

This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being

Volume Author/Editor: Alan B. Krueger, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-45456-8

Volume URL: <http://www.nber.org/books/krue08-1>

Conference Date: December 7-8, 2007

Publication Date: October 2009

Title: Well-Being Measurement and Public Policy

Author: Richard Layard

URL: <http://www.nber.org/chapters/c5058>

Well-Being Measurement and Public Policy

Richard Layard

Chapter 1 of this volume represents an excellent use of time by its five authors. It is a high point in an important program that has provided major insights into what people enjoy and do not enjoy in their daily lives. Equally important, intensive study of the experience of daily living provides an important way of assessing the *overall* quality of an individual's life. This method of overall assessment can provide a valuable addition to the answers to the global questions now routinely asked about life satisfaction and happiness.

My comments will focus mainly on this latter issue and will be concerned only with the ways in which well-being data can best contribute to public policy debate. I shall begin by questioning whether the U-index is the best way to represent the overall quality of a day lived. I shall then discuss how far data on enjoyment in different activities can contribute to policy debate. Finally, I shall report on some parallel developments in Britain.

6.1 The U-Index and Public Policy

6.1.1 Public Policy Usefulness

A major reason for much social science is to illuminate the public debate. It is not, of course, the only reason, but in the end much of social science gets used in policy debate. Given this, it is best to set up the inquiry so that its findings are as explicitly helpful for policymaking as possible.

Among economists, the standard approach to public policy is to think of social welfare (W) as an additive aggregate of individual happiness, H_i ,

Richard Layard is the director of the Well-Being Programme in the Centre for Economic Performance, London School of Economics.

perhaps with diminishing marginal social welfare attaching to increments of individual happiness:¹

$$W = \sum_i f(H_i) \quad (f' > 0, f'' \leq 0),$$

where H_i is an empirically measurable value, but $f(\)$ reflects the ethical perspective of whoever uses these data in public debate. There are two extreme versions of the $f(\)$ function. In the Benthamite version, we just add up everyone's happiness:

$$W = \sum_i H_i,$$

and in the Rawlsian version, we just look at the happiness of the least happy person:

$$W = \text{Maximin} (H_i).$$

If we want to compare two situations, corresponding, for example, to two different policies, we examine the sign of the change in welfare, given by:

$$(1) \quad \Delta W = \sum_i \frac{\partial f}{\partial H_i} \cdot \Delta H_i.$$

Clearly this involves comparing the magnitude of the changes in happiness experienced by different people.² So, we must be able to measure happiness in a cardinal fashion, on an interval scale, where a change of one unit in happiness means the same at different points of the scale. And these units need to be comparable between people. To the economist who objects to these ideas, I would say that we all use these ideas regularly in how we describe the world:

Cardinality: we say A's mood improved a lot today, but much less than it did yesterday.

Comparability: the bad news upset A much more than it upset B.

However, there is still the ethical issue about the function $f(\)$. A strong egalitarian might say that we can ignore changes affecting any but the least happy people, on the grounds that public policy mainly exists to protect the weak (its equity function). But this, of course, is quite wrong: much public policy exists to improve the efficiency of society, since externalities, information problems, or economies of scale raise problems that require collective action for the benefit of all. In fact, most policy actions involve equity and efficiency considerations simultaneously, which is why the perspective provided by equation (1) is important.

1. See, for example, Atkinson and Stiglitz (1980, part 2).

2. The exception is the Rawlsian case, where a purely ordinal measure of H will suffice if it is comparable across people. But most people find the Rawlsian function too extreme.

6.1.2 Properties of the U-Index

How well would changes in the U-index provide a proxy for what is needed according to equation (1)? For egalitarians, one attraction might be the focus on the lower end of the happiness distribution (though not, of course, going as far as Rawls). But even over this part of the range, much information is discarded—we do not record *how* miserable an episode was, but only whether it was miserable or not. And again, if the experience was not miserable, we do not record how good it was, but simply that it was not miserable.

So, one might think that the most natural measure for each episode would be a scalar measure of how happy the person was. However, the authors claim that these affect measures (as reported) are purely ordinal and vary between individuals, so they cannot readily be used in that way. However, they say that by comparing two of the measures, something can be learned. For example, let us take a simple example (simpler and perhaps more intuitive than the one they use). Suppose we compare the answers on the Happy scale and the Blue scale and set U equal to one iff

$$\text{Happy} < \text{Blue},$$

and otherwise zero. The argument is that if person A is more emotional in his reporting than person B, he will use higher values of both “Happy” and “Blue” to report the same state when compared with B. Thus, if we put the true ordinal measures on the horizontal axis and the reported measures on the vertical axis, we get the position shown in figure 6.1.

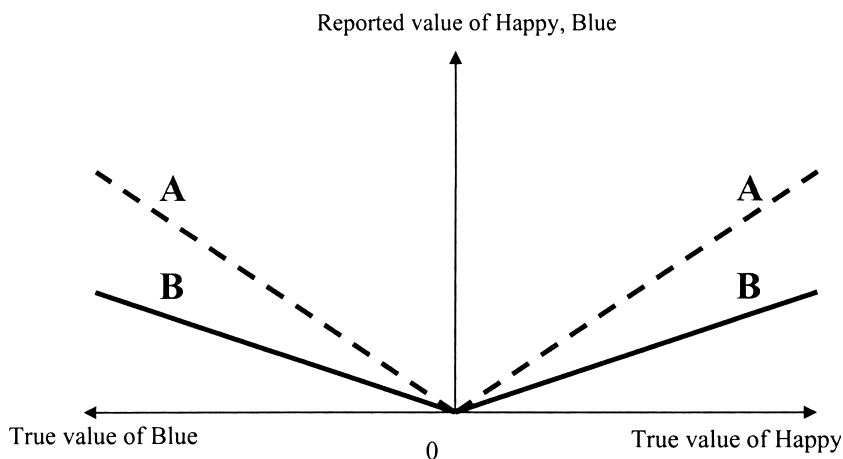


Fig. 6.1 Pattern assumed by authors

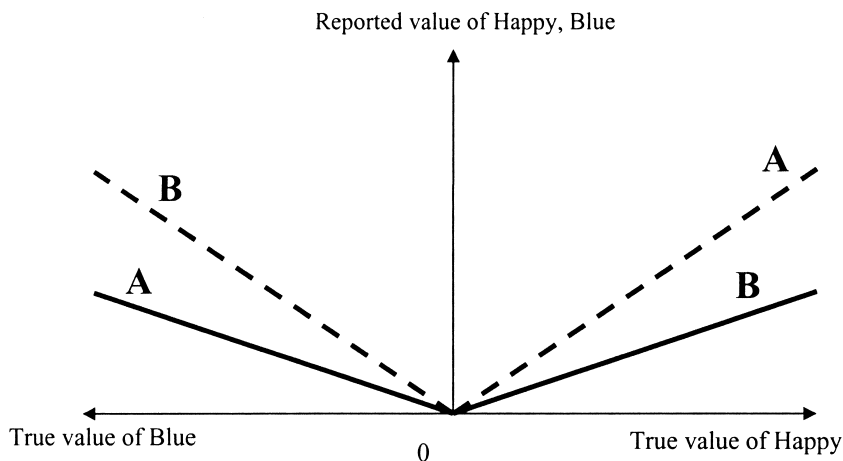


Fig. 6.2 Alternative pattern

Thus, any given state will lead to the same value of U , whether it is reported by A or B.

But the reporting pattern shown in figure 6.2 is just as likely. Here, A is more optimistic than B: he overreports “Happy” and underreports “Blue.” Now the two people may have different U values for the same state. Indeed, the chapter has in table 1.3 an illustration of this opposite mechanism: when H is asked about before Pain (rather than the other way around), the mean of H is higher and of Pain is lower.

So, if all the affect measures are truly ordinal, I do not see that the U procedure overcomes the problem. (Moreover, the procedure also requires that a person can compare on a scale of zero to six how “Happy” he is with how “Angry” he is. This is asking quite a lot. And how bad is extreme righteous anger if a person is at the same time quite high on the happiness scale?)

So, if the measures are truly ordinal, the procedure only partly handles the problem. Moreover, by comparing two numbers, it adds to problems of measurement error, while it loses so much of the information along the whole scale of H .

6.1.3 Is Happiness Purely Ordinal?

However, if we do want to use the whole scale, it cannot be purely ordinal in the sense that economists use the word—meaning that it can be arbitrarily subjected to any increasing monotonic transformation. Such a scale cannot be used to compare the magnitude of a change at one point of the scale with that at another. To do that we need a cardinal scale, which can be subjected only to an increasing *linear* transformation.

If we look at the ways in which people use the scales, they are surely not

perfectly cardinal nor perfectly comparable,³ but they do approximate to those conditions. The basic evidence, much of which is quoted in section 1.3, is that if H_i is treated as a scalar variable and is regressed on possible causal factors, the size of many causal effects is well determined. This is inconceivable with a purely ordinal scale that varies widely across individuals. Again, if we compare different studies of different populations, the effect of different experiences (like unemployment) is very similar across studies once we adjust for the length of the scale between “Very Happy” and “Very Unhappy.” It certainly looks as if respondents try to divide up that range into intervals reflecting a standard difference in intensity.

This applies not only to estimates of the first order effect of a causal variable but also to some (second order) estimates of the curvature of an effect. An example of this is a recent study by Layard, Mayraz, and Nickell (2008b). The aim of the study was to see how quickly the marginal utility of income falls as income rises—a key parameter for all public policy (including cost-benefit analysis and optimal taxation). The six surveys used are shown in table 6.1 and cover fifty countries and thirty-three years. The happiness or life-satisfaction variable in each study was put through a linear transformation to fit into a scale from zero to ten. All the analyses included (besides income) country \times year dummies, as well as sex, age, education, marital status, and employment status. The estimated equation was

$$(2) \quad H_{it} = \alpha_{ct} \frac{y_{it}^{1-\rho} - 1}{1 - \rho} + \sum_j \beta_j x_{jit} + y_{ct} + \varepsilon_{it},$$

where i is individual, t is time, c is country, j is characteristic, and ρ is the elasticity of marginal utility with respect to real income.

Table 6.2 shows the estimated values of ρ . The results of the very different surveys are remarkably close, and they do not differ significantly between subgroups of the population. Since ρ is not so far from unity, the following logarithmic formulation is a reasonably accurate approximation:

$$(3) \quad H_{it} = \alpha \log y_{it} + \sum_j \beta_j x_{jit} + y_{ct} + \varepsilon_{it}.$$

Table 6.2 shows the values of α obtained from the different surveys. The estimates are less similar than for ρ , but still remarkable, given the diversity of sources.

Clearly the finding about ρ is influenced by the assumption that people use the happiness scales in a truly cardinal way. This is not easy to check. We do ordered logit and probit analysis and obtain almost identical estimates of ρ , but this procedure depends crucially on the assumption of symmetrical cardinal errors.

3. Moreover, if we are wanting an aggregate measure of happiness across a number of time periods (episodes), there are further requirements that the scale be a ratio scale (Kahneman, Wakker, and Sarin 1997). We have no evidence that people’s replies satisfy that requirement.

Table 6.1 Surveys used by Layard et al. (2008b)

Survey	Countries	Years	Observations	Happiness variable	Income variable
General Social Survey	United States	1972–2004	17,603	Happiness (3 levels)	Yearly gross
World Values Survey	Worldwide	1981–2003	37,288	Life satisfaction (1–10)	Varies
European Social Survey	Europe	2002, 2004	26,687	Both (0–10)	Monthly net
European Quality of Life Survey	Europe	2003	8,175	Both (1–10)	Monthly net
German Socio-Economic Panel	Germany	1984–2005	78,877	Life satisfaction (0–10)	Monthly net
British Household Panel Survey	Britain	1996–2004	43,484	Life satisfaction (1–7)	Monthly net

Table 6.2 Parameter estimates for equations (1) and (2)

	ρ	α
General Social Survey	1.20 (0.91–1.48)	0.70 (0.61–0.80)
World Values Survey	1.25 (1.05–1.45)	0.62 (0.57–0.66)
European Social Survey	1.34 (1.12–1.55)	0.60 (0.55–0.64)
European Quality of Life Survey	1.19 (0.87–1.52)	0.82 (0.73–0.91)
German Socio-Economic Panel	1.26 (0.90–1.63)	0.55 (0.51–0.59)
British Household Panel Survey	1.30 (0.97–1.62)	0.35 (0.30–0.40)
Overall	1.26 (1.16–1.37)	
Subgroups		
Men	1.22 (1.06–1.39)	
Women	1.26 (1.11–1.40)	
30–42	1.27 (1.12–1.42)	
43–55	1.26 (1.10–1.41)	
Low education	1.13 (0.85–1.40)	
Mid education	1.21 (1.01–1.42)	
High education	1.26 (1.16–1.37)	
Couples	1.27 (1.11–1.43)	
Never married	1.44 (1.13–1.77)	
Others	1.34 (1.12–1.55)	

Note: Ninety-five percent confidence intervals in brackets.

The better approach is to ask what one would mean by a true interval scale of happiness. One might suppose that each unit on the scale should be proportional to a Just Noticeable Difference (JND). If this were the case and people were retested on their replies, people who scored low on the scale should have the same degree of difference between their two replies as do people high on the scale. In other words, a regression of test 2 values on test

1 values should exhibit homoscedastic errors. In a test-retest study of net affect, Krueger and Schkade (2008) did not reject homoscedasticity.

6.2 The Role of Time Use

It is also extremely interesting to know how happy people are when they are doing different things. It can aid reflection on lifestyle, and it can help with public policy.

6.2.1 Work-Life Balance

But as Chapter 1 correctly says, there are no simple public policy conclusions. For example, if people do not much enjoy their work, it does not follow that they should work less, since the marginal money they earn may justify the comparative disutility.

In discussing optimal work-life balance, the more important information would be about distortions affecting choice: that is, about externality and misforecasting. If we can show that people's happiness depends on relative income as well as on absolute income, then there is a negative externality. A number of studies have investigated the impact of other people's income upon individual happiness,⁴ but the estimates are not yet precise enough to yield estimates of optimal tax. Similarly, there is evidence that people underpredict the (negative) effect of current consumption on future happiness (Loewenstein et al. 2000). Again, the parameter estimates are not yet well defined. But studies of these issues are at least as relevant for public policy as studies of time use are.

6.2.2 Explanatory Power

But what about the explanatory power of time-use patterns in explaining the average happiness of different people or groups? I had expected the explanatory power to be greater. For example, the U-index is 2.8 points higher for Americans than for the French. But only one point of this is due to time use. Similarly, changing patterns of time use in the United States have predicted a one point fall in the U-index since the 1960s. But did it happen? And how much do differences in time use explain the differences between individuals in the sample?

One fascinating aspect of the France and United States comparison (Rennes, France versus Columbus, Ohio) is that while the U-index is higher for Americans, so is average life satisfaction. But these apparent differences are readily reconciled once we look at the *distribution* of life satisfaction (see table 6.3).

So, to explain the life satisfaction results, it does not seem necessary to

4. See Layard, Mayraz, and Nickell (2008a) and Layard (2005); see also Annex 2.5, available at <http://cep.lse.ac.uk/layard/annex.pdf>.

Table 6.3 Results reported in Krueger et al.

	United States (%)	France (%)
% not very satisfied or not at all satisfied	23.0	17.2
U-index (average)	18.8	16.0
U-index for bottom quartile	58.0	48.0

invoke differences in reporting habits, since in both types of data, the United States has a bigger tail of unhappy people. It would, however, be interesting to see how this looked if we used not the U-index, but instead used numbers below a certain level of happiness.

As table 1.21 shows, the bigger U.S. tail of unhappy people is not mainly due to greater income inequality.⁵ It must be due to other aspects of inequality, perhaps more closely related to human relationships.

6.3 Britain

In Britain, it has become a matter of practical urgency to resolve these issues of measurement, because policymakers are demanding it. There are four main clients.

The Office of National Statistics has chosen well-being as one of the three main areas for statistical development over the next year. If successful, this would put measures of well-being at the center of national government. Meanwhile, the central government department that has so far been responsible for coordinating the Whitehall approach to well-being is the Department of the Environment (because opponents of gross domestic product [GDP] maximization are either promoters of well-being or promoters of the environment). This year, the department included in its annual Indicators of Sustainable Development the results of a well-being survey covering overall life satisfaction, domain satisfaction, “feelings experienced every day or most days in last two weeks,” social activity, physical activity, and cultural activity. The measures “are presented on a provisional basis and as a starting point for possible future development.” A regular national survey of positive mental health will also be done by the Department of Health.

Finally, local governments are demanding ways of measuring well-being locally—partly to monitor trends, and equally important, to identify where the real problems are in their communities.

All this reflects, of course, the policy interest in well-being. At least three departments have a policy-making section called “X and Well-Being,” where

5. It is interesting that here, income does affect feelings measured by the U-index (see also table 6.1). This contrasts with the finding in Kahneman et al. (2006) for the Columbus sample, where feelings are measured by net affect and are broadly unrelated to income.

X includes health, work, or education. Major spending commitments have already resulted—for example, \$600 million to provide evidence-based psychological therapy in the National Health Service (NHS). Similarly, local government is responding to a statutory duty put on them to promote the well-being of their population. The interest is bipartisan, and Conservative leader David Cameron has proposed General National Well-Being (GNW) as an alternative national goal to the gross national product (GNP).

6.4 Conclusion

Let me list a few bald conclusions.

1. Detailed measurement of affect over the day provides excellent information for monitoring well-being and its distribution in the population. Both the Day Reconstruction Method (DRM) and the Princeton Affect and Time Survey (PATS) can play a great role. The team has performed a service to the world in developing these tools, and I hope the U.S. government will adopt one of them.

2. The most useful analytical measures for each individual would be scalar averages over the day, especially of happiness.

3. The feeling that well-being is fuzzy is similar to the feeling that once prevailed that depression is fuzzy. But clinical psychology has successfully developed scales (like the Beck Depression Inventory) that are no longer controversial. I have no doubt that the same can be achieved for well-being, even using scalar variables. If we worry about the measurement error involved in single questions, we should bring in other closely related questions (as in the measurement of depression). Questions about anger and stress remain interesting but may not be near enough to the basic concept of well-being to be included in the scale.

4. Determined and repetitive presentation of results from these scales will eventually result in popular understanding of the scales, just as people now understand Fahrenheit and Celsius.

5. Congratulations on a fascinating study.

References

- Atkinson, A. B., and J. E. Stiglitz. 1980. Household decisions, income taxation and labour supply. In *Lectures on public economics*, lecture 2. London: McGraw-Hill.
- Kahneman, D., A. Krueger, D. Schkade, N. Schwarz, and A. Stone. 2006. Would you be happier if you were richer? A focusing illusion. *Science* 312 (5782): 1908–10.
- Kahneman, D., P. Wakker, and R. Sarin. 1997. Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics* 112 (2): 375–405.

- Krueger, A., and D. Schkade. 2008. The reliability of subjective well-being measures. Special issue, *Journal of Public Economics* 92 (8/9): 1833–45.
- Layard, R. 2005. *Happiness: Lessons from a new science*. New York: Penguin Press.
- Layard, R., G. Mayraz, and S. Nickell. 2008a. Does relative income matter? Are the critics right? London School of Economics. Discussion Paper no. 918. London: Centre for Economic Performance.
- . 2008b. The marginal utility of income. Special issue, *Journal of Public Economics* 92 (8/9): 1846–57.
- Loewenstein, G., T. O'Donoghue, and M. Rabin. 2000. Projection bias in predicting future utility. CBDR Working Paper no. 258. Pittsburgh: Center for Behavioral Decision Research.