

Chapter 1

Concept and validity

The main focus of this chapter is to set the conceptual scope for the measurement of subjective well-being and to provide an overview of what is currently known about the statistical quality of subjective well-being measures. The chapter covers what is meant by subjective well-being, its relevance and why it should be measured, and reviews the evidence on the validity of different types of subjective well-being measure.

In the first section of the chapter a conceptual framework is proposed for measures of subjective well-being, which outlines the main elements of subjective well-being and how these relate to each other. This is necessary both because a clear view of what is being measured is logically prior to decisions about how to measure them and because subjective well-being covers a number of distinct concepts; it is therefore important to be clear about what exactly is covered by the guidelines.

The remainder of the chapter focuses on issues of statistical quality, particularly the *relevance* and *accuracy* of measures of subjective well-being. Relevance addresses the issue of why measures of subjective well-being are important and how they can be used. Accuracy, on the other hand, is concerned with the degree of validity of measures of subjective well-being. In particular, the chapter considers the validity of measures of subjective well-being, focusing on the notion of “fitness for purpose” with respect to specific user needs. There are a number of issues where the evidence on measures of subjective well-being is insufficient to form a clear view of fitness for purpose. For this reason, the final section of the chapter concludes by summarising known issues with data on subjective well-being, including problems with measures of subjective well-being and areas where further research is needed.

1. Conceptual framework

In measurement, it is important to be clear about the nature and scope of the concept being measured. This is particularly the case for a topic such as subjective well-being where the precise concept being measured is less immediately obvious than is the case for a more straight-forward concept such as income, consumption, age or gender. The validity of a statistical measure – as will be discussed in the following sections of this chapter – can be understood as the degree to which the statistical measure in question captures the underlying concept that it is intended to measure. A clear conceptual framework for subjective well-being is therefore essential before it is possible to discuss validity in any meaningful sense.

The first element of a conceptual framework for the measurement of subjective well-being is to define exactly what is meant by subjective well-being. This is important because there are potentially a wide range of subjective phenomena on which people could report, not all of which would necessarily fall under the heading of “well-being”. It is also important to define subjective well-being in order to be able to communicate clearly what is being measured. Often, the measurement of subjective well-being is conflated with measuring “happiness”; however, this is both technically incorrect (there is more to

subjective well-being than happiness) and misleading, and thus lends support to sceptics who characterise the measurement of subjective well-being in general as little more than “happiology”.¹

Most experts characterise subjective well-being as covering a number of different aspects of a person’s subjective state (Diener et al., 1999; Kahneman, Diener and Schwarz, 1999). However, there is room for some debate about exactly what elements should be included. For example, some analysts, such as Kahneman and Krueger (2006), focus primarily on the hedonic aspect of subjective experience, while others, such as Huppert et al. (2009), opt for a definition that includes measures of good psychological functioning as well as purpose in life. For the purposes of these guidelines, a relatively broad definition of subjective well-being is used. In particular, subjective well-being is taken to be:²

Good mental states, including all of the various evaluations, positive and negative, that people make of their lives, and the affective reactions of people to their experiences.

This definition is intended to be inclusive in nature, encompassing the full range of different aspects of subjective well-being commonly identified. In particular, the reference to good mental functioning should be considered as including concepts such as interest, engagement and meaning, as well as satisfaction and affective states. Thus, in the terms of Diener (2006), “subjective well-being is an umbrella term for the different valuations people make regarding their lives, the events happening to them, their bodies and minds, and the circumstances in which they live”. Such valuations are subjective, in that they are experienced internally (i.e. they are not assessments of some external phenomenon); they constitute aspects of well-being in that they relate to the pleasantness and desirability or otherwise of particular states and aspects of people’s lives.

While the definition of subjective well-being used here is broad and potentially reflects the influence of a wide range of people’s attributes and circumstances, it does not imply that subjective well-being is proposed as the single all-encompassing measure of people’s well-being, with all other aspects having only instrumental value in achieving this. On the contrary, this definition is explicitly consistent with approaches that conceive of people’s well-being as a collection of different aspects, each of them having intrinsic value. In measuring overall human well-being then, subjective well-being should be placed alongside measures of non-subjective outcomes, such as income, health, knowledge and skills, safety, environmental quality and social connections.

The definition of subjective well-being outlined above is relatively broad, and could give the impression that subjective well-being is a hopelessly vague concept. This is not the case. There is, in fact, general agreement among experts on the specific aspects that comprise subjective well-being (Dolan and White, 2007; Sen, Stiglitz and Fitoussi, 2009; ONS, 2011). In particular, a distinction is commonly made between life evaluations, which involve a cognitive evaluation of the respondent’s life as a whole (or aspects of it), and measures of affect, which capture the feelings experienced by the respondent at a particular point in time (Diener, 1984; Kahneman et al., 1999). In addition to the distinction between evaluation and affect, a number of researchers argue that there is also a clear eudaimonic aspect of subjective well-being, reflecting people’s sense of purpose and engagement (Huppert et al., 2009). The framework used here covers all three concepts of well-being:

- Life evaluation.
- Affect.
- Eudaimonia (psychological “flourishing”).

Elements of subjective well-being

Life evaluation

Life evaluations capture a reflective assessment on a person's life or some specific aspect of it. This can be an assessment of "life as a whole" or something more focused. Such assessments are the result of a judgement by the individual rather than the description of an emotional state. Pavot and Diener et al. (1991) describe the process of making an evaluation of this sort as involving the individual constructing a "standard" that they perceive as appropriate for themselves, and then comparing the circumstances of their life to that standard. This provides a useful way to understand the concept of life evaluation, although in practice it is not clear whether the process of comparison is a conscious one if respondents more commonly use a heuristic to reach a decision.

There is evidence that the construct captured by life evaluation is closely related to that used by people when they make a conscious judgement that one course of action is preferable to another (Kahneman et al., 1999; Helliwell and Barrington-Leigh, 2010). The underlying concept being measured is thus, in some senses, relatively close to an economist's definition of utility. However, economists usually assume (at least implicitly) that the remembered utility on which people base their decisions is equivalent to the sum of momentary utilities associated with moment-by-moment experiences. This is not the case. Life evaluations are based on how people remember their experiences (Kahneman et al., 1999) and can differ significantly from how they actually experienced things at the time. In particular, the so-called "peak-end rule" states that a person's evaluation of an event is based largely on the most intense (peak) emotion experienced during the event and by the last (end) emotion experienced, rather than the average or integral of emotional experiences over time. It is for this reason that life evaluations are sometimes characterised as measures of "decision utility" in contrast to "experienced utility" (Kahneman and Krueger, 2006).³ Despite this limitation, the fact that life evaluations capture the same sort of construct that people use when making conscious decisions and align closely to the conception of individual welfare that is grounded in the conventional economic paradigm makes them of high interest to researchers and policy-makers.

The most commonly used measures of life evaluation refer to "life as a whole" or some similar over-arching construct. However, in addition to global judgements of life as a whole, it is also possible for people to provide evaluations of particular aspects of their lives such as their health or their job. In fact, there is good evidence that a strong relationship exists between overall life evaluations and evaluations of particular aspects of life. One of the most well documented measures of life evaluation – the *Personal Wellbeing Index* – consists of eight questions, covering satisfactions with eight different aspects of life, which are summed using equal weights to calculate an overall index (International Wellbeing Group, 2006). Similarly, Van Praag, Frijters and Ferrer-i-Carbonell (2003) use panel data from the German Socio-Economic Panel to estimate overall life satisfaction as a function of satisfaction with six specific life domains (job satisfaction, financial satisfaction, house satisfaction, health satisfaction, leisure satisfaction and environmental satisfaction), while controlling for the effect of individual personality. These approaches are important because they establish that evaluations of specific aspects of life have a meaningful relationship with overall life evaluations; this therefore suggests that the scope of life evaluations covered in these guidelines needs to encompass specific as well as general measures.

Affect

Affect is the term psychologists use to describe a person's feelings. Measures of affect can be thought of as measures of particular feelings or emotional states, and they are typically measured with reference to a particular point in time. Such measures capture how people experience life rather than how they remember it (Kahneman and Krueger, 2006). While an overall evaluation of life can be captured in a single measure, affect has at least two distinct hedonic dimensions: positive affect and negative affect (Kahneman et al., 1999; Diener et al., 1999). Positive affect captures positive emotions such as the experience of happiness, joy and contentment. Negative affect, on the other hand, comprises the experience of unpleasant emotional states such as sadness, anger, fear and anxiety. While positive affect is thought to be largely uni-dimensional (in that positive emotions are strongly correlated with each other and therefore can be represented on a single axis of measurement), negative affect may be more multi-dimensional. For example, it is possible to imagine at one given moment feeling anger but not fear or sadness.

The multi-dimensional nature of affect raises an interesting question about the relationship of affective states to life evaluation. Life evaluations are uni-dimensional in that different experiences can be rated unambiguously as better or worse. Kahneman et al. (1999), argues for the existence of a "good/bad" axis on which people are able to place experiences based on the emotional states they are experiencing. In effect, he argues, people are able to make an overall judgement about the net impact of their affective state at a particular point in time. In principle, this is the same process that is involved in forming life evaluations from remembered affective states. Kahneman's point is that affective states can be compared, and that one can therefore reasonably aggregate measures of current affect. For this reason, affect measures are sometimes reported in terms of affect balance, which captures the net balance between positive and negative affect (Kahneman and Krueger, 2006).

The measurement of affect poses different challenges to the measurement of life evaluation. It is difficult to ask people to recall affective states in the past, since responses will be affected by recall biases such as the peak/end rule mentioned above. The gold standard for measuring affect is the experience sampling method (ESM), where participants are prompted to record their feelings and perhaps the activity they are undertaking at either random or fixed time points, usually several times a day, throughout the study period, which can last several weeks. To maximise response rates and ensure compliance throughout the day, electronic diaries are often used to record the time of response. While the ESM produces an accurate record of affect, it is also expensive to implement and intrusive for respondents.

A more viable approach is the use of the day reconstruction method (DRM), in which respondents are questioned about events from a time-use diary recorded on the previous day. Research has shown that the DRM produces results comparable with ESM, but with a much lower respondent burden (Kahneman et al., 2004). Experience Sampling, the DRM and similar methods for collecting affect data in time-use studies allow for analysis that associates particular affective states with specific activities. It is also possible to collect affect data in general household surveys.⁴ However, affect measures collected in general household surveys lose some detail due to the need to recall affect (even if only what affective states the respondent experienced on the previous day) and also cannot easily capture information linking affect to particular activities.

Eudaimonia

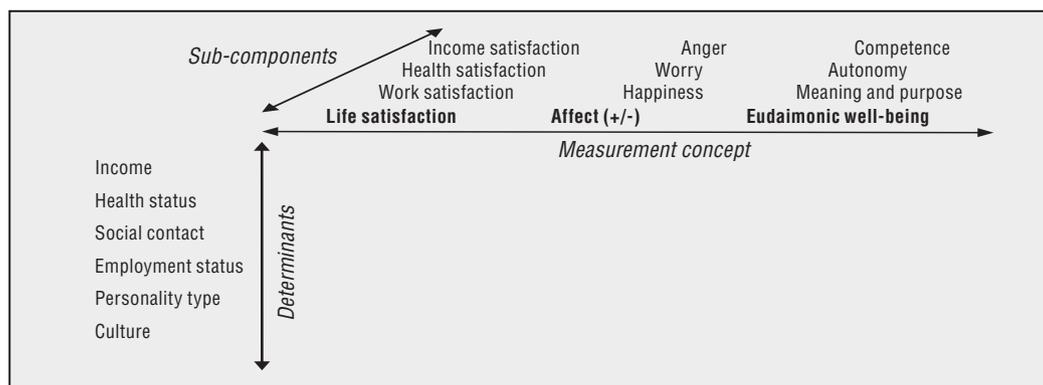
In addition to life evaluations and affect, which focus on a person's experiences (current or recalled), some definitions of subjective well-being found in the psychological literature include other aspects of a person's psychological processes as well. In particular, there is a substantial literature focused on the concept of good psychological functioning, sometimes also referred to as "flourishing" or "eudaimonic" well-being (Huppert et al., 2009; NEF, 2009; Clark and Senik, 2011; Deci and Ryan, 2006). Eudaimonic well-being goes beyond the respondent's reflective evaluation and emotional states to focus on functioning and the realisation of the person's potential. In developing the questionnaire on psychological well-being for the European Social Survey, for example, Huppert et al. (2009) characterise the "functioning" element of well-being as comprising autonomy, competence, interest in learning, goal orientation, sense of purpose, resilience, social engagement, caring and altruism. Eudaimonic conceptions of subjective well-being thus differ significantly from the evaluative and affective components in that they are concerned with capabilities as much as with final outcomes and thus have a more instrumental focus. Because measuring eudaimonia draws on both psychological and humanist literature, which identifies key universal "needs" or "goals", the approach represents a useful response to the criticism that the measurement of subjective well-being is "happiology", or built purely on hedonistic philosophy, and also aligns itself with many people's perceptions of what it is important to value in life.

While there is now a general consensus on the distinction between life evaluations and affect, the conceptual structure of eudaimonic well-being is less well fleshed out. It is not clear, for example, whether eudaimonic well-being describes a uni-dimensional concept in the sense of life evaluation, or whether the term is used to cover a range of different concepts. It is, however, clear that eudaimonic measures of well-being capture important aspects of people's subjective perceptions about their own well-being that are not covered by either life evaluations or affect. For example, having children has a negligible (or even mildly negative) correlation with average levels of life evaluation (Dolan, Peasgood and White, 2008), while child care (even of one's own children) is associated with relatively low levels of positive affect (Kahneman et al., 2004). This conflicts with the intuitive assumption that children, at least for those who choose to have them, contribute in some way to their parent's well-being. Indeed, people with children report much higher average levels of meaning or purpose in their lives than other respondents (Thompson and Marks, 2008).

The relationship between life evaluation, affect and eudaimonia

Life evaluation, positive and negative affect and eudaimonic well-being are all conceptually distinct. However, it is helpful to have a conceptual model of how they might relate to each other. Figure 1.1 provides a simple model of the different elements of a measurement framework for subjective well-being. The model emphasises three dimensions involved in the measurement of subjective well-being. These are: 1) the measurement concept; 2) the sub-components of well-being; and 3) determinants. It should be noted that the list of determinants and sub-components in Figure 1.1 is illustrative rather than exhaustive. The purpose of the conceptual model presented in Figure 1.1 is not to provide a comprehensive framework covering all possible elements of subjective well-being. Rather, it is intended to serve as an organising framework for thinking about the scope of the topics covered by these guidelines. This is necessarily focused on a narrower range of measures than might be found in an academic survey of

Figure 1.1. A simple model of subjective well-being



human well-being, and reflects the topics most likely to be of relevance for official statistics and policy-making. There is also a significant bias towards those concepts that underpin the measures traditionally used in large-scale surveys.

Figure 1.1 outlines the various elements of a simple measurement framework for subjective well-being. However, it is also useful to briefly review the empirical relationship between the three types of measures. There is extensive evidence on the relationship between measures of affect and overall measures of life evaluation. Diener, Kahneman, Tov and Arora (in Diener, Helliwell and Kahneman, 2010) show that there is a high correlation (0.82) across countries between the most commonly used average measures of life evaluation, but a much lower correlation (0.55-0.62) between average affect balance and either of two life evaluation measures (life satisfaction and the Cantril Ladder). Similarly, at the individual level, Kahneman and Krueger (2006) report only a moderate correlation (0.38) between life satisfaction (an evaluative measure) and net affect.

There is also a body of evidence on the empirical relationship between eudaimonic well-being and the other aspects of subjective well-being, which suggests that this correlation is smaller than in the case of the relationship between affect and life evaluations. Clarke and Senik (2011), for example, report a correlation between life satisfaction and four different aspects of eudaimonic well-being of between 0.25 and 0.29. Diener et al. (2009) report a correlation of 0.62 ($N = 563$, $p < 0.001$) between their Psychological Well-Being Scale and the evaluative Satisfaction with Life Scale, and correlations of 0.62 and 0.51 respectively between the Psychological Well-Being Scale and the positive and negative subscales of the Scale of Positive and Negative Experience ($N = 563$, $p < 0.001$ in all cases). Huppert and So (2009) found a correlation of 0.32 between flourishing and life satisfaction in European Social Survey data. Among the European Social Survey sample overall, 12.2% met the criteria for flourishing, and 17.7% met the criteria for high life satisfaction, but the percentage for both flourishing and high life satisfaction was 7.2%.

Table 1.1 gives the correlations between individual measures of life evaluation derived from the Gallup World Poll (life satisfaction), positive affect, negative affect and eudaimonic well-being (purpose) across 362 000 respondents in 34 OECD countries. The correlation is highest between the two measures of affect, at -0.3855, and lowest between purpose and negative affect, at -0.091. Life satisfaction has a correlation of about 0.23 with both measures of affect, and of 0.13 with purpose. While all the coefficients in Table 1.1 show the expected sign and all are significant at the 0.1% level, none of the measures have a correlation near 1, indicating that the different measures capture different underlying phenomena.

Table 1.1. **Correlation coefficients among purpose, life satisfaction, positive affect and negative affect at the individual level, 2006-10**

	Purpose	Life satisfaction	Positive affect	Negative affect
Purpose	1.000			
Life satisfaction	0.134	1.000		
Positive affect	0.142	0.229	1.000	
Negative affect	-0.091	-0.231	-0.3855	1.000

Note: The precise measures used are the so-called “Cantril Ladder” for life satisfaction, an “important purpose” in life for purpose, and the sum of “yes” responses to smiled yesterday, experienced joy yesterday, and was well rested yesterday for positive affect and an equivalent index based on experience of sadness, worry and depression for negative affect.

Source: Gallup World Poll.

2. The quality of subjective well-being measures

Quality is crucial to any statistical measure. Unless data captures the concept being measured with a sufficient degree of accuracy to draw reasonable inferences from it, there is little point in collecting it. This is particularly true for official statistics, which are expected to be of the highest quality. As the *United Nations Fundamental Principles of Official Statistics* states, “official statistics provide an indispensable element in the information system of a society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation” (OECD, 2008). It is therefore essential that decisions about the measurement of subjective well-being through official statistics are solidly grounded in a clear understanding of the reliability and validity of such measures.

The *Quality Framework and Guidelines for OECD Statistical Activities* (OECD, 2008) sets out the OECD’s approach to dealing with issues of statistical quality. Under the *Framework*, quality is defined as “fitness for use” in terms of user-needs. The ultimate benchmark as to the quality of statistics is essentially whether they meet the needs of the user in terms of providing useful information. Because users must often make decisions about a course of action whether or not statistical information is available, a focus on “fitness for purpose” may involve accepting the use of data that is less than perfectly accurate provided that the data is of sufficient quality that it improves rather than detracts from the quality of decision-making.

Evaluating a concept as broad as “fitness for purpose” is challenging. For this reason, the *Framework* identifies seven dimensions of statistical quality. These seven dimensions define the characteristics of high-quality data and provide a structured way of assessing the quality of a particular set of statistics. The seven dimensions of quality are:

- *Relevance*, i.e. the degree to which data serves to address the purposes for which they are sought by users.
- *Accuracy*, i.e. the degree to which data correctly estimate or describe the quantities or characteristics they are designed to measure.
- *Credibility*, i.e. the confidence that users place in statistics based on their image of the data producer.
- *Timeliness*, i.e. the length of time between the availability of data and the phenomenon or event that the data describe.
- *Accessibility*, i.e. how readily data can be located and retrieved by users.

- *Interpretability*, i.e. the ease with which the user can understand and properly use and analyse the data.
- *Coherence*, i.e. the degree to which the data is mutually consistent with other similar measures and logically integrated into a system of statistics.

These seven criteria, along with the more general principle of cost effectiveness in producing/collecting such data, provide the OECD's overall framework for assessing statistical quality. However, most of these criteria relate to how statistics are measured and collected rather than what is collected. For the purposes of these guidelines, the concern is more narrowly focused on what should be collected rather than the more general principles of how an official statistical agency should operate. Thus, the main focus for assessing the quality of measures of subjective well-being will be the principles of relevance, accuracy and, to a lesser degree, coherence.

3. The relevance of measures of subjective well-being: Why are they important?

It is important to be clear about why subjective well-being should be measured. Official statistics are produced to meet the needs of policy-makers in planning and assessing the impact of policy decisions, and to inform the general public about the state of society. Academics and the media are also important users of official statistics, contributing to a better understanding of society and informing the public and decision-makers. The demand for official statistics is thus, ultimately, a derived demand; statistics are collected because they are of use to someone, rather than for their own sake.

The principles of official statistics generally reflect the view that information is collected only when there is good reason and for a clear purpose. The OECD framework for data quality identifies relevance as the first of the seven key dimensions of quality. Relevance implies that the value of data "is characterised by the degree to which that data serves to address the purposes for which they are sought by users" (OECD, 2003). Similarly, the *United Nations Fundamental Principles of Official Statistics* asserts that the role of official statistical agencies is to compile and make available "official statistics that meet the test of practical utility... to honour citizens' entitlement to public information".

There are sound ethical and practical reasons why official statistical agencies insist on having a clear understanding of the uses of any proposed statistical measures. Many official statistical agencies have the power to compel responses from respondents. That is, respondents are legally required to provide information when approached by a national statistical agency. The corollary of such authority is the requirement for national statistical offices to use data responsibly. From an ethical standpoint, only information that is sufficiently important to justify the intrusion into respondents' lives should be collected. The International Statistical Institute's *Guidelines on Professional Ethics* notes that:

Statisticians should be aware of the intrusive potential of some of their work. They have no special entitlement to study all phenomena.

Over and above this ethical concern is also a practical concern. Even if compliance is legally mandated, the quality of compliance depends heavily on preserving a good relationship between respondents and the official statistical agency. This, in turn, is undermined if the statistical agency cannot articulate why the data being collected is important and how it will be used.

Official statistical agencies are also under increasing resource pressures. This takes the form of both budget cuts, which preclude collecting all the information for which there is a potential demand, and issues of response burden. Even where funding exists to collect information, official statistical agencies must be careful not to over-burden respondents and jeopardise the good will on which high-quality responses depend. Because of this, collecting measures of subjective well-being will have an opportunity cost in terms of other data that will not be collected in order to produce such measures. If subjective well-being measures are to be included in official statistics, therefore, it is essential to be clear about how they will be used.

It is also important to be clear about how subjective well-being measures will be used for purely technical reasons. The field of subjective well-being covers a wide range of different concepts and measures. Choosing which measures should be the focus of collection efforts requires knowing what the measures will be used for. Different measures of subjective well-being will be better suited to different purposes, and it is therefore important that these guidelines identify the right measures needed given the core policy- and public-uses for the data.

The intended use for measures of subjective well-being also affects judgements about the validity of such measures. No statistical measure captures the concept it is intended to measure perfectly. Whether any particular measure can be considered valid, therefore, ultimately involves a judgement about whether the quality of the measure is sufficient to support its intended use. A measure that is valid for one purpose may not be valid for other purposes. For example, a measure could provide valid information about the distribution of outcomes within a country, but be subject to significant bias due to cultural or linguistic factors. While this would be a significant limitation if the intended use of the data is to rank countries compared to each other, it is less important for purely domestic uses.

Measures of subjective well-being have a wide variety of potential uses and audiences. For the purposes of these guidelines it is useful to classify the possible uses of subjective well-being measures under a general framework. The following framework identifies four main ways in which measures of subjective well-being are used. In particular, they can:

- *Complement other outcome measures.*
- *Help better understand the drivers of subjective well-being.*
- *Support policy evaluation and cost-benefit analysis, particularly where non-market outcomes are involved.*
- *Help in identifying potential policy problems.*

Complement other outcome measures

Measures of subjective well-being provide an alternative yardstick of progress that is firmly grounded in people's experiences. These subjective measures may differ in important respects from the picture provided by more conventional metrics that focus on access to resources. This is desirable, since if measures of subjective well-being duplicated the picture provided by other social and economic indicators, there would be few additional gains in using them.⁵ In particular, being grounded in peoples' experiences and judgements on multiple aspects of their life, measures of subjective well-being are uniquely placed to provide information on the net impact of changes in social and economic conditions on the perceived well-being of respondents, taking into account the

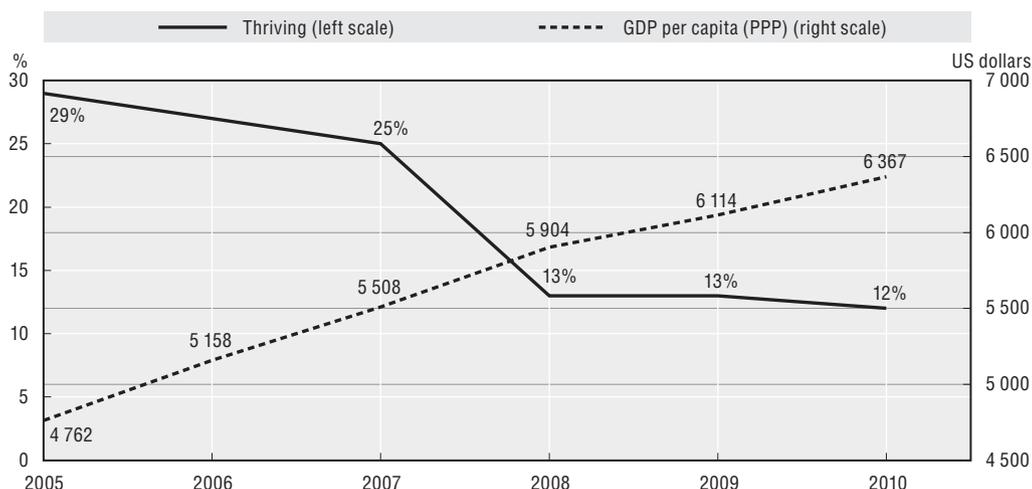
Box 1.1. Subjective well-being, GDP growth and the “Arab Spring”

For policy-makers, measures of subjective well-being are valuable as an indicator of progress when they can alert them to issues that other social and economic indicators might fail to identify. One recent example where measures of subjective well-being clearly demonstrate their ability to capture important elements of well-being not captured by more traditional measures is the decline in country-average measures of subjective well-being that occurred in Egypt and Tunisia in the years leading up to 2011, a decline that contrasts with the much more favourable evolution of GDP data. For example, Tunisian real GDP per capita increased from USD 8 891 in 2008 to USD 9 489 in 2010, i.e. a real gain of around 7%. However, the proportion of the population indicating a high level of satisfaction with their life as a whole fell from 24% to 14% over the same period (Gallup, 2011). Egypt (shown in the picture below) shows a similar pattern from 2005 to 2010, with a real gain in GDP per capita of around 34% and a decline in the share of respondents classified as “thriving” by almost half.* This illustrates how subjective perceptions can provide information on very significant outcomes in societies that other conventional indicators such as GDP growth do not provide.

* “Thriving” is a composite measure of subjective well-being calculated by the Gallup Organisation. It is based on answers to the Cantril ladder measure of life satisfaction for life at the moment and how people expect life to be in five years.

Figure 1.2. Trends in subjective well-being and GDP in Egypt: 2005-10

Recent trends in percentage “thriving” and GDP per capita (PPP)



Source: Subjective well-being data are from Gallup. GDP per capita (PPP) estimates are from the *International Monetary Fund's World Economic Outlook Database*.

impact of differences in tastes and preferences among individuals. An example of how these measures can change perceptions about progress in individual countries is provided by Box 1.1, in respect of the “Arab Spring”.

In addition to providing information on aggregate changes at the national level, measures of subjective well-being can also provide a picture of which groups in society are most (dis)satisfied or experience the best or worse life. Again, because measures of subjective well-being capture the impact of taste and aspirations as well as the distribution of other life circumstances, such measures provide useful additional information for policy-makers in situations where comparisons are made across sub-groups of the

population. Migrants, for example, may be more motivated than the rest of the population by income relative to other factors (Bartram, 2010), as this is a primary motive for their decision to move abroad. An attempt to assess migrant well-being compared to the rest of the population is therefore challenging, given that there is good reason to believe that there will be systematic differences in the importance that the two groups attach to different aspects of quality of life. Because measures of subjective well-being incorporate the impact of the different weights that various people may attach to the different aspects of their quality of life, they have the potential to add an important dimension to analysis in situations involving comparisons between population groups.

The final policy use of measures of subjective well-being in the context of measuring progress is for cross-country comparisons of aggregate measures of subjective well-being, such as those included in *How's Life?* (OECD, 2011). Due to the impossibility of performing controlled experiments across countries, cross-country comparisons of subjective well-being outcomes are one way to learn about the strengths and weaknesses of different policies. Because measures of subjective well-being are sensitive to a different range of drivers than are other social and economic indicators, they can provide additional information about the consequences of a particular policy. A crucial issue in using measures of subjective well-being in this way, however, is the degree to which cross-cultural comparisons of measures of subjective well-being are valid. This issue is considered in more depth later in this chapter.

The interest of the general public and the media in using measures of subjective well-being as complements of measures of progress is generally similar to that of policy-makers. For these users, the key contributions that subjective well-being measures can potentially make are in highlighting how different groups fare compared to each other, what can be learned from the experiences of other countries, and perhaps whether things are getting better or worse overall – all of which are of potential interest to the general public and the media.

Better understand the drivers of subjective well-being

The second major use of subjective well-being measures is to contribute to a better understanding of the drivers of well-being at an individual level. If it is accepted that measures of subjective well-being are valid, and that they accurately capture the concepts that they claim to measure – an overall evaluation of life or the experienced moods and emotions of an individual over a period of time – then it follows that such measures can be used to provide information on the relative contribution of different factors and circumstances to a person's well-being – albeit with some noise due to both measurement error and the fact that a person's subjective perception of their well-being is not necessarily quite the same thing as their overall well-being (for examples see Dolan, Peasgood and White, 2008; Helliwell and Wang, 2011; Boarini, Comola, Smith, Manchin and De Keulenaer, 2012).

Measures of subjective well-being can be used to help identify what factors are critical aspects of people's well-being. In particular, such measures can be used to test intuitions about what factors matter most to people. This is potentially important to the broad agenda of measuring progress, since it provides an empirical way to test whether the outcomes used to measure progress align well with the factors that determine people's perceptions of their well-being. Although people's subjective perceptions are not necessarily equivalent to overall well-being for a number of reasons,⁶ measures of subjective well-being are unique in that they provide a relatively robust empirical source of

information on such preferences, especially when non-market outcomes are involved. Without using subjective views of what matters the most to people, we would be left to essentially *a priori* judgements and anecdotal focus group research.

Subjective well-being measures are, however, unique in that they provide a relatively robust empirical source of information on what affects how people feel about their lives, which is an important component of overall well-being. By examining the level of subjective well-being *actually achieved* as a result of different decisions or approaches, policy-makers and individuals can better understand what matters to people on an empirical (rather than anecdotal) level. For example, subjective measures can be used to test more specific hypotheses about what aspects of policy are most important to people. Halpern (2010), for example, refers to an instance where the Merseyside police, in the United Kingdom, used data on how satisfied members of the public were with the service provided by the local police, alongside more traditional performance measures on crimes committed and offence resolutions. In contrast to the expected hypothesis – which was that minimising the response time from the police was of crucial importance for public satisfaction – the evidence showed that it was much more important that police arrived when they said they would. For minor issues not involving safety, what mattered was punctuality rather than speed.

Going beyond just identifying what matters for well-being, measures of subjective well-being can assist in developing a better understanding of the trade-offs between different outcomes. Many policy problems require taking a decision about how to compare two fundamentally different types of outcome (see Box 1.2). Dolan and White (2010) note that these types of issue characterise many attempts to encourage “joined-up government”, where there is a need for different government agencies to consider the costs and benefits of a particular intervention not just on their primary outcome of concern, but also in terms of how these affect the outcomes of other government agencies.

Because measures of subjective well-being can capture the combined effect of all different changes in life circumstances on an individual’s perception of their well-being in a single measure, they can be used as a common metric for assessing the relative impact of fundamentally different outcomes. For example, Ferrer-i-Carbonell and Frijters (2004) use measures of overall satisfaction with life and satisfactions with specific outcome domains to assess the relative weights to attach to different outcome areas. Comparing the magnitude of the impact of health satisfaction on overall life satisfaction with the impact associated with housing satisfaction gives a way of quantifying the relative importance of each dimension within a particular sample, given where they started on each measure.⁷ Similarly, Di Tella, Oswald and Macculloch (2003) used the coefficients from a regression on life satisfaction to investigate the inflation/unemployment trade-off. While the so-called “misery index” weights the unemployment rate and inflation rate equally as indicators of the negative impact of macroeconomic outcomes, Oswald and Macculloch’s analysis suggests that the impact of unemployment on subjective well-being is significantly greater than that of inflation.

Policy evaluation and cost-benefit analysis

The third main use of subjective well-being measures is to assist in the evaluation of policies. This includes both the direct use of measures of subjective well-being in formal policy evaluations as well as the more indirect – but possibly more important – role that they can play in cost-benefit analysis.

Box 1.2. Using measures of subjective well-being to value life events

People intuitively compare different life events on a daily basis and make judgements about how bad or good things might be. However, trying to put a number on the relative magnitude of the impact of different life events such as marriage or divorce, on a person's well-being – much less a monetary value – might seem counter-intuitive to many people. Nonetheless, such values are of potentially high interest from the perspective of thinking about how much to invest in preventing or encouraging a particular outcome.

Measures of subjective well-being provide a relatively straight-forward way of comparing the relative impact of fundamentally different life events in a quantitative way and, based on this, assigning such events a monetary value. Clark and Oswald (2002) present a method for valuing life events and, although the literature on using measures of subjective well-being to value life events has expanded significantly since 2002, the basic methodology remains largely unchanged. Consider the results below from a regression of a range of possible determinants of subjective well-being against life satisfaction (Boarini et al., 2012). The coefficients for the (base two) logarithm of household income, being married, and being unemployed are shown, and express the change in life satisfaction (on a scale of 0 to 10) associated with a doubling of income, being married, or being unemployed, respectively, holding all else constant.

Event	Coefficient
Log household income	0.1482
Married	0.2584
Unemployed	-0.4643

Using these coefficients, it is possible to calculate the relative impact of being married compared to being unemployed on life satisfaction as $0.2584/0.4643 = 0.5565$. Or, put more simply, being unemployed has almost twice the impact on life satisfaction as does being married.

Going beyond this, the monetary value of being married or being unemployed can be calculated by comparing the relevant coefficients with that associated with the coefficient for household income. Using the values presented above, the coefficient on being married is $0.2584/0.1482 = 1.7435$ times larger than the impact of a doubling of household income. For a person with a household income equal to the OECD per capita household disposable income (USD 17 286 at PPP, 2008), this is equivalent to $1.7435 \times \text{USD } 17\,286 = \text{USD } 30\,138$. For unemployment the comparable value is $2.930 \times \text{USD } 17\,286 = \text{USD } 50\,647$.

These values are intended to illustrate the techniques involved, and need to be treated with caution. In particular, better measures would use panel data to capture the causal relationship (as do Clark and Oswald) rather than just correlation, and need to consider any potential biases in the data as well as the structure of the regression equations used to calculate the coefficients (Fujiwara and Campbell, 2011).

In formal policy evaluations, measures of subjective well-being can complement other social and economic indicators as a measure of the outcomes achieved by a policy. Here, as is the case with monitoring the progress of entire communities, measures of subjective well-being can add additional information over and above that captured by more traditional indicators. For some initiatives – where the impact on the subjective

experiences of the population is the main object of the programme – measures of subjective well-being may even be suitable as the primary metric for assessing the programme's success.

Many policy evaluations already include subjective measures of client satisfaction and questions on the respondent's perceptions of what elements of the programme were most valuable. More general measures of overall subjective well-being, however, have some significant advantages over and above these more focused measures. Most importantly, measures of subjective well-being provide information on the actual impact of an initiative on the respondent's subjective well-being, rather than the impact that the respondent consciously identifies. These values can differ because people's judgements about the impact of a programme may be influenced by the fact that they have participated in the programme (i.e. they might be more prone to assign the cause of any recent changes in their well-being to the programme rather than to other factors, knowing that this is what he/she is being asked about). Also, people may not be aware of all of the various feedback loops via which a policy programme affects them. For example, in evaluating an active employment programme, respondents might consider the direct effect on their well-being of both having a job and gaining additional income, but not the flow on well-being that would stem from changes in their time-use due to longer commuting. Because measures of subjective well-being can capture the overall impact of a change on life circumstances, without requiring a cognitive judgement by the respondent on which causal pathways are being asked about, such measures provide useful additional information on the overall impact of a programme.

In some cases, measures of subjective well-being can be better than conventional cost-benefit analysis at treating non-monetary outcomes. Examining the relative costs and benefits of a proposal is relatively straight-forward when the proposal is aimed at strictly economic outcomes, and the costs and benefits of the proposal can be obtained from the relevant market prices. However, where the aim of a proposal is to achieve outcomes that do not have an obvious market price, it is much more challenging to obtain meaningful values for analysing the relevant costs and benefits. Because much government policy is concerned with market failures, many government policies are correspondingly focused on achieving non-market outcomes.

The traditional economic approaches to cost-benefit analysis for non-market outcomes depend on either revealed preference or contingent valuation techniques to estimate "prices" for such outcomes. A revealed preference approach involves calculating values based on the shadow prices implied by observed behaviour, while contingent valuation techniques calculate values based on the "willingness to pay" for the outcome in question, as expressed by respondents to a hypothetical question in a survey. Clarke and Oswald (2002) note that measures of subjective well-being can provide the framework for such valuations by comparing the impact of a particular outcome on subjective well-being with the impact of a change in income on subjective well-being. By making such a comparison, it is possible to calculate the amount of money required to achieve the same increase or decrease in well-being as that caused by the outcome under assessment.

There is good reason to believe that, in several circumstances, measures of subjective well-being have advantages over both revealed preference and contingent valuation for the purposes of cost-benefit analysis (see Box 1.3). An obvious advantage is that many measures of subjective well-being – such as overall life satisfaction – are relatively easy and

Box 1.3. *The Green Book and life satisfaction*

The Green Book is the formal guidance from the Treasury of the United Kingdom to other UK government agencies on how to appraise and evaluate policy proposals. The current edition of *The Green Book* dates to 2003, and provides advice on how officials should provide justification for a proposed government intervention, set objectives for the proposal, appraise the various options, and evaluate the effectiveness of the final action that results. In July 2011, *The Green Book* was updated to reflect the results of a review of valuation techniques for social cost-benefit analysis jointly commissioned by the Treasury and the Department for Work and Pensions (Fujiwara and Campbell, 2011). The review specifically focuses on the contribution that can be played by measures of subjective well-being – particularly life satisfaction – alongside more traditional approaches to cost-benefit analysis. In summarising the conclusions of the review, *The Green Book* states (p. 58):

A newer, “subjective well-being approach” has been gaining currency in recent years. The “life satisfaction approach” looks at people’s reported life satisfaction in surveys such as the ONS’s Integrated Household Survey, which began including questions on respondents’ subjective well-being in April 2011. The life satisfaction approach uses econometrics to estimate the life satisfaction provided by certain non-market goods, and converts this into a monetary figure by combining it with an estimate of the effect of income on life satisfaction.

At the moment, subjective well-being measurement remains an evolving methodology and existing valuations are not sufficiently accepted as robust enough for direct use in Social Cost-benefit Analysis. The technique is under development, however, and may soon be developed to the point where it can provide a reliable and accepted complement to the market based approaches outlined above. In the meantime, the technique will be important in ensuring that the full range of impacts of proposed policies are considered, and may provide added information about the relative value of non-market goods compared with each other, if not yet with market goods.

While the amendment to *The Green Book* stops short of fully endorsing the use of life satisfaction measures for use in formally evaluating government programmes, the decision to make an interim amendment in itself signals strongly the importance that UK central agencies attach to obtaining improved measures of the value of non-market outcomes.

cheap to collect. However, there are also more substantive methodological advantages that may be associated with using measures of subjective well-being in this way. Revealed preference relies on strong assumptions about people’s ability to know how an outcome will affect them in the future, and on the assumptions that markets are in equilibrium. Diener, Lucas, Schimmack and Helliwell (2009) note that for market prices for houses to reflect the disutility of airport noise accurately would require that house purchasers are able to forecast how much the noise will impact them before buying the house. Similarly, in this example, it is difficult to disentangle the differences in house prices due to noise from differences in other aspects of house quality.

Contingent valuation also relies strongly on people’s ability to make accurate judgements about how something will make them feel in the future. Dolan and Peasgood (2006) note that people have difficulty imagining how good or bad different circumstances are actually going to be. Indeed, the “willingness to pay” surveys commonly used for contingent valuation are, to a large degree, measures of the subjective well-being associated with a hypothetical scenario. Using measures of subjective well-being to calculate the costs based on the actual impact of different life circumstances on subjective

well-being removes the hypothetical element from the equation. In addition, contingent valuation surveys tend to produce very different estimates of the value of outcomes for people at different points on the income distribution. This tends to result in either weighing the desires of the rich more heavily than the poor when assessing the costs and benefits associated with the proposal under consideration or taking account of the marginal utility of income in calculating the final cost. The latter approach is difficult in the absence of robust estimates of the marginal utility of income (Dolan and White, 2007).

Identifying potential policy problems

An important feature of measures of subjective well-being is their ability to provide an insight into human behaviour and decision-making. In particular, measures of subjective well-being can help researchers to understand the difference between the *ex ante* beliefs that people hold about their future well-being (which form the basis for decisions) and the *ex post* outcomes that people achieve in terms of their subjective well-being. A better understanding of these issues is important both for policy-makers and for the broader public. Policy-makers have an interest in understanding why people make the decisions that they do, because much public policy involves dealing with the consequences of systematic poor decision-making by individuals. Similarly, businesses and the general public have an interest in understanding how people's subjective well-being shapes their behaviour.

One way in which measures of subjective well-being are useful to businesses and the broader public is by providing information on the characteristics of good places to live and work. There is clear evidence that subjective well-being predicts future behaviour. Clark (2001), for example, has shown that measured job satisfaction predicts the probability of an employee going on to leave their job. Thus businesses might well have an interest in the measured job satisfaction of their employees and in understanding the determinants of job satisfaction.

Measures of subjective well-being can also help shed light on various biases in the way people make decisions. Although people are generally able to predict whether events are likely to be pleasant or unpleasant, Wilson, Gilbert and colleagues have described various ways in which affective forecasting can be biased or faulty, particularly with regard to the intensity and duration of emotional reactions to future events (e.g. Wilson, Wheatley, Meyers, Gilbert and Axsom, 2000; Wilson and Gilbert, 2006). Kahneman et al. (2006), for example, show that people are prone to over-estimate the impact of income gains on their life satisfaction. When evaluating those factors that people expect to contribute to a positive mood, people tend to focus on conventional achievements, thus over-estimating the role of income relative to other factors. By way of contrast, other activities that are less commonly used as a reference for conventional measures of status get under-estimated with respect to their impact on subjective well-being. Commuting, for example, has been found to have a strong negative impact on both measures of affect (Kahneman et al., 2006) and life evaluations (Frey and Stutzer, 2008). This suggests that people may be prone to over-estimating the positive impact of, for example, a new job with a higher salary but a longer commute.

Faulty affective forecasting is significant in this context because it suggests that decisions reflected in market choices will not always serve to maximise subjective well-being in practice. Individuals may have a substantial interest in better understanding the factors affecting the level of well-being that they actually achieve. Hence, a sound evidence base derived from measures of subjective well-being is of potential interest to the general public. There are also direct policy applications for better understanding the

human decision-making process and the various biases and heuristics involved in it. Consider the case of policy options that incorporate a “default” option, for example, workplace retirement schemes that are set up on a basis of either “opt in” clauses, where a new employee does not join the scheme unless he/she ticks a box to join, or “opt out” clauses, where the reverse is the case. The fact that people respond differently depending on which default is selected – despite the fact that in neither case is there any compulsion – has raised policy interest in the idea of “libertarian paternalism”, which focuses on achieving better outcomes by setting policy defaults to influence people’s behaviour in positive directions. Dolan and White (2007) note that information on subjective well-being can be used to help set policy default options appropriately, by indicating which default options contribute most to subjective well-being.

4. The accuracy of subjective well-being measures

Accuracy is concerned with whether the measure in question accurately describes the qualities of the concept being measured. This, in turn, is usually assessed in terms of reliability and validity. Reliability concerns the extent to which a measure yields consistent results (i.e. whether it has a high signal-to-noise ratio). Validity, on the other hand, is about the extent to which it actually captures the underlying concept that it purports to measure (i.e. whether it is measuring the right thing). Some degree of reliability is a necessary but not sufficient condition for validity.

Reliability

Reliability is a fundamental component of accuracy. For any statistical measure it is desirable that the measure produce the same results when carried out under the same circumstances. This is essential if the measure is to be able to be used to distinguish between changes in the measure due to a genuine change in the condition being measured as opposed to changes that simply represent measurement error. While no statistical measure is completely reliable, it would be of concern if measures of subjective well-being performed significantly worse than other commonly-used measures.

There are two main ways to measure reliability. Internal consistency reliability concerns the extent to which different items on an overall scale or measure agree with one another, and is assessed through examination of inter-item correlations. If the correlation between the two items is high, this suggests that the two measures capture the same underlying concept. On the other hand, if the correlation is low, it is not necessarily the case that both measures are poor but at least one of them must be.

The second approach involves looking at test-retest reliability, where the same question is administered to the same respondent more than once, separated by a fixed period of time. Test-retest reliability places a lower bound on the overall reliability of the measure, but not an upper bound. For example, a low test-retest score could indicate that a measure lacks reliability, but it could also be associated with a high level of actual reliability and a genuine change in the subject of interest.

Both aspects of reliability have been extensively tested for measures of life evaluation and affect over the past twenty years. There is strong evidence for convergence between different life satisfaction measures. In a meta-review of the reliability and validity of subjective well-being measures, Diener (2011) reports a Cronbach’s alpha⁸ for multiple item measures of life satisfaction (including the Satisfaction With Life Scale) of between 0.8

and 0.96. A Cronbach's alpha of 0.7 is typically taken to be the threshold of acceptable convergence, and the scores Diener reports indicate a very high degree of convergence between the different questions used in the life satisfaction scales.

Whilst it is not possible to compute Cronbach's alpha for single item measures, other estimation procedures can be employed. Comparisons across countries using different measures of the same construct generally show slightly lower correlations, but are still relatively high given that the scores not only represent different questions, but are also sourced from different surveys. Bjørnskov (2010), for example, finds a correlation of 0.75 between the average Cantril Ladder measure of life evaluation from the Gallup World Poll and life satisfaction as measured in the World Values Survey for a sample of over 90 countries.

Test-retest results for single item life evaluation measures tend to yield correlations of between 0.5 and 0.7 for time periods of 1 day to 2 weeks (Krueger and Schkade, 2008). Michalos and Kahlke (2010) report that a single-item measure of life satisfaction had a correlation of 0.67 for a one-year period and of 0.65 for a two-year period. In a larger study, Lucas and Donellan (2011) estimated the reliability of life satisfaction measures in four large representative samples with a combined sample size of over 68 000, taking into account specific errors. They found test-retest correlations in the range of 0.68 to 0.74, with a mean of 0.72 over a period of one year between reports.

Multiple item measures of subjective well-being also generally do better than single questions on test-retest reliability. Krueger and Schkade (2008) report test-retest scores in the range of 0.83 to 0.84 for a period of 2 weeks to 1 month between tests, with correlations declining to 0.64 at 2 months and to 0.54 over 4 years. The pattern of decline here is as expected, with longer periods of time showing lower reliability due to a higher likelihood that there has been a genuine change in the respondent's circumstances. In another study, the "satisfaction with life" scale (a multi-item measure of life satisfaction) showed a correlation coefficient of 0.56, dropping to 0.24 after 16 years (Fujita and Diener, 2005).

Generally speaking, country averages of measures of subjective well-being show higher levels of stability than do those for individuals. Diener (2011), for example, reports a correlation coefficient of 0.93 for the Cantril Ladder in the Gallup World Poll over 1 year, and a correlation of 0.91 across 4-year intervals.

There is less information available on the reliability of measures of affect and eudaimonic well-being than is the case for measures of life evaluation. However, the available information is largely consistent with the picture for life satisfaction. In terms of internal consistency reliability, Diener et al. (2009) report that their Psychological Well-Being scale has a Cronbach's alpha of 0.86 ($N = 568$), whilst the positive, negative and affective balance subscales of their Scale of Positive and Negative Experience (SPANE) have alphas of 0.84, 0.80 and 0.88, respectively ($N = 572, 567$ and 566).

In the case of test-retest reliability, which one might expect to be low in the case of momentary affect, but higher in the case of longer-term affective experiences, Krueger and Schkade (2008) report test-retest scores of between 0.5 and 0.7 for a range of different measures of affect over a 2-week period. Watson, Clark and Tellegen (1988) report slightly lower scores of between 0.39 and 0.71 for a range of different measures over an 8-week period. The lower scores are recorded by measures of momentary affect, while the upper scores are for questions focusing on affective states over a longer period of time, so the range of scores is consistent with expectations. Diener et al. (2009) meanwhile report a correlation of 0.71 ($N = 261$) between measures of Psychological Well-Being issued one

month apart, whilst the positive, negative and affect balance measures of the SPANE had coefficients of 0.62, 0.62 and 0.68, respectively ($N = 261$). Clark and Senik (2011) meanwhile report a Cronbach's alpha of 0.63 in their eudaimonia measure, which is derived from items in the well-being module of the European Social Survey ($N =$ over 30 000 respondents).

Summary: Reliability

Taken as a whole, test-retest scores for measures of subjective well-being are generally lower than is the case for commonly collected demographic and labour market statistics such as education and income. These variables generally show a test-retest score in the region of 0.9 (Krueger and Schkade, 2008) – although one would of course expect education and income to vary less over very short time periods. However, the scores for measures of subjective well-being are higher than those found for some more cognitively challenging economic concepts and for those that one would expect to be more variable over time. For example, an analysis of expenditure information found test-retest values of 0.6 over a period of one week (Carin, G., D. Evans, F. Ravndal and K. Xu, 2009).

In general, a cut-off at 0.7 is considered an acceptable level of internal consistency reliability for tests based on comparing results when using different measures (Nunnally and Bernstein, 1994; Kline, 2000). By this criterion, the more reliable multi-item measures of subjective well-being, such as the satisfaction with life scale, exhibit high reliability, although they are not as reliable as demographic statistics or educational status. The case for single item measures is more ambiguous, although the analysis of Lucas and Donellan, which has the best measures and largest dataset of any of the studies considered here, suggests that single item measures of life satisfaction also have an acceptable degree of reliability. Looking at country averages, the reliability of life satisfaction measures is generally well above the required threshold for acceptable reliability.

Measures of affect would be expected to have lower levels of reliability than is the case for evaluative measures, simply because moods change more frequently. The available evidence is generally consistent with this, and suggests that affect measures are reliable enough for use. There is less evidence on measures of eudaimonia. Although the Diener/Wirtz Psychological Well-being Scale performs relatively well, this cannot necessarily be generalised to other measures of eudaimonic well-being, and further research is warranted in this area.

Validity

While the reliability of a measure is largely a function of the degree to which it produces consistent results, and is therefore relatively easy to test, validity is more challenging to establish. This is particularly the case for subjective measures. Although there is a broad consensus around the range of concepts to be measured, this does not, in and of itself, mean that establishing the validity of a measure is straight-forward. If a measure is subjective, how can we know whether it is a good measure of the underlying concept?

At this point some precision is required about what is meant by a subjective measure. There are two senses in which one can talk about a subjective measure of something. A subjective measure can refer either to the measure itself, or to the concept being measured. In the first sense, it is possible to have a subjective measure of an objective concept (as in the case of the question, "who do you think is older, John or Marama?"). The measure is subjective in that it seeks a person's opinion, but the subject being measured (the ages of John and Marama) can be objectively verified, i.e. by checking the dates of birth for both John and Marama, different people will provide the same outcome.

When the concept itself is subjective, however, things become more complicated. In the case of the question “how much do you like the colour blue”, the concept being measured is itself subjective. There is no way for a person other than the respondent to provide the correct answer. This makes testing the validity of such measures much more challenging than in the first instance.

Measures of subjective well-being are subjective in this second sense, and this means that we cannot compare measures of subjective well-being with objective measures of the same concept in order to reassure us of their validity. However, this does not mean that we cannot meaningfully analyse the validity of measures of subjective well-being at all. There is an extensive psychological literature on the validity of subjective measures, and this literature suggests three types of validity that a good subjective measure should demonstrate:

- Face validity, i.e. do respondents and/or data users judge that the items are appropriate, given what they are told about the assessment objectives?
- Convergent validity, i.e. does the measure correlate well with other proxy measures for the same underlying concept?
- Construct validity, i.e. does the measure perform in the way theory would suggest with respect to the construct being measured?

Face validity

The face validity of measures is important because it can affect both respondent motivation and the acceptability of resulting data to potential users of that data. The face validity of subjective well-being is relatively straight-forward to establish. The standard questions used have a clear intuitive relationship to the concept being measured. It is not a great stretch, for example, to suggest that asking a person whether they experienced sadness during the previous day is a plausible way to find out whether they felt sad during that day. Although it is relatively unusual to ask respondents about face validity directly, there are a number of pieces of evidence that suggest that respondents find questions on subjective well-being easy to understand. For example, the time it takes respondents to reply to questions on subjective well-being is low, with median response times well under thirty seconds for single item questions (ONS, 2011). This indicates that respondents do not generally find such measures conceptually difficult to understand. Cognitive testing by the ONS also supports the view that respondents do not generally find subjective questions difficult or upsetting to answer, nor does the inclusion of such questions negatively impact the response rates to subsequent questions or to the survey as a whole (ONS, 2011; ONS, 2012). Measures of subjective well-being also have low item-specific non-response rates (Rässler and Riphahn, 2006), suggesting that respondents do not find these types of question difficult to answer.

In a large analysis by Smith (2013) covering three datasets (the Gallup World Poll, the European Values Survey and the World Values Survey) and over 400 000 observations, item-specific non-response rates for measures of life evaluation and affect were found to be similar to those for measures of educational attainment, marital status and labour force status. The acceptability of the subjective well-being measures, however, appeared to be higher than that of income, which is commonly collected as part of official statistics, and had an item-specific non-response rate of between 10 and 100 times higher than subjective well-being measures, depending on the country. The results also held when item-specific

non-response rates were broken down by cause (into “don’t know” and “refused to answer”) and regardless of whether the measure tested was evaluative (life satisfaction or the Cantril Ladder) or affective.

Convergent validity

Convergent validity involves examining whether a measure correlates well with other proxy measures for the same concept. Although measures of subjective well-being are focused on an inherently subjective concept, a range of information can be used as proxy measures for people’s subjective states. One option is to look at how ratings from the respondent compare to ratings from other people, such as friends, families, or even the interviewer. Similarly, one can observe the behaviour of the respondent to see whether it is consistent with their reported subjective state. Finally, one can use biophysical measures related to emotion states. All of these approaches have been applied to measures of subjective well-being and provide strong support for convergent validity.

Ratings of a person’s subjective well-being from friends and family have been shown to correlate well with self-reports of life satisfaction (Frey and Stutzer, 2002). A review by Pavot and Diener (1993) found correlations of between 0.43 and 0.66 between interviewer ratings and self-ratings, and correlations of between 0.28 and 0.58 between self-reports and other informants, such as friends and families. In a meta-analysis of self-informant ratings, Schneider and Schimmack (2009) found a mean correlation of 0.42 between self-reports of life satisfaction and informant reports. Similarly, for momentary affect, strangers shown a video or pictures of the respondent are able to accurately identify the subject’s dominant emotion at a particular point in time (Diener, Suh, Lucas and Smith, 1999). This latter finding also held when the informant was a person whose culture differed fundamentally from that of the respondent.

Subjective assessments of well-being are also reflected in behaviour. People who rate themselves as happy tend to smile more. This applies particularly to so-called “Duchenne” or “unfakeable” smiles, where the skin around the corners of the eye crinkles through a largely involuntary reflex (Frey and Stutzer, 2002; Diener, 2011). There is also good evidence that people act in ways that are consistent with what they say about their subjective well-being, i.e. people avoid behaviours that they associate with a low level of subjective well-being (Frijters, 2000). Diener (2011), summarising the research in this area, notes that life satisfaction predicts suicidal ideation ($r = 0.44$), and that low life satisfaction scores predicted suicide 20 years later in a large epidemiological sample from Finland (after controlling for other risk factors such as age, gender and substance use). Self-reports of job satisfaction have been shown to be a strong predictor of people quitting a job, even after controlling for wages, hours worked and other individual and job-specific factors (Clark, Georgellis and Sanfey, 1998).

There have been a number of studies that look at the correlation between various bio-physical markers and subjective well-being. Measures of subjective well-being have been shown to be correlated with left/right brain activity (Urry et al., 2004; Steptoe, Wardle and Marmot, 2005). Activity in the left prefrontal cortex of the brain has been shown to be strongly correlated with processing approach and pleasure, while activity in the corresponding part of the right hand side of the brain is correlated with processing avoidance and aversive stimuli. Steptoe, Wardle and Marmot (2005) also investigated the association between the level of the stress hormone cortisol in the bloodstream and self-reported happiness, finding a 32% difference in cortisol levels between people in the

highest and lowest quintiles of happiness. People reporting high levels of subjective well-being also recover more quickly from colds and minor injuries (Kahneman and Krueger, 2006).

There is no clear cut-off point for what constitutes an acceptable level of convergent validity. This is because the measured relationship depends as much on the quality of the proxy variable to which the measure is being compared as it does on the validity of the measure itself. However, the available evidence on convergent validity does allow for claims to be made about subjective well-being measures from the perspective of falsifiability. In particular, if it were found that a plausible proxy measure of subjective well-being showed no or a negative relationship with a measure of subjective well-being, this would be taken as good evidence that the measure in question lacked validity until the finding was over-turned, either because a good explanation for the relationship was found or because the result failed to be replicated. The consistent positive relationship found between measures of life satisfaction and the wide range of proxy measures considered above suggests strongly that such measures can be considered as displaying adequate convergent validity. The evidence for the convergent validity of affective measures is also persuasive. However, there is little to draw on with respect to convergent validity and eudaimonic measures.

Construct validity

Where convergent validity involves assessing the degree to which a measure correlates with other proxy measures of the same concept, construct validity focuses on whether the measure performs in the way that theory would predict. Construct validity concerns itself with whether the measure shows the expected relationship with the factors thought to determine the underlying concept being measured, and with outcomes thought to be influenced by the measure in question. There is an extensive literature relevant to the assessment of the construct validity of measures of subjective well-being. Economists in particular, driven in part by the desire to understand how well such measures perform as a potential measure of utility, have looked in depth at the economic and social drivers of subjective well-being. Meanwhile, psychologists have often explored individual, psychological and psycho-social determinants.

Measures of subjective well-being broadly show the expected relationship with other individual, social and economic determinants. Among individuals, higher incomes are associated with higher levels of life satisfaction and affect, and wealthier countries have higher average levels of both types of subjective well-being than poorer countries (Sacks, Stevenson and Wolfers, 2010). At the individual level, health status, social contact, education and being in a stable relationship with a partner are all associated with higher levels of life satisfaction (Dolan, Peasgood and White, 2008), while unemployment has a large negative impact on life satisfaction (Winkelmann and Winkelmann, 1998). Kahneman and Krueger (2006) report that intimate relations, socialising, relaxing, eating and praying are associated with high levels of net positive affect; conversely, commuting, working, childcare and housework are associated with low levels of net positive affect. Boarini et al. (2012) find that affect measures have the same broad sets of drivers as measures of life satisfaction, although the relative importance of some factors changes.

Further, it is clear that changes in subjective well-being – particularly life evaluations – that result from life events are neither trivial in magnitude, nor transient. Studies have shown that change in income, becoming unemployed, and becoming disabled have a

long-lasting impact on life satisfaction (e.g. Lucas, 2007; Lucas, Clark, Georgellis and Diener, 2003; Diener, Lucas and Napa Scollon, 2006), although there can also be substantial individual differences in the extent to which people show resilience, or are able to adapt to, adversity over time. In the case of negative life experiences, Cummins et al. (2002) note that extreme adversity is expected to result in “homeostatic defeat” – thus, life experiences such as the chronic pain of arthritis or the stress of caring for a severely disabled family member at home can lead to stably low levels of subjective well-being. Similarly, Diener, Lucas and Napa Scollon (2006) describe evidence of partial recovery from the impacts of widowhood, divorce and unemployment in the five years following these events, but subjective well-being still fails to return to the levels observed in the five years prior to these events. Thus although there is evidence of partial adaptation to changes in life circumstances, adaptation is not complete, and the impact of these life events on life evaluations is long-lasting.

Summary: Validity

Over the last two decades an extensive body of evidence has accumulated on the validity of measures of life evaluation and affect. This evidence strongly supports the view that measures of both life evaluation and affect capture valid information. This does not mean that these measures are universally valid or devoid of limitations. However, these limitations do not suggest that measures of subjective well-being should be regarded as not fit for purpose if used with appropriate caveats. The evidence base for eudaimonic measures is less clear. While psychologists have studied concepts related to eudaimonic well-being such as good psychological functioning for some time, it has proved more difficult to pull together a summary of the literature addressing validity in the terms set out above. This does not mean that eudaimonic measures are not valid, but suggests that further work is needed before a definitive position can be taken on the validity of these measures. Table 1.2 provides a summary of the evidence for the validity of subjective well-being outlined above.

Limits of validity

Although the evidence for the reliability, validity and usefulness of subjective well-being measures is strong, like all measures they are not perfect, and there are limitations that need to be considered by both producers and users of subjective well-being data.

Subjective well-being measures have been found to have a relatively high noise-to-signal ratio. For example, in reviewing the evidence, Diener (2011) states that around 60-80% of the variability in life satisfaction scales is associated with long-term factors and that the remaining 20-40% is due to occasion-specific factors and errors of measurement. These occasion-specific factors can include one-off occurrences that affect large numbers of people simultaneously, such as major news events or Valentine’s Day (Deaton, 2011), or circumstantial events that may affect individuals’ momentary mood prior to the survey (Schwarz and Strack, 2003). Whilst the latter effect should be sufficiently random to wash out of large representative data sets, the former implies that a reasonable number of *days*, as well as people, need to be sampled to reduce the risk of systematic error. This is further supported by work demonstrating that the day of the week (e.g. Taylor, 2006; Helliwell and Wang, 2011), the season (Harmatz et al., 2000) and the weather (e.g. Barrington-Leigh, 2008) can also influence certain subjective well-being measures,⁹ although results do tend to be more mixed in these areas.

Table 1.2. Evidence on the validity of measures of subjective well-being

Type of evidence	Sources
Face validity	
<ul style="list-style-type: none"> ● Item-specific non-response rates. ● Time to reply. 	Rässler and Riphahn (2006); Smith (2013); ONS (2011);
Convergent validity	
<ul style="list-style-type: none"> ● Ratings by friends and family. 	Frey and Stutzer (2002); Pavot and Diener (1993); Schneider and Schimmack (2009).
<ul style="list-style-type: none"> ● Ratings by interviewers. 	Pavot and Diener (1993).
<ul style="list-style-type: none"> ● Emotion judgements by strangers. 	Diener, Suh, Lucas and Smith (1999).
<ul style="list-style-type: none"> ● Frequency/intensity of smiling. 	Frey and Stutzer (2002); Kahneman and Krueger (2006); Seder and Oishi (2012).
<ul style="list-style-type: none"> ● Changes in behaviour. 	Frijters (2000); Diener (2011); Clark, Georgellis and Sanfrey (1998).
<ul style="list-style-type: none"> ● Biophysical measures. 	Urry et al. (2004); Steptoe, Wardle and Marmot (2005); Kahneman and Krueger (2006)
<ul style="list-style-type: none"> ● Relationships among different evaluative, affective and/or eudaimonic measures. 	Diener, Helliwell and Kahneman (2010); Kahneman and Krueger (2006), Clark and Senik (2011); Diener, Wirtz, Biswas-Diener, Tov, Kim-Prieto, Choi and Oishi (2009); Huppert and So (2009)
Construct validity	
<ul style="list-style-type: none"> ● Association with income (individual and national level). 	Sacks, Stevenson and Wolfers (2010).
<ul style="list-style-type: none"> ● Life events (e.g. impact of becoming unemployed, married, disabled, divorced or widowed). 	Diener, Lucas and Napa Scollon (2006); Lucas (2007); Lucas, Clark, Georgellis and Diener (2003); Winkelmann and Winkelmann (1998).
<ul style="list-style-type: none"> ● Life circumstances (health status, education, social contact, being in a stable relationship). 	Dolan, Peasgood and White (2008); NEF (2009).
<ul style="list-style-type: none"> ● Daily activities (e.g. commuting, socialising, relaxing, eating, praying, working, childcare, housework). 	Kahneman and Krueger (2006); Frey and Stutzer (2008); Helliwell and Wang (2011); Stone (2011).

Subjective well-being measures can also be sensitive to specific aspects of the survey content. For example, Deaton (2011) found that asking questions about whether or not the country is going in the right direction immediately before an evaluative subjective well-being measure exerted a strong downward influence on the data. Similarly, a number of authors have shown a question order effect when life satisfaction and dating or marriage satisfaction questions are asked (e.g. Strack, Martin and Schwarz, 1988; Schwarz, Strack and Mai, 1991; Tourangeau, Rasinski and Bradburn, 1991). Pudney (2010) finds some evidence that the survey mode impacts the relationship between different satisfaction domains and their objective drivers. These effects are real and can have significant implications for measurement and fitness for use. However, they are largely factors that have the potential to be managed through consistent survey design. Chapter 2 discusses these issues and the best way of handling them.

Differences may also exist among respondents in terms of how questions are interpreted, how different response formats and scale numbers are used, and the existence of certain response styles, such as extreme responding and acquiescence. Socially desirable responding may also impact the mean levels of reported subjective well-being. The evidence for these effects, and the methodological implications, are discussed in further detail in Chapter 2. To the extent that these differences are randomly distributed among populations of interest, they will contribute to random “noise” in the data without necessarily posing a fundamental challenge to data users. However, where they systematically vary across different survey methods, and/or where they affect certain groups, nationalities or cultures differently, this can make the interpretation of group and sample differences in subjective well-being problematic.

Group differences in scale use may arise for a number of reasons, including translation issues (e.g. Veenhoven, 2008; Oishi, 2010), differences in susceptibility to certain response styles (Hamamura, Heine and Paulhus, 2008; Minkov, 2009; van Herk, Poortinga and Verhallen, 2004), or the cultural relevance and sensitivity of certain subjective well-being questions (Oishi, 2006; Vittersø, Biswas-Diener and Diener, 2005). Various methods do exist to detect and control for these effects, and survey design can also seek to minimise this variability, as is described in Chapter 2. However, further research is needed to inform the best approach to international comparisons in particular. That said, there is evidence to suggest that these responses do not extensively bias the analysis of determinants in micro-data (Helliwell, 2008; Fleche, Smith and Sorsa, 2011). At the national level, there remain clear and consistent relationships between objective life circumstances and subjective well-being (Helliwell, 2008), and these differences are reflected in mean scores – such that, for example, the distribution of life satisfaction scores in Denmark and Togo are almost non-overlapping (Diener, 2011).

A final consideration is the extent to which individual, cultural or national fixed effects influence subjective well-being measures – including differences in personality and dispositional affect (Diener, Oishi and Lucas, 2003; Diener, Suh, Lucas and Smith, 1999; Schimmack, Diener and Oishi, 2002; Suls and Martin, 2005). These differences may be due to genetic factors (e.g. Lykken and Tellegen, 1996) or environmental factors in a person's development that produce chronically accessible and stable sources of information on which subjective well-being assessments may be based (e.g. Schimmack, Oishi and Diener, 2002). To the extent that public policy can shape a person's life experiences, particularly in developmental phases, the existence of large and relatively stable individual fixed effects does not necessarily mean that measures are insensitive to the effects of policy interventions – but the time frames over which these experiences take effect may be quite long.

It is clear that individual fixed effects are real, and they account for a significant proportion of variance across individuals. For example, in two national longitudinal panel studies, Lucas and Donnellan (2011) found that 34-38% of variance in life satisfaction was due to stable trait differences, and a further 29-34% of additional variance was due to an autoregressive component that was moderately stable in the short-term, but could fluctuate over longer time periods. However, this study did not include measures of the objective life circumstances that might impact on both stable trait-like and autoregressive components. Different cultures and nations also vary in both mean levels (e.g. Bjørnskov, 2010; OECD, 2011) and in the composition or construction of reported subjective well-being (e.g. Schimmack, Oishi and Diener, 2002; Diener, Napa Scollon, Oishi, Dzokoto and Suh, 2000; Oishi, 2006) – although again, it is rare to find research that explicitly documents the contribution of national or cultural fixed effects over and above objective life circumstances.

It is possible that individual, cultural and national fixed effects are substantive, in so far as they have a genuine impact on how people subjectively feel about their well-being. If this is the case, they should not be regarded as measurement error. In practice, however, it is not always easy to disentangle fixed effects in actual experienced subjective well-being from differences in translation, response styles, question meaning and retrospective recall biases – although Oishi (2010) suggests that measurement artefacts (such as differences in number use, item functioning and self-presentation) play a relatively small role in explaining overall national differences at the mean level. What we do know is that, although there is a reasonably stable component to life evaluations, these measures are still sensitive to life circumstances, and do change over time in response to life events (Lucas, 2007; Lucas, Clark, Georgellis and Diener, 2003; Diener, Lucas and Napa Scollon,

2006). Thus, the main drivers of subjective well-being of interest to policy-makers, including long-term life circumstances that may influence resilience to the impact of negative life events, can still be examined. Panel data can also be used so that the impact of changes in life circumstances (including policy interventions) can be examined whilst controlling for fixed effects where necessary.

Summary: Limits of validity

While there are a wide range of issues that potentially limit the validity of subjective measures of well-being, many of these are either of marginal impact in terms of fitness for purpose (i.e. they do not substantively affect the conclusions reached) or can be dealt with through appropriate survey design. For example, any contextual effects from the preceding question will not bias the analysis of changes over time if the survey itself does not change over time. In practical terms, the impact of these limitations on fitness for purpose depends very much on the purpose for which the data is being used. These issues will be dealt with in Chapter 2. One major limit does, however, need to be acknowledged. Despite evidence that cultural factors do not substantively bias multi-variate analysis, there is good reason to be cautious about cross-country comparisons of levels of subjective well-being – particularly life satisfaction.

Coherence and the measurement of subjective well-being

Coherence addresses the degree to which different measures of the same underlying construct tell the same story. Two similar statistics might reflect slightly different concepts, and hence not be comparable even though they are both highly accurate. Coherence depends on the use of common concepts, definitions and classifications. For this reason, the issue of coherence in measures of subjective well-being essentially reduces to the case for a common measurement framework. As discussed in the first section of this chapter, this is the primary rationale for producing a set of guidelines. While the guidelines themselves will not initially constitute a standard, they are a necessary initial step in the process that might later lead to a formal standard.

Conclusion

The case for national statistical agencies producing measures of subjective well-being depends on the relevancy and accuracy of such measures. If it is possible to produce measures of subjective well-being that are of sufficient quality to be fit for purpose, and which are of use either in formulating policy or in informing the broader public in useful ways, then there is a strong argument in favour of collecting and publishing such measures. Over the last ten years, measures of subjective well-being have moved increasingly from the realm of academic discourse into the realm of informing policy-making and public debate. There is now a sufficient body of knowledge on how measures of subjective well-being will be used to make a strong *prima facie* case for their relevance. While ultimately the proof of relevance will depend on up-take and use, this cannot occur until measures are produced, at least on an experimental basis. For this reason it is important that at least some official statistical agencies start producing regular series of subjective well-being data to support policy-makers and allow more informed decisions about the ultimate usefulness of such data. Only when such data is available for some time can more definite judgements be reached on relevance.

There is a large body of evidence on the reliability and validity of measures of subjective well-being and the methodological challenges involved in collecting and analysing such data. Indeed, given the academic interest in the topic and the challenging nature of the subject, the body of evidence on the strengths and weaknesses of measures of subjective well-being exceeds that for many measures regularly collected as part of official statistics (Smith, 2013). While measures of subjective well-being have some important limitations, there is no case for simply considering subjective measures of well-being “beyond the scope” of official statistics. Although subject to some methodological limitations, it is clear that for many of the purposes for which they are desired, measures of subjective well-being are able to meet the basic standard of “fit for purpose”.

However, there are also a number of areas where measures of subjective well-being are more strongly affected by issues to do with how the measure is collected and with potentially irrelevant characteristics of the respondent than is the case for many other official statistics. This does not suggest that measures of subjective well-being should be rejected outright, but points to two important steps. First, there is a need for official measures of subjective well-being to be collected – possibly as experimental data series – in order to provide the basis to resolve some of the outstanding methodological issues. Second, it is important that information on the nature of the most significant methodological issues associated with collecting measures of subjective well-being is available to potential producers of subjective well-being data, and that a common approach to dealing with these issues is developed. This is the focus of the second chapter.

Notes

1. E.g. “A New Gauge to See What’s Beyond Happiness”, *New York Times*, 16 May 2011.
2. The definition used is taken largely from Diener et al. (2006).
3. By “decision utility” Kahneman refers to the sort of evaluation used by individuals to make choices between different options. He distinguishes this from “experienced utility” which is the sum of felicitic experience for an individual over time. The former approaches what economists mean by utility in standard microeconomic models, while the latter is closer to Jeremy Bentham’s original notion of utility in the context of utilitarianism (Bentham, 1789).
4. For example, the Gallup World Poll contains a range of questions on affect during the previous day, which have been extensively tested. The UK Office of National Statistics has collected similar measures of affect in its Integrated Household Survey programme.
5. One concern sometimes raised about subjective measures is that they are unlikely to change as fast over time as more traditional indicators. In fact, this is not strictly true (see Box 1.2). However, even for those circumstances where measures of subjective well-being do not change as much as, say, resource-based measures, this should be regarded as information rather than as a problem with the measure.
6. There are two primary reasons why subjective well-being might be considered to differ substantially from overall well-being. First, subjective well-being is affected by a number of factors, such as personality and culture, which might be considered a source of bias in terms of measuring actual well-being. Second, most theories of well-being are not strictly utilitarian in nature and recognise standards that are important to well-being regardless of their association with a person’s subjective state. For example, Sen (1979) defines well-being in terms of achieved “capabilities” to do certain things, explicitly rejecting a subjective (utilitarian) alternative.
7. Consideration of initial sample variance in each measure is important here: if the sample has uniformly high levels of health satisfaction, but variable levels of housing satisfaction, housing satisfaction may look more important in a regression analysis, simply because it has more variation to associate with variation in the outcome measure.

8. Cronbach's coefficient alpha is considered to be the most stable and accurate index of internal consistency reliability (Kline, 2000; Nunnally and Bernstein, 1994). Provided all item standard deviations are equal, alpha is mathematically equivalent to the mean of all split-half reliabilities (i.e. dividing test items into two halves and correlating those halves with one another); it is slightly lower than the mean where item standard deviations vary (Cortina, 1993). Alpha is calculated by multiplying the mean average inter-item co-variance by the number of items in a test (which estimates the true score variance) and dividing this figure by the sum of all the elements in the variance-covariance matrix, which equals the observed test variance (Nunnally and Bernstein, 1994).
9. Not all of this "noise" is strictly error, however. The sensitivity of affect measures to the day of the week, for example, validates these measures to the extent that individuals participate in more pleasurable activities, such as time spent with friends and family, on weekends (Heliwell and Wang, 2011; Stone, 2011).

Bibliography

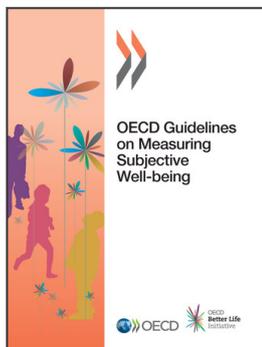
- Barrington-Leigh, C.P. (2008), "Weather as a Transient Influence on Survey-Reported Satisfaction with Life", *Munich Personal RePEc Archive (MPRA) Paper*, No. 25736, University of British Columbia, Vancouver, available online at: <http://mpra.ub.uni-muenchen.de/25736/>.
- Bartam, D. (2010), "Economic Migration and Happiness: Comparing Immigrants' and Natives' Happiness Gains From Income", *Social Indicators Research*, Vol. 103, No. 1, pp. 57-76.
- Bentham, J. (1789), *An Introduction to the Principles of Morals and Legislation*, Oxford, Clarendon Press.
- Bjørnskov, C. (2010), "How Comparable are the Gallup World Poll Life Satisfaction Data?", *Journal of Happiness Studies*, Vol. 11, pp. 41-60.
- Boarini, R., M. Comola, C. Smith, R. Manchin and F. De Keulenaer (2012), "What makes for a better life: the determinants of subjective well-being in OECD countries, evidence from the Gallup World Poll", *OECD STD Working Paper*.
- Carrin, G., D. Evans, F. Ravndal and K. Xu (2009), "Assessing the reliability of household expenditure data: Results of the World Health Survey", *Health Policy*, Vol. 91(3), pp. 297-305.
- Clark, A. (2001), "What really matters in a job? Hedonic measurement using quit data", *Labour Economics*, No. 8, pp. 223-242.
- Clark, A.E., Y. Georgellis and P. Sanfey (1998), "Job Satisfaction, Wage Changes and Quits: Evidence from Germany", *Research in Labor Economics*, Vol. 17.
- Clark, A. and A. Oswald (2002), "A simple statistical method for measuring how life events affect happiness", *International Journal of Epidemiology*, No. 31, pp. 1139-1144.
- Clark, A. and C. Senik (2011), "Is happiness different from flourishing? Cross-country evidence from the ESS", *Working Paper 2011-04*, Paris, School of Economics.
- Clifton, J. and L. Morales (2011), "Egyptians', Tunisians' Wellbeing Plummet Despite GDP Gains", Gallup Organisation.
- Cortina, J.M. (1993), "What is coefficient alpha? An examination of theory and applications", *Journal of Applied Psychology*, Vol. 78(1), pp. 98-104.
- Cummins, R., R. Eckersley, J. Pallant, J. Vugt and R. Misajon (2003), "Developing a National Index of Subjective Wellbeing: The Australian Unity Wellbeing Index", *Social Indicators Research*, Vol. 64, No. 2, pp. 159-190.
- Deaton, A. (2011), "The financial crisis and the well-being of Americans", *Oxford Economic Papers*, No. 64, pp. 1-26.
- Deci, E. and R. Ryan (2006), "Hedonia, Eudaimonia, and Well-being: An introduction", *Journal of Happiness Studies*, No. 9, pp. 1-11.
- Di Tella, R., R. MacCulloch and A. Oswald (2003), "The Macroeconomics of Happiness", *The Review of Economics and Statistics*, Vol. 85(4), pp. 809-827.
- Diener, E. (2011), *The Validity of Life Satisfaction Measures*, unpublished working paper.
- Diener, E. (2006), "Guidelines for National Indicators of Subjective Well-Being and Ill-Being", *Applied Research in Quality of Life*, No. 1, pp. 151-157.
- Diener, E. (1984), "Subjective Well-Being", *Psychological Bulletin*, Vol. 95, No. 3, pp. 542-575.

- Diener, E. and M. Chan (2011), "Happy People Live Longer: Subjective Well-being Contributes to Health and Longevity", *Applied Psychology: Health and Well-being*, March.
- Diener, E., R. Emmons, R. Larsen and S. Griffin (1985), "The Satisfaction With Life Scale", *Journal of Personality Assessment*, Vol. 49(1), pp. 71-75.
- Diener, E., J.F. Helliwell and D. Kahneman (eds.) (2010), *International Differences in Well-Being*, Oxford University Press.
- Diener, E., R.E. Lucas, U. Schimmack and J. Helliwell (eds.) (2009), *Well-Being For Public Policy*, Oxford University Press.
- Diener, E., R.E. Lucas and C. Napa Scollon (2006), "Beyond the hedonic treadmill: Revising the adaptation theory of well-being", *American Psychologist*, Vol. 61(4), pp. 305-314.
- Diener, E., S. Oishi and R.E. Lucas (2003), "Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life", *Annual Review of Psychology*, Vol. 54, pp. 403-425.
- Diener, E., C.K.N. Scollon, S. Oishi, V. Dzokoto and E.M. Suh (2000), "Positivity and the construction of life satisfaction judgements, global happiness is not the sum of its parts", *Journal of Happiness Studies*, No. 1, pp. 159-176.
- Diener, E., E. Suh, R. Lucas and H. Smith (1999), "Subjective Well-Being: Three Decades of Progress", *Psychological Bulletin*, Vol. 125, No. 2, pp. 276-302.
- Diener, E., D. Wirtz, R. Biswas-Diener, W. Tov, C. Kim-Prieto, D. Choi and S. Oishi (2009), "New Measures of Well-Being", in E. Diener (ed.) (2009), *Assessing Well-Being: The Collected Works of Ed Diener*, Social Indicators Research Series, Vol. 39, Dordrecht: Springer, pp. 247-266.
- Dolan, P., R. Layard and R. Metcalfe (2011), *Measuring Subjective Well-being for Public Policy*, Office for National Statistics.
- Dolan, P. and T. Peasgood (2006), "Valuing non-market goods: Does subjective well-being offer a viable alternative to contingent valuation", *Imperial College Working Paper*, Imperial College, London.
- Dolan P., T. Peasgood and M. White (2008), "Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being", *Journal of Economic Psychology*, Vol. 29, pp. 94-122.
- Dolan, P. and M. White (2007), "How Can Measures of Subjective Well-Being Be Used to Inform Public Policy?", *Perspectives on Psychological Science*, Vol. 2(1), pp. 71-85.
- Easterlin, R. (1974), "Does Economic Growth Improve the Human Lot? Some Empirical Evidence", in David, P.A. and M.W. Reder, *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz*, New York, Academic Press Inc, pp. 89-125.
- Ferrer-i-Carbonell, A. and P. Frijters (2004), "How important is methodology for the estimates of the determinants of happiness?", *The Economic Journal*, No. 114, pp. 641-659.
- Fleche, S., C. Smith and P. Sorsa (2011), "Exploring determinants of subjective wellbeing in OECD countries-evidence from the World Value Survey", *OECD Working Paper*, No. 921.
- Frey, B.S. and A. Stutzer (2011), "The use of happiness research for public policy", *Social Choice Welfare*, Vol. 38(4), pp. 659-674.
- Frey, B.S. and A. Stutzer (2008), "Stress that Doesn't Pay: The Commuting Paradox", *The Scandinavian Journal of Economics*, Vol. 110(2), pp. 339-366.
- Frey, B.S. and A. Stutzer (2005), "Happiness Research: State and Prospects", *Review of Social Economy*, Vol. 63(2), pp. 207-228.
- Frey, B.S. and A. Stutzer (2002), "What Can Economists Learn from Happiness Research?", *Journal of Economic Literature*, Vol. 40(2), pp. 402-435.
- Frijters, P. (2000), "Do individuals try to maximize general satisfaction?", *Journal of Economic Psychology*, No. 21, pp. 281-304.
- Fujita, F. and E. Diener (2005), "Life satisfaction set point: stability and change", *Journal of Personality and Social Psychology*, Vol. 88(1), p. 158.
- Fujiwara, D. and R. Campbell (2011), "Valuation Techniques for Social Cost-Benefit Analysis: Stated Preference, Revealed Preference and Subjective Well-Being Approaches", *Green Book Discussion Paper*.
- Gallup (2011), "Egyptians', Tunisians' Wellbeing Plummet despite GDP Gains", available online at: www.gallup.com/poll/145883/egyptians-tunisians-wellbeing-plummet-despite-gdp-gains.aspx.

- Halpern, D. (2010), *The Hidden Wealth of Nations*, Polity Press.
- Hamamura, T., S.J. Heine and D.L. Paulhus (2008), "Cultural differences in response styles: The role of dialectical thinking", *Personality and Individual Differences*, No. 44, pp. 932-942.
- Harmatz, M.G., A.D. Well, C.E. Overtree, K.Y. Kawamura, M. Rosal and I.S. Ockene (2000), "Seasonal Variation of Depression and Other Mood: A Longitudinal Approach", *Journal of Biological Rhythms*, Vol. 15, No. 4, pp. 344-350.
- Helliwell, J.F. (2008), "Life Satisfaction and the Quality of Development", *NBER Working Paper*, No. 14507, National Bureau of Economic Research.
- Helliwell, J.F. and C.P. Barrington-Leigh (2010), "Measuring and Understanding Subjective Well-being", *NBER Working Paper*, No. 15887, National Bureau of Economic Research.
- Helliwell, J. and S. Wang (2011), "Trust and Well-being", *International Journal of Well-being*, Vol. 1(1), pp. 42-78.
- Hicks, S. and L. Tinkler (2011), "Measuring Subjective Well-being", *ONS Supplementary Paper*.
- HM Treasury, (2011), *The Green Book: Appraisal and Evaluation in Central Government*.
- Huppert, F.A. and T.T.C. So (2009), *What percentage of people in Europe are flourishing and what characterises them?*, Well-Being Institute, University of Cambridge, mimeo prepared for the OECD/ISQOLS meeting on *Measuring subjective well-being: an opportunity for NSOs?*, Florence, 23/24 July, available online at: www.isqols2009.istitutodeglinnocenti.it/Content_en/Huppert.pdf.
- Huppert, F.A., N. Marks, A. Clark, J. Siegrist, A. Stutzer, J. Vitterso and M. Wahrendorf (2009), "Measuring Well-being Across Europe: Description of the ESS Well-being Module and Preliminary Findings", *Social Indicators Research*, No. 91, pp. 301-315.
- International Wellbeing Group (2006), *Personal Wellbeing Index*, 4th edition, Melbourne, Australian Centre on Quality of Life, Deakin University, available online at: www.deakin.edu.au/research/acqol/instruments/wellbeing_index.htm.
- Kahneman, D. and A. Deaton (2010), "High income improves life evaluation but not emotional well-being", *Proceedings of the National Academy of Sciences*, Vol. 107(38), pp. 16489-16493.
- Kahneman, D., E. Diener and N. Schwarz (1999), *Well-being: The Foundations of Hedonic Psychology*, Russel Sage Foundation, New York.
- Kahneman, D. and A. Krueger (2006), "Developments in the measurement of subjective well-being", *The Journal of Economic Perspectives*, No. 20, pp. 3-24.
- Kahneman, D., A. Krueger, D. Schkade, N. Schwarz and A. Stone (2004), "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method", *Science*, Vol. 306(5702), pp. 1776-1780.
- Kline, P. (2000), *The Handbook of Psychological Testing*, 2nd edition, Florence, KY, US: Taylor and Frances/Routledge.
- Krueger, A. and D. Schkade (2008), "The reliability of subjective well-being measures", *Journal of Public Economics*, No. 92, pp. 1833-1845.
- Krueger, A., D. Kahneman, C. Fischler, D. Schkade, N. Schwarz and A. Stone (2009), "Time Use and Subjective Well-Being in France and the US", *Social Indicators Research*, No. 93, pp. 7-18.
- Krueger, A., D. Kahneman, D. Schkade, N. Schwarz and A. Stone (2009), "National Time Accounting: The Currency of Life", in A. Krueger (ed.), *Measuring the Subjective Well-being of Nations: National Accounts of Time Use and Well-Being*, NBER.
- Larson, R.J. and B.L. Fredrickson (1999), "Measurement Issues in Emotion Research", in D. Kahneman, E. Diener and N. Schwarz (eds.), *Well-being. The Foundations of Hedonic Psychology*, Russel Sage Foundation, New York, pp. 40-60.
- Lucas, R.E. (2007), "Long-term disability is associated with lasting changes in subjective well-being: Evidence from two nationally representative longitudinal studies", *Journal of Personality and Social Psychology*, Vol. 92(4), pp. 717-730.
- Lucas, R.E., A.E. Clark, Y. Georgellis and E. Diener (2003), "Re-examining adaptation and the set point model of happiness: Reactions to changes in marital status", *Journal of Personality and Social Psychology*, Vol. 84, pp. 527-539.
- Lucas, R.E. and M. Donnellan (2012), "Estimating the Reliability of Single-Item Life Satisfaction Measures: Results from Four National Panel Studies", *Social Indicators Research*.

- Lykken, D. and A. Tellegen (1996), "Happiness is a stochastic phenomenon", *Psychological Science*, No. 7, pp. 186-189.
- Michalos, A. and P. Kahlke (2010), "Stability and Sensitivity in Perceived Quality of Life Measures: Some Panel Results", *Social Indicators Research*, No. 98, pp. 403-434.
- Minkov, M. (2009), "Nations with more dialectical selves exhibit lower polarization in life quality judgements and social opinions", *Cross-Cultural Research*, No. 43, pp. 230-250.
- New Economics Foundation, (2009), *National Accounts of Well-being*.
- Nunnally, J.C. and I.H. Bernstein (1994), *Psychometric Theory (Third Edition)*, New York: McGraw-Hill.
- OECD (2011), *How's Life? Measuring Well-being*, OECD Publishing.
- OECD (2008), *Quality Framework and Guidelines for OECD Statistical Activities*, OECD Publishing.
- Oishi, S. (2010), "Culture and well-being: Conceptual and methodological issues", in E. Diener, J.F. Helliwell and D. Kahneman (eds.), *International Differences in Well-Being*, Oxford: Oxford University Press.
- Oishi, S. (2006), "The concept of life satisfaction across cultures: An IRT analysis", *Journal of Research in Personality*, Vol. 40(4), pp. 411-423.
- ONS (2012), *Subjective Well-being User Guide: 12 Month Data Set*, Office for National Statistics.
- ONS (2011a), *Initial investigation into Subjective Wellbeing from the Opinions Survey*, Office for National Statistics.
- ONS (2011b), *Measuring Subjective Well-being*, Office for National Statistics, United Kingdom.
- Pavot, W. and E. Diener (1993), "Review of the Satisfaction with Life Scale", *Psychological Assessment*, Vol. 5, No. 2, pp. 164-172.
- Pavot, W., E. Diener, C.R. Colvin and E. Sandvik (1991), "Further validation of the satisfaction with life scale: Evidence for the cross-method convergence of well-being measures", *Journal of Personality Assessment*, Vol. 57(1), pp. 149-161.
- Pudney, S. (2010), *An experimental analysis of the impact of survey design on measures and models of subjective wellbeing*, Institute for Social and Economic Research.
- Rässler, S. and R.T. Riphahn (2006), "Survey item nonresponse and its treatment", *Allgemeines Statistisches Archiv*, No. 90, pp. 217-232.
- Sacks, W.D., B. Stevenson and J. Wolfers (2010), "Subjective Well-being, Income, Economic Development and Growth", *NBER Working Paper*, No. 16441.
- Schimmack, U., E. Diener and S. Oishi (2002), "Life satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources", *Journal of Personality*, Vol. 70(3), pp. 345-382.
- Schimmack, U., S. Oishi and E. Diener (2002), "Cultural influences on the relation between pleasant emotions and unpleasant emotions: Asian dialectic philosophies or individualism-collectivism?", *Cognition and Emotion*, Vol. 16(6), pp. 705-719.
- Schneider, L. and U. Schimmack (2009), "Self-Informant Agreement in Well-Being Ratings: A Meta-Analysis", *Social Indicators Research*, Vol. 94(3), pp. 363-376.
- Schwartz, N. and F. Strack (2003), "Reports of Subjective Well-Being: Judgemental Processes and their Methodological Implications", in D. Kahneman, E. Diener and N. Schwartz (eds.), *Well-being: The foundations of hedonic psychology*, Russell Sage Foundation, New York.
- Schwarz, N., F. Strack and H. Mai (1991), "Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis", *Public Opinion Quarterly*, Vol. 55, pp. 3-23.
- Seder, J.P. and S. Oishi (2012), "Intensity of smiling in facebook photos predicts future life satisfaction", *Social Psychological and Personality Science*, Vol. 3(4), pp. 407-413.
- Sen, A. (1979), *Equality of What?*, Tanner Lecture on Human Values, Stanford University.
- Smith, C. (2013), "Making happiness count: Four myths about subjective well-being", *OECD Working Paper*, forthcoming.
- Steptoe, A., J. Wardle and M. Marmot (2005). "Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102(18), pp. 6508-6512.

- Stevenson, W. (2008), "Economic Growth and Subjective Wellbeing: Reassessing the Easterlin Paradox", NBER Working Paper, No. 14282.
- Stiglitz, J.E., A. Sen and J.P. Fitoussi (2008), *Report by the Commission on the Measurement of Economic Performance and Social Progress*.
- Stone, A.A. (2011) "A rationale for including a brief assessment of hedonic well-being in large-scale surveys", *Forum for Health Economics and Policy*, Vol. 14(3), Article 7.
- Strack, F., L. Martin and N. Schwarz (1988), "Priming and Communication: The Social Determinants of Information Use in Judgments of Life Satisfaction", *European Journal of Social Psychology*, Vol. 18, pp. 429-42.
- Suls, J. and R. Martin (2005). "The Daily Life of the Garden-Variety Neurotic: Reactivity, Stressor Exposure, Mood Spillover, and Maladaptive Coping", *Journal of Personality*, Vol. 73(6), pp. 1485-1510.
- Taylor, M.P. (2006), "Tell Me Why I don't Like Mondays: Investigating Day of the Week Effects on Job Satisfaction and Psychological Well-Being", *Journal of the Royal Statistical Society Series A*, Vol. 169, No. 1, pp. 127-142.
- Thompson, S. and N. Marks (2008), *Measuring well-being in policy: issues and applications*, New Economics Foundation.
- Tourangeau, R. (1999), "Context Effects on Answers to Attitude Questions", in G. Monroe (ed.), *Cognition and Survey Research*, Sirken, Wiley, New York, pp. 111-131.
- Tourangeau, R., K.A. Rasinski and N. Bradburn (1991), "Measuring Happiness in Surveys: A Test of the Subtraction Hypothesis", *Public Opinion Quarterly*, No. 55, pp. 255-266.
- Urry, H., J. Nitschke, I. Dolski, D. Jackson, K. Dalton, C. Mueller, M. Rosenkranz, C. Ruff, B. Singer and R. Davidson (2004), "Making a Life Worth Living: Neural Correlates of Well-Being", *Psychological Science*, Vol. 15, No. 6, pp. 367-372.
- Van Herk, H., Y.H. Poortinga and T.M.M. Verhallen (2004), "Response styles in rating scales: Evidence of method bias in data from six EU countries", *Journal of Cross-Cultural Psychology*, No. 35, pp. 346-360.
- Van Praag, B.M.S., P. Frijters and A. Ferrer-i-Carbonell (2003), "The anatomy of subjective well-being", *Journal of Economic Behaviour and Organisation*, No. 51, pp. 29-49.
- Veenhoven, R. (2008), "The International Scale Interval Study: Improving the comparability of responses to survey questions about happiness", in V. Moller and D. Huschka (eds.), *Quality of life and the millennium challenge: Advances in quality-of-life studies, theory and research*, Social Indicators Research Series, Vol. 35, Springer, pp. 45-58.
- Vittersø, J., R. Biswas-Diener and E. Diener (2005), "The Divergent Meanings of Life Satisfaction: Item Response Modeling of the Satisfaction With Life Scale in Greenland and Norway", *Social Indicators Research*, Vol. 74, pp. 327-348.
- Watson, D., L.A. Clark and A. Tellegen (1988), "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales", *Journal of Personality and Social Psychology*, Vol. 54(6), pp. 1063-1070.
- Weinhold, D. (2008), "How big a problem is noise pollution? A brief happiness analysis by a perturbable economist", *MPRA Working Paper*, No. 10660.
- Wilson, T.D. and D.T. Gilbert (2006), "Affective forecasting: Knowing what to want", *Current Directions in Psychological Science*, Vol. 14(3), pp. 131-134.
- Wilson, T.D., T. Wheatley, J.M. Meyers, D.T. Gilbert and D. Axsom (2000), "Focalism: A source of durability bias in affective forecasting", *Journal of Personality and Social Psychology*, Vol. 78(5), pp. 821-836.
- Winkelman, L. and R. Winkelman (1998), "Why Are the Unemployed So Unhappy? Evidence from Panel Data?", *Economica*, Vol. 65, pp. 1-15.



From:
OECD Guidelines on Measuring Subjective Well-being

Access the complete publication at:
<https://doi.org/10.1787/9789264191655-en>

Please cite this chapter as:

OECD (2013), "Concept and validity", in *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264191655-5-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.