4 ACCEPTED MEASURES OF HAPPINESS

The practice of measuring happiness described in chapter 3 of this introductory text has been criticized. Doubts have been expressed about the validity and the reliability of self-reported happiness. I will review these objections in section 4/1_of this chapter and conclude that happiness is well measurable using self-reports. This is not to say that all self-reports ever used measure happiness equally well. Some questions meant to measure happiness address slightly different matters. Since this World Database of happiness focuses on a specific concept of happiness (cf. chapter 2) all measures are screened for fit with this concept. This selection is reported in section 4/2. In section 4/3 we will see that several commonly used measures of happiness fail that test. In section 4/4 I argue why selectiveness makes sense.

4/1 Measurability of happiness

Since happiness polls became part of our methodology in the 1960's, there has been an ongoing debate about its measurability. One issue in that discussion is whether responses to survey questions reflect how much people enjoy their life rather than something else; in other words, how 'valid' they measure the concept of happiness defined in chapter 2 of this text. Another issue is about the precision of these measures, in technical terms their 'reliability'. I have discussed these questions elsewhere (Veenhoven 1984: Ch. 4).

4/1.1 Validity

Critics have suggested that responses to questions on happiness measure other phenomena. Rather than indicating how much the respondent enjoys life, answers will reflect the respondents' normative notions and desires.

No Notion

One of the misgivings is that most people have no opinion at all about their happiness. They will be more aware of how happy they are supposed to be and report that instead. Although this may happen incidentally, it does not appear to be the rule. Most people know quite well whether they enjoy life. Eight out of ten Americans think about this every week. Responses on questions about happiness tend to be prompt. Non-response on these items is low, both absolutely (±1 %) and relatively to other attitudinal questions. "Don't know" responses are also infrequent.

A related assertion is that respondents mix up how happy they are, with how happy other people think they are, given their situation. If so, people considered to be well-off will typically report they are very happy, and people regarded as disadvantaged should characterize themselves as unhappy. This pattern is observed sometimes, but it is not general. For instance, in the Netherlands, a good education is seen as a prerequisite for a good life, but the highly educated appears to be slightly less happy in comparison to their less educated counterparts.

Coloured Answers

Another objection concerns the presence of systematic bias in responses. It is assumed that questions on happiness are interpreted correctly, but that responses are often false. People who are dissatisfied with their life will tend to answer that they are quite happy. Both ego-defence and social desirability would cause such distortions.

This bias is seen to manifest in over-report of happiness; most people claim to be happy, and most perceive themselves as happier than average. Another indication of bias is seen in the finding that psychosomatic complaints are not uncommon among the happy; however, these findings allow other interpretations as well.

Firstly, the fact that more people say they are happy than unhappy does not imply over-reporting of their happiness. It is quite possible that most people are truly happy.

Secondly, there are also good reasons why most people think that they are happier than average. One such reason is that most people are like critical scientists and think that unhappiness is the rule.

Thirdly, the occurrence of headaches and worries among the happy does not prove response distortion. Life can be a sore trial sometimes but still be satisfying on balance.

The proof of the pudding is in demonstrating the response distortion. Some clinical studies have tried to do so by comparing responses to single direct questions with ratings based on depth interviews and projective tests. The results generally do not differ from responses to single direct questions posed by an anonymous interviewer.

4/1.2 Reliability

Though single questions on happiness seem to measure what they are supposed to measure, they measure it rather imprecisely. When the same question is asked twice in an interview, responses are not always identical. Correlations are about +.70. Over a period of a week, test- retest reliability drops to circa +.60. Though responses seldom change from “happy” to “unhappy,” switches from “very” to “fairly” are rather common. The difference between response options is often ambiguous. The respondent’s notion about his/her happiness tends to be global. Thus, the choice for one answer-category or the next is sometimes haphazard.

Because choice is often arbitrary, subtle differences in interrogation can exert a considerable effect. Variations in the place where the interview is held, the characteristics of the interviewer, sequence of questions, and precise wording of the key item can tip the scale to one response or the other. Such effects can occur in different phases of the response process, during consideration of the answer and during the process of communicating the answer.

Bias in Appraisal

Though most people have an idea of how much they enjoy life, responding to questions on this matter involves more than just bringing up an earlier judgment from memory. For the most part, memory only indicates a range of happiness. Typically, the matter is reassessed in an instant judgment. This reappraisal may be limited to recent change: are there any reasons to be happy than I used to be? But it can also involve quick re-evaluation of life: what are my blessings and frustrations? In making such instant judgments, people use various heuristics.

These mental simplifications are attended with specific errors. For instance, the “availability” heuristic involves orientation on pieces of information that happen to be readily available. If the interviewer is in a wheelchair, the benefit of good health will be more salient. Respondents in good health will then rate their happiness

somewhat higher, and the correlation of happiness ratings with health variables will be more pronounced. Several of these heuristic effects have been demonstrated by Schwarz and Strack (1991).

Bias in Response

Once a respondent has formed a private judgment, the next step is to communicate it; at this stage, reports can also be biased in various ways. One source of bias is inherent to semantics; respondents interpret words differently, and some interpretations may be emphasized by earlier questions. For example, questions on happiness are more likely to be interpreted as referring to "contentment" when preceded by questions on success in work, rather than items on mood.

Another source of response bias is found in considerations of self-presentation and social desirability. Self-rating of happiness tends to be slightly higher in personal interviews than on anonymous questionnaires; however, direct contact with an interviewer does not always inflate happiness reports. Modest self-presentation is encouraged if the interviewer is in a wheelchair.

Much of these biases are random and balanced out in large samples. So, in large samples, random error does not affect the accuracy of happiness averages. Yet it does affect correlations; random error "attenuates" correlations. Random error can be estimated using multiple-trait-multiple- method (MTMM) studies, and correlations can be corrected (dis-attenuated) on this basis. A first application on satisfaction measures is reported by Saris, Scherpenzeel, and Veenhoven (1996).

Some biases may be systematic, especially bias produced by technique of interrogation and sequence of questions. Bias of this kind does affect the reliability of the distributional data. In principle it does not affect correlations, unless the measure of the correlate is biased in the same way, i.e., correlated error. To some extent, systematic error can also be estimated and corrected. See also Saris et al. (1996).

4/1.3 Comparability across nations

Average happiness differs markedly across nations. Russians currently score 5.4 on a 0-10 scale, while in Canada the average is 7.7. Does this mean that Russians really take less pleasure in life? Several claims to the contrary have been advanced. Elsewhere I have checked these doubts (Veenhoven, 1993). The results of that inquiry are summarized below.

Words for happiness

The first objection is that differences in *language* hinder comparison. Words like "happiness" and "satisfaction" will not have the same connotations in different tongues. Questions using such terms will therefore measure slightly different matters. I checked this hypothesis by comparing the rank orders produced by three kinds of questions on life satisfaction: a question about "happiness," a question about "satisfaction with life," and a question that invites respondents to give a rating between "best and worst possible life." The rank orders appeared to be almost identical. I also compared responses on questions on happiness and satisfaction in two bilingual countries and found no evidence for linguistic bias.

Valuation of happiness

A second objection is that responses are differentially distorted by *desirability bias.* In countries where happiness ranks high in value, people will be more inclined to overstate their enjoyment of life. I inspected that claim by checking whether reported happiness is indeed higher in countries where hedonic values are most endorsed. This appeared not to be the case.

As a second check, I looked at whether reports of general happiness deviated more from feelings in the past few weeks in these countries, the former measure being more vulnerable to desirability distortion than the latter. This also appeared not to be true.

Response style

A third claim is that *response styles* distort answers to questions about happiness dissimilarly in different countries. For instance, a collectivistic orientation in a country will discourage “very” happy responses because modest self-presentation is more appropriate within such a cultural context. I tested this hypothesis by comparing happiness in countries differing in value collectivism but found no effect in the predicted direction. The hypothesis also failed several other tests.

Concept

A related claim is that happiness is typically a Western concept. Unfamiliarity with it in non-Western nations would lead to lower scores. If so, we can expect more “don’t know” and “no answer” responses in non-Western nations; however, this appears not to be the case.

The issue of *cultural bias* in the measurement of happiness must be distinguished from the question of *cultural influence* on appraisal of the quality of life. Russians can be truly less happy than Canadians but be so because of a gloomier outlook on life, rather than because they have an inferior quality of life.

4/2 Accepted measures of happiness

Having established that happiness can be measured in principle, we can proceed to consider the specific methods of assessing it. We now meet a great variety of questionnaires and interrogation techniques. During the last decades more than a hundred methods have been proposed; some of them presented under impressive names such as 'Life Satisfaction Index', 'General Satisfaction Score' or 'Happiness Scale'. Many of these methods labour under rather obvious defects.

Those methods depend on questioning. Hence the most current defect is that questions are inappropriate. Several do not ask about happiness as defined here but solicit responses about subtly different things. Close reading shows that many questions in so called 'happiness scales' refer to things like 'optimism', 'frustration tolerance' and 'social adjustment'. Investigators who use such questionnaires typically fail to define happiness formally.

Another current defect is that methods are not sufficiently specific. Some 'expert-ratings' for example, do not clearly define what the expert regards as happiness. Similarly, methods based on 'content analysis' sometimes lack clear instructions for interpretation. Again, this is often a result of slovenly conceptualization. Sometimes even more basic defects appear for example when happiness is assessed based on estimates by peers who do not know the individual's private thoughts and therefore base their estimate on his overt behaviour and living conditions.

Elsewhere I have screened all the current measures for applicability to the concepts defined in chapter 2. The following indicators were deemed acceptable (Veenhoven 1984: chapter 4).

4/2.1 Measures of overall happiness

Overall happiness can only be assessed by direct questioning. It cannot be measured indirectly by questions that tap essentially different matters that are assumed to be related to happiness, such as the related concepts discussed in chapter 2 of this introductory text, section 2.3.

Direct questions on overall happiness can use various key terms. One of the appropriate words is 'happiness', provided that the context of the question makes clear that happiness-in-life is concerned, rather than happiness-of-the-moment. Another acceptable term is 'satisfaction-with-life'. Questions can be framed in different ways: as closed questions, as open-ended questions and as focused interviews. In the latter two cases, clear instructions for content analysis of responses are required.

Overall happiness cannot be assessed by peer-ratings, because peers do not know precisely what the subject has on his mind and rather tend to imagine how they themselves would feel if they were in the subject's shoes.

4/2.2 Measures of hedonic level of affect

Hedonic level of affect can be assessed in three ways: 1) by direct questioning, 2) by projective tests and 3) by ratings based on non-verbal behaviour. Again, the method of direct questioning is to be preferred, when the individual is asked several times during a certain period how pleasant he/she feels at that given time (experience sampling).

Though generally less dependable, indirect methods can sometimes suffice. Some projective tests at least seem to be reasonably valid. Ratings by others based on non-verbal behaviour are also acceptable, if rating instructions are sufficiently specific. Unlike cognitive judgments, affective conditions may manifest reliably in non-verbal behaviour.

4/2.3 Measures of contentment

Contentment can be measured only by using direct questions. Like overall happiness, it cannot be validly assessed by indirect questioning or by peer-ratings. Direct questions must again be specific. In this case this means that the question must clearly focus on realization of wants in a life-perspective. Such questions are probably best understood when preceded by an enumeration of one's major aspirations. Questions can again be framed in various formats.

4/2.4 Mixed measures

Finally, there are several acceptable indicators that cover two or more of the above happiness variants. The majority of these consist of single direct questions, which by wording or answer format refer to overall happiness as well as to hedonic level. If they do not labour specific deficits, these questions are acceptable.

Some indicators work with multiple questions. Characteristically these questions cover both overall happiness and one or both discerned components. When all separate questions meet the demands outlined above, such composite indicators are accepted.

A last method to be mentioned in this context is the focused interview of which the 'depth interview' is a variant. Such interrogations tend to cover all three happiness variants. A lack of clear reports on the themes of inquiry and on rating procedures mostly makes it difficult to assess their validity.

4/3 Rejected measures

Many of the currently used questions in the field of subjective wellbeing do not to fit the concept of happiness as defined here. Such indicators are therefore not included in this collection of happiness measures and hence observations yielded by these questions are also not included in the finding collections of this database. This implies rigorous selection: about 80% of the research literature in the field of subjective wellbeing will be left out, for instance almost all the questionnaires on ‘health related quality of life’.

The selection is based on an inspection for face validity. That is, close reading of the questions or instructions to assess whether happiness as defined here is assessed. This process is reported in full detail in Veenhoven 1984: chapter 4 and Veenhoven 2000: section 3). I present some illustrative cases below.

4/3.1 Questionnaires involving questions on other qualities of life

Many currently used measures of wellbeing consist of lists of questions, part of which refer to happiness and part to related concepts. Scheme 4/3.1a is a good help to take stock of the substantive contents of such inventories. As an illustration it is applied to the 24 items in Sheeney's 'Wellbeing Scale. See scheme 4/3.1b. Clearly only some of the questions are in the right-bottom enjoyment quadrant.

Questionnaires on 'adjustment to old age'

Such questionnaires are commonly used in gerontological research. The inventories mix questions about happiness typically with items on 'social participation', 'future orientation' and 'activity' i.e. Lawton's (1975) PGMC and the often used 'Life-satisfaction Scale' of Neugarten et all (1961).

These questions are rejected, because it is not at all sure that high social participation, future orientation and vigour always mark a high appreciation of life. There are always socially active, future orientated and vigorous people who are profoundly dissatisfied with their life. Moreover, one can question an orientation on the future in the third age.

Responses to questions of this kind cause contamination in correlational analysis: e.g. when vigoro is an item in a happiness index, scores on this index correlate with vigorous behaviours. For the purpose of comparison through time and between nations, such scores are also problematic, because concomitants of happiness are typically not the same in all countries at all times. Social activity is more crucial in modern individualistic society than in the context of embedded collectivism.

Scheme 4/3.1a
Difference with other qualities of life

	Outer qualities	*Inner qualities*
Life chances	Livability of environment	Life-ability of the person
Life results	Usefulness of life	**Satisfaction with life**

Scheme 4/3.1b
Illustrative use of scheme 4/3.1a to sort contents in a well-being questionnaire: Sheeney's (1982) 'Wellbeing Scale'

	Outer qualities	*Inner qualities*
Life chances	Has love relation	In control over life
Life results	Lives usefully (vs. ordinary) Contributes to society	Interested in life Satisfied with life-domains **Satisfied with life as a whole** Feels to realize dreams

Health related QOL-questionnaires
A comparable generation of questionnaires has developed in research on the outcomes of medical treatment. There are general purpose questionnaires and questionnaires that focus on the sequel of specific illnesses. An example of the former is the much-used SF-12 (Ware et al 1996). This inventory is largely about physical capability (e.g. climbing the stairs) and functioning in social roles (e.g. work). It also involves questions about general health, vitality and mood. One item is about recent happiness. An example of a special illness inventory is the 'Life Satisfaction Questionnaire (LSQ) by Carlson et. all. (1996). This 43-item questionnaire focuses on the condition of breast-cancer patients. The questions concern physical complaints, daily performance, quality of social relations and several domain-satisfactions. Fear for death is not included in this list, but figures in several other inventories.

Happiness as defined here is at best a side issue in these questionnaires. Hence, they are also rejected.

4/3.2 Questionnaires involving other satisfactions

In chapter 2 of this introductory text, I distinguished happiness from other satisfactions, using scheme 4/3.2a. A lot of 'happiness' inventories cover a wide variety of satisfactions. Scheme 4/3.2b presents an example of the assortment in the 40-item SUBI questionnaire. Items are found in all the boxes. In the sum-score items about overall happiness have the same weight as of sleeping badly. Next to these enjoyment items, the inventory has also questions, that belong in the top-quadrants of scheme 4/3.1a, for instance questions on availability of social support (livability) and about stress resistance (life-ability). There are many such muddy measures. Though they typically involve acceptable items on happiness the sum-scores cannot be accepted as a valid measure of happiness as defined here.

Scheme 4/3.2a
Four kinds of satisfaction

	Passing	*Enduring*
Life aspects	Pleasure	Domain satisfaction
Life as a whole	Top experience	**Life-satisfaction**

Scheme 4/3.2b
Illustrative use of scheme 4/3.2a to sort contents of a satisfaction questionnaire:
Sell (1994) 'Subjective Well-Being Inventory

	Passing	*Enduring*
Life aspects	Agitated Pain Tired	Satisfaction with life-aspects Satisfaction with life domains Fulfilment of expectations Specific worries
Life as a whole	Ecstatic experiences Oceanic feeling	**Overall happiness** **Contentment**

Summed life-domain satisfactions
A common variant is to measure overall satisfaction (right bottom quadrant) by aggregating satisfaction with various life-domains (right top quadrant). For instance, by computing the average of satisfaction with 'work', 'marriage' and 'leisure'. Andrews & Withey (1976) presented several such sum-scores, long and short ones, weighted and unweighted.

This method has several drawbacks. Firstly, it does not adequately reflect the individuals 'overall evaluation'. Such sum-scores tap selected aspects of life only, not the ones selected by the subject. Secondly, not all aspect-satisfactions apply equally well to everybody, how about marriage-satisfaction of the unmarried and the job-satisfaction of the unemployed? Thirdly, the significance of life-aspects such as 'work' and 'marriage' is not the same across time, culture and social categories. Comparison is therefore often not possible using such indicators.

Part of these problems can be met by asking respondents to rate domains by importance and then compute a weighted average. Yet this does not solve the problem of missing domains and it is not sure that perceived importance equals actual impact. In fact, this method assumes a 'bottom up' evaluation of life, while there is growing evidence that 'top down' appraisal is most common (Veenhoven 1997: 59-62).

Summed life-aspect satisfactions
These objections also apply to 'semantic-differential scales', which involve the ratings of one's life on various evaluation criteria, such as 'boring/interesting', 'lonely/friendly' and 'hard/easy'. Again, each of the questions falls short as an indicator of overall happiness, an interesting life is not necessarily a satisfying life, and neither is an easy life always more gratifying than a hard life. Taken together several such questions do not provide a good estimate of the overall evaluation either, because the weights are likely to differ across persons and to be variable across time and culture. Such a semantic differential scale is part of the much-used 'Index of Wellbeing' of Campbell et al. (1976), and this index is therefore not acceptable.

4/3.3 Deficient questions on happiness

There are also many questions that do focus on happiness as defined here, but that fail to use sufficiently sharp questions. Some illustrative examples are presented below. This check pans out negatively for many multiple item questionnaires on happiness. The more items the greater the chance of one being incorrect, and if one item is wrong, the whole questionnaire is rejected.

Comparison with others
Several investigators have asked their subjects how happy they think they are compared to others, rather than how they feel themselves. Such questions are rejected. Even if one is happier than one's neighbour is, one can still be unhappy. This invalidates the four-item 'Subjective Happiness Scale' (SHS) of Lyubomirsky and Lepper (1999), the second item of which is perceived happiness relative to peers.

Better than in the past
For the same reason the question 'I have been happier than I am now' is deemed unacceptable. Being less happy than before does not imply that one is unhappy.

Preference for another life
Some investigators derive happiness from responses about questions on appreciation of alternative ways of life. For instance: one of the questions in Diener's (1985) 'Satisfaction With Life Scale' (SWLS) is 'If I could live my life over, I would change nothing'. This item is also rejected, and thereby the scale as a whole. An individual could be quite happy, but still opt to try life another way if one could live one's life over. There are clearly many roads to happiness and most people know this.

4/4 Sense of selectiveness

Altogether I reject more than half of the questions that claim to tap happiness or life satisfaction. Since this involves many commonly used inventories, this selection leaves out some 80% of the research literature in the field of subjective wellbeing. Why be so choosy?

The reason is we will never get any wiser if we go on using sloppy concepts. That is why in chapter 2 of this introductory text, I defined happiness in a much narrower way than understood in common language. The aim of this database is to gather the empirical findings on this specific kind of happiness, with the very purpose of reducing the conceptual ambiguity found in current research. Obviously, this requires that we limit only to items that deal with matter as defined in section 1.

Why then focus on the enjoyment of life and not on another quality of life? I have given several reasons in section 2/6 of chapter 2. Happiness is the most comprehensive indicator of quality of life. This concept does not bring us into circular reasoning when evaluating society and our way of life. In this section we have seen that it is also a matter that can be measured full stop.

I would not be so choosy if this 'Collection of Happiness Measures' were a mere test bank, such as such as the ones mentioned in the introductory section. Yet this collection is part of the wider finding archive is used to define which investigations will be included, and hence what kind of findings will become available for comparative analysis. Since one cannot compare apples and oranges, the findings must pertain to the same subject matter.

REFERENCES

Andrews, F.M. & Withey, S.B. (1976)
Social indicators of well-being New
York, U.S.A., Plenum Press.

Campbell, A & Converse, P. & Rodgers, W.L. (1976)
The Quality of American Life New York, U.S.A.,
Sage.

Carlsson, M. & Hamrin, E. (1996)
Measurement of quality of life in women with breast cancer. Development of a Life Satisfaction Questionnaire (LSQ-32) and a comparison with the EORTC QLQ-C30
Quality of Life Research, vol. 5, pp. 265-274

Diener, E., Emmons, R.A., Larsen, R.J. & Griffin, S. (1985)
The satisfaction with life scale
Journal of Personality Assessment, Vol. 49, 71-75.

Lawton, M.P. (1975)
Philadelphia geriatric centre morale scale-revision
Journal of Gerontology, Vol. 30, 85-89.

Lyubomirsky, S. & Lepper, H.S. (1999)
A measure of subjective happiness: Preliminary reliability and construct validation
Social Indicators Research, vol. 46, pp. 137-155

Neugarten, B.L., Havighurst, R.J. & Tobin, S.S. (1961)
The measurement of life-satisfaction
Journal of Gerontology, Vol.16, 134-143.

Sell, H. (1994)
Subjective Well-Being Inventory (SUBI)
International Journal of Mental Health 23 (3) 89-102

Sheeney, G. (1982)
Pathfinders
Sidwick & Jackson, London

Veenhoven, R. (1984)
Conditions of happiness
Kluwer, Dordrecht, Netherlands

Veenhoven, R. (1993)
Happiness in nations, RISBO.
Rotterdam: Erasmus University.

Veenhoven, R. (1997)

Progrès dans la compréhension du bonheur
Revue Québécoise de Psychologie, vol 18, pp 29-74
Also available in English 'Advances in the understanding of happiness'

Veenhoven, R. (2000)
The four qualities of life
Journal of Happiness Studies, Vol. 1, pp. 1-39

Veenhoven, R. (2008)
Comparability of happiness across nations
School of Sociology and Social Work Journal, no 104, 2008, pp. 211-234, Kwansei Gakuin University, Japan, ISSN 0452-9456

Ware, J.E., Kosinski, M. & Keller, M.D. (1996)
A 12-item short form Health Survey: construction of scales and preliminary tests of reliability and validity
Medical Care, vol. 34, pp. 220-