competing ways to use resources. The transparency in the final weighting scheme could contribute to including people's own priorities in the development decision-making process.

¹⁹ See Coulthard et al (2014) for more details.

References

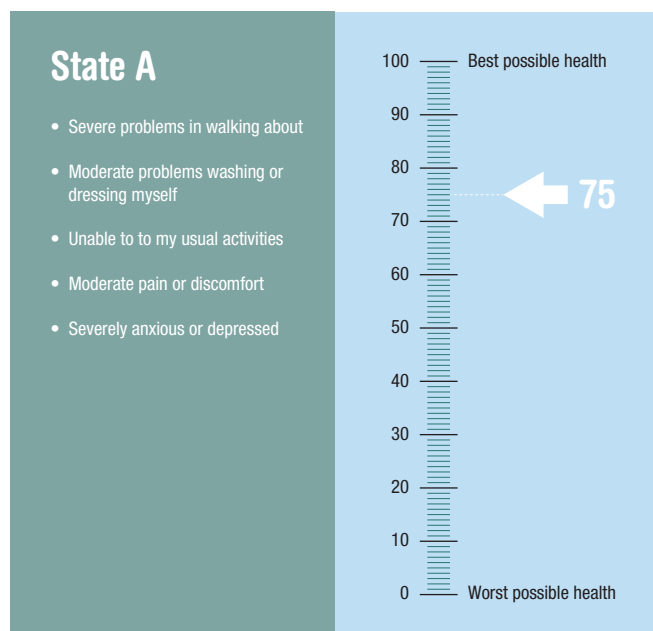
- Alkire, S. (2008) 'The capability approach to the Quality of Life'. Working paper prepared for the Working Group 'Quality of Life'. Available at: http://www.stiglitz-sen-fitoussi.fr/documents/capability_approach.pdf.
- Alkire, S. and Foster, J. (2011) 'Counting and multidimensional poverty measurement', *Journal of Public Economics* 95(7–8): 476–487.
- Anand, S. and Sen, A.K. (1997) 'Concepts of human development and poverty: A multidimensional perspective', *Human Development Papers*: 1–19.
- Angulo Salazar, R.C., Diaz Cuervo, Y. and Pardo Pinzon, R. (2011) 'Indice de pobreza multidimensional para Colombia 1997–2010', *Archivos de Economia* (382): 1–56.
- Arnesen, T. and TROMMALD, M. (2005) 'Are QALYs based on time trade-off comparable? A systematic review of TTO methodologies', *Health Economics* 14(1): 39–53.
- Arnesen, T. and Nord, E. (1999) 'The value of a DALY life: problems with ethics and validity of disability adjusted life years', *British Medical Journal* 319(7222): 1423–1425.
- Atkinson, A.B., Cantillon, B., Marlier, E. and Nolan, B. (2002) *Social Indicators: The EU and Social Inclusion*. Oxford: Oxford University Press.
- Buckingham, K.J. and Devlin, N.J. (2009) 'A note on the nature of utility in time and health and implications for cost utility analysis', *Social Science Medicine* 68(2): 362–367.
- Camfield, L. and Ruta, D. (2007) "Translation is not enough": using the Global Person Generated Index (GPGI) to assess individual quality of life in Bangladesh, Thailand, and Ethiopia', *Quality of Life Research* 16: 1039–1051.
- Camfield, L. and Tafere, Y. (2009) "Children with a good life have to have school bags": Diverse understandings of well-being among older children in three Ethiopian communities'. *Young Lives Working Paper* 37. Oxford: Young Lives, Department of International Development, University of Oxford.
- Chowdhury, S. and Squire, L. (2006) 'Setting weights for aggregate indices: an application to the commitment to development index and human development index', *Journal of Development Studies* 42(5): 761–771.
- CoRWM (2006) 'Managing our radioactive waste safely'. CoRWM's recommendations to Government. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/294118/700_-_CoRWM_July_2006_Recommendations_to_Government.pdf
- Coulthard, S., Camfield, L. and McGregor, J.A. (2014) 'Valuing development progress by focusing on what matters most to people: a demonstration of wellbeing assessment across three development sectors'. Outline proposal, Draft 4, 14 February.
- Damschroder, L.J. et al. (2005) 'Trading people versus trading time: what is the difference?', *Population Health and Metrics* 3(10): 1–13.
- Damschroder, L.J., Roberts, T.R., Zikmund-Fisher, B. and Ubel, P.A. (2007) 'Why people refuse to make tradeoffs in person tradeoff elicitation: a matter of perspective?', *Medical Decision Making* 27(3): 266–280.
- Decancq, K. and Lugo, M.A. (2013) 'Weights in multidimensional indices of well-being: an overview', *Econometric Reviews* 32(1): 7–34.
- Decancq, K., Van Ootegem, L. and Verhofstadt, E. (2013) 'What if we voted on the weights of a multidimensional well-being index? An illustration with Flemish data', *Fiscal Studies* 34(3): 315–332.
- Devlin, N. (2012) 'Weighting measures of health'. Paper presented at 'Valuing Development Workshop', London, 22 March.
- Dibben, C., Atherton, I.M., Cox, M. et al. (2007) *Investigating the impact of changing the weights that underpin the Index of Multiple Deprivation 2004*. London: Department of Communities and Local Governments.
- Flynn, T.N., Louviere, J.J., Marley, A.A.J., et al. (2008) 'Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale'. *Population Health Metrics* 6(6): 1–11.
- Flynn, T.N. (2010) 'Using conjoint analysis and choice experiments to estimate QALY values: issues to consider', *Pharmaco Economics* 28(9): 711–722.
- Gilbert, D. (2004) 'The surprising science of happiness'. TED Talks. Available at: http://www.ted.com/talks/dan_gilbert_asks_why_are_we_happy
- Green, C. (2001) 'On the societal value of health care: what do we know about the person trade-off technique?', *Health Economics* 10(3): 233–243.
- Green, C., Brazier, J. and Deverill, M. (2000) 'Valuing health-related quality of life: a review of health state valuation techniques', *PharmacoEconomics* 17(2): 151–165.
- Jones, N. and Presler-Marshall, E. (2012) *Charting the future: empowering girls to prevent early pregnancy*. London: Overseas Development Institute and Save the Children.
- Kahneman, D. and Krueger, A.B. (2006) 'Developments in the measurement of subjective well-being', *Journal of Economic Perspectives* 20(1): 3–24.
- Loomes, G. and McKenzie L., (1989) 'The use of QALYs in health care decision making', *Social Science & Medicine* 28(4): 299–308.
- Louviere, J.J. (2006) 'What you don't know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments', *Environmental and Resource Economics* 34(1): 173–188.
- Marra, C.A., Marion, S.A., Guh, D.P., et al. (2007) 'Not all "quality-adjusted life years" are equal', *Journal of Clinical Epidemiology* 60(6): 616–624.
- Mangham, L.J. et al. (2009) 'How to do (or not to do)... designing a discrete choice experiment for application in a low-income country', *Health Policy and Planning* 24 (2): 151–158.
- McGregor, A. and Sumner, A. (2010) 'Beyond business as usual: what might 3-D Wellbeing contribute to MDG momentum?', *IDS Bulletin* 41(1): 104–112.
- Melamed, C. (2011) 'Does development give poor people what they want?'. Background Note. London: Overseas Development Institute.
- Melamed, C., Devlin, N., and Appleby, J. (2012) "Valuing development": Could approaches to measuring outcomes in health help make development more accountable?. Research Reports and Studies. London: Overseas Development Institute.
- Narayan, D., with Patel, R., Schaffte, K., Rademacher, A., and Koch-Schulte, S. (2000) *Voices of the Poor Vol. 1: Can Anyone Hear Us?* New York: Oxford University Press for the World Bank.
- Njong, A.M. and Ningaye, P. (2001). 'Characterizing weights in the measurement of multidimensional poverty: An application of data-driven approaches to Cameroonian data'. *OPHI Working Paper* 21. Oxford: Oxford Poverty & Human Development Initiative, Department of International Development, University of Oxford.
- Nussbaum, M.C. (2001) 'Adaptive preferences and women's options', *Economics and Philosophy* (17): 67–88.
- Oswald, A.J. (1997) 'Happiness and economic performance', *The Economic Journal* 107(445): 1815–1831.

-
- Parkin, D. and Devlin, N.J. (2006) 'Is there a case for using visual analogue scale valuations in cost-utility analysis?' *Health Economics* 15(7): 653–664.
- Phillips, C. (2009) 'What is a QALY?' What is...? series, 2nd edn. Oxford: Medical Sciences Division, University of Oxford.
- Russell, L.B. and Hubley, A.M. (2005) 'Importance ratings and weighting: old concerns and new perspectives', *International Journal of Testing* 5(2): 105–130.
- Santos, M.E. and Ura, K. (2008) 'Multidimensional Poverty in Bhutan : Estimates and Policy Implications'. *OPHI Working Paper* 14. Oxford: Oxford Poverty & Human Development Initiative, Department of International Development, University of Oxford.
- Sen, A. (1999a) *Development as Freedom*. Oxford: Oxford University Press.
- Sen, A. (1999b) 'The possibility of social choice', *American Economic Review* 89(3): 349–378.
- Stiglitz, J.E., Sen, A. and Fitoussi, J.-P. (2009) 'Report by the Commission on the Measurement of Economic Performance and Social Progress'. Paris: Commission on the Measurement of Economic Performance and Social Progress.
- Tafere, Y. and Camfield, L. (2009) 'Community Understandings of Children's Transitions in Ethiopia: Possible Implications for Life Course Poverty'. *Young Lives Working Paper* 41. Oxford: Young Lives, Department of International Development, University of Oxford.
- Tolley, K. (2009) 'What are health utilities?' What is...? series. Oxford: Medical Sciences Division, University of Oxford.
- Torrance, G.W. and Feeny, D. (1989) 'Utilities and quality-adjusted life years', *International Journal of Technology Assessment in Health Care* 5(4): 559–575.
- UNDP (2010). *Human Development Report 2010. The Real wealth of Nations: Pathways to Human Development*. New York: UNDP
- Weinstein, M.C., Torrance, G., and McGuire, A. (2009) 'QALYs: the basics', *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 12 (s1) S5-S9, pp.S5–S9.
- White, S.C. (2009). 'Analysing wellbeing: a framework for development practice', *Development in Practice* 20(2): 158–172.
- White, S.C., Gaines, S.O. and Jha, S. (2012) 'Beyond subjective well-being: a critical review of the Stiglitz report approach to subjective perspectives on quality of life', *Journal of International Development* 24(6): 763–776.
- Woodcock, A., Camfield, L., McGregor, J.A., and Martin, F. (2009) 'Validation of the WeDQoL-Goals-Thailand measure: culture-specific individualized quality of life', *Social Indicators Research* 94(1): 135-171.
- Wright, G. (2008) 'Findings from the Indicators of Poverty and Social Exclusion project: A profile of poverty using the Socially Perceived Necessities Approach'. Key Report 7. Department of Social Development, Republic of South Africa. Available at: [http://www.casasp.ox.ac.uk/docs/IPSE Key Report 7 DSD-CASASP.pdf](http://www.casasp.ox.ac.uk/docs/IPSE%20Key%20Report%207%20DSD-CASASP.pdf).
- Young Lives (2011) *Young Lives methods guide: The qualitative research component*. Oxford: Young Lives, Department of International Development, University of Oxford.

Appendix 1

Sample questions

VAS



Source: Devlin (2012). EQ-5D.

DCE

Example choice set

	Person A	Person B
Crime	Not a victim of crime in last 4 years	Victim of crime in last 4 years
Employment	Unemployed	Employed, retired or looking after home/family
Income	Less than £100 per adult	At least £100 per adult
Health	No limits on daily activity and work	Limits on daily activity and work
Housing quality	Decent	Non-decent
Education	No educational qualifications	Educational qualifications
Convenience of services	Inconvenient	Convenient
Who needs most support?	Person A	Person B

Source: Dibben et al. (2007); UK Index of Multiple Deprivation 2004.

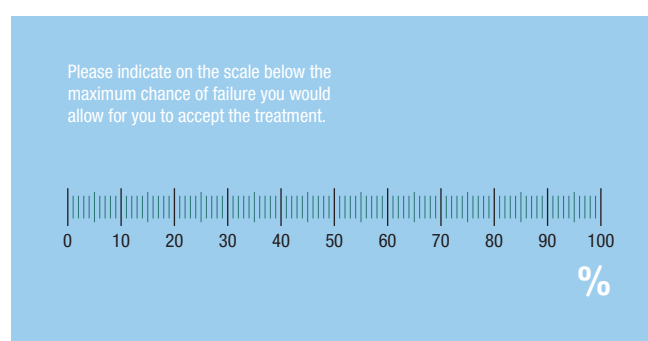
SG

Imagine that you go to your doctor for a routine medical examination. To your surprise, your doctor informs you that you have an unusual health condition. In this condition, your doctor informs you that, without treatment, you will live for 10 more years in good health, and then you will die.

However, your doctor also informs you that there is a treatment for your condition, which, if taken, would give you a chance of living for 40 years in good health before death. However, there is also a chance that the treatment would kill you immediately.

Your doctor tells you that the size of the chance that the treatment will fail is not known.

Please indicate on the scale below the maximum chance of failure you would allow for you to accept the treatment.



Source: Oliver (2004).

TTO

Imagine that you have one friend, Mr./Mrs. Adams, who is 30 years old and has **paraplegia**. This means that he/she is paralyzed from the waist down. You can imagine that his/her bowel, bladder, and sexual function is normal, but he/she cannot move his/her legs at all, and has to use a wheelchair to get around. Mr./Mrs. Adams will live 50 more years with paraplegia, and then die in his/her sleep.

Now imagine that you have another friend, Mr./Mrs. Brown, who is 30 years old and in **perfect health**. This means he/she has no health problems at all. Mr./Mrs. Brown will live 50 more years in perfect health, and then die in his/her sleep.

Which friend do you think is better off or would you consider them to be equally well off? Remember to think out loud as you answer.

If the subject chooses Mr./Mrs. Brown, we ask, “OK, imagine again that Mr./Mrs. Adams will live for 50 more years with paraplegia. Now imagine that Mr./Mrs. Brown will live for a few more days in perfect health. Which friend do you think is better off or would you consider them to be equally well off?”

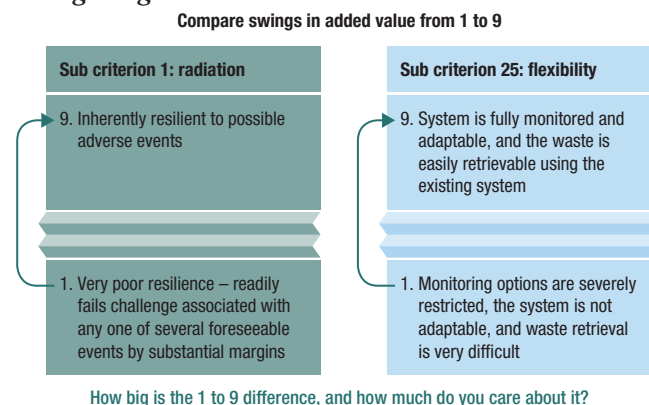
Source: Damschroder et al. (2005). TTO Scenario Used in Save-Save and Cure-Save Groups.

PTO

You are a decision maker who has enough money to buy only one of two mutually exclusive health interventions. If you purchase intervention A, you will extend the life of 1000 healthy [non-disabled] individuals for exactly one year, at which point they will all die. If you do not purchase intervention A, they will all die today. The alternative use of your scarce resources is intervention B, with which you can extend the life of n individuals with a particular disabling condition for one year. If you do not buy intervention B they will all die today; if you do purchase intervention B, they will die at the end of exactly one year.

Source: Arnesen and Nord (2000). *Global Burden of Disease 1996*, PTO question 1.

Swing weights



Source: CoRWM (2006).

Overseas Development Institute

203 Blackfriars Road
London SE1 8NJ

The Institute is limited by guarantee
Registered in England and Wales
Registration no. 661818
Charity no. 228248

Contact us

developmentprogress@odi.org.uk
T: 020 7922 0300

Sign up for our e-newsletter

developmentprogress.org/
sign-our-newsletter

Follow us on Twitter

twitter.com/dev_progress

Project Notes

Outline the project approach and provide background and methodological information.

Disclaimer

The views presented in this paper are those of the author(s) and do not necessarily represent the views of ODI.

© Overseas Development Institute 2014. Readers are encouraged to quote or reproduce material for non-commercial use. For online use, please link to the original resource on the Development Progress website. As copyright holder, ODI requests due acknowledgement and a copy of the publication.

Acknowledgements

The author thanks Koen Decancq for his valuable and detailed peer review, and Laura Camfield, Sarah Coulthard, Claire Melamed, Emma Samman and Liesbet Steer for comments they made in the course of drafting this paper. The Overseas Development Institute's Measuring and Valuing Development workshops (21-22 March 2012 and 22 November 2013) generated useful ideas, and the author would like to thank all participants. The Project Note was funded by the Bill & Melinda Gates Foundation as part of 'Development Progress', a four-year research project which aims to better understand, measure and communicate what has worked in development and why. The findings and conclusions contained within are those of the author and do not necessarily reflect the positions or policies of the Bill & Melinda Gates Foundation.