

THE INTERNATIONAL SCALE INTERVAL STUDY

Improving the comparability of responses to survey questions about happiness

Ruut Veenhoven, Erasmus University Rotterdam, The Netherlands,
E-mail: veenhoven@fsw.eur.nl

Published in: Mølle, V. & Huschka, D. 'Quality of Life in the new millennium Advances-in-quality studies, theory and research', Springer Press, Social Indicator Research Series 2008, Dordrecht, The Netherlands.

*Part 2: Refining concepts and measurement to assess cross-cultural quality-of-life.
Chapter 5*

ABSTRACT

This study is about survey questions on happiness using verbal response options, such as 'very happy' and 'fairly happy'. The aim is to estimate what degrees of happiness are denoted by such terms in different questions and languages. These degrees are expressed in numerical values on a 0 to 10 scale, which are then used to compute 'transformed' means and standard deviations.

Native speakers read survey questions on happiness that have been used in their country. For each question separately, they rate the relative value of each of the response options in their language. They do that using an electronic 'Scale Interval Recorder'. On their computer screen, participants see a vertical bar scale that they can partition into sections by shifting separation lines. The response options are presented next to the scale and move with the bars. Their task is to move the separation lines until they feel that the intervals on the scale correspond with the degree of happiness denoted by each of the verbal response options. The aim is to cover 74 languages.

This study will allow a better use of the available survey data on happiness

1 INTRODUCTION

The subject of happiness has long been a playground for speculative philosophy, but in the last decades it became also a subject of social scientific research. This research focuses on happiness in the sense of overall life satisfaction and assesses that matter in large-scale surveys. This has resulted in a considerable body of empirical findings; much of which is stored in the World Database of Happiness (Veenhoven 2006). At this moment (July 2006) that database contains the results of some 2500 survey studies in 112 nations and this provides a rich source for comparative analysis across time and nations.

Happiness is typically measured using single direct questions. Some examples are presented in [scheme 1](#). Though all these questions aim at the same concept, they differ in wording and response options. The 'Item Bank' of the above mentioned World Database of Happiness contains no less than 850 variations. ([Scheme 1](#))

This variation in questions used limits the comparability of the available data. As a result, only part is effectively used in comparative analysis and in synthetic studies. For instance, studies on the trend of happiness in the USA typically restrict to answers to one particular question (3-step GSS item) leaving 80% of the available data aside. Likewise, comparison of happiness across nations limits typically to cases for which a particular item on life satisfaction is available (10-step WVS item), leaving out some 30 countries where happiness has not been assessed in that way.

There are several ways to cope with this heterogeneity, which I have described elsewhere (Veenhoven 1993 chapter 7). One of these ways is transforming scores on different questions to a same scale. In the case of numerical scales this can be done using linear transformation, e.g. stretching a score on a 1 to 5 scale to range 0-10. A variant of that approach is Cummins' (1995) percentage of the scale maximum (%SM).

Another method is transforming scores on scales with verbal response options to a common numerical range. This requires that experts rate the numerical equivalents of verbal response options, e.g. a rating of 5 for the verbal response option 'neither happy nor unhappy' on a 0 to 10 numerical scale. This method has been proposed by Thurstone as early as in the 1950s (Jones & Thurstone 1955) and is recently also applied by Smith et. al. (2005)¹.

This paper describes a web-based variation of this rating method and reports a large-scale application that involves all the questions on happiness ever used in general population surveys.

2 APPROACH

This study limits to survey questions on happiness using verbal response options, such as the first three items in [scheme 1](#). Results yielded with such questions can hardly be compared with results yielded by slightly different questions. Below I explain why and then show how comparability can be improved.

2.1 Rate numerical value of verbal response options

As we have seen, happiness is commonly measured using single questions that such as "Taking all together, how happy would you say you are these days?" Such questions are answered by choosing from a list of response options that are ordered from more to less happiness. Often these options are denoted by words such as 'very happy', 'pretty happy' and 'not too happy'. Such response scales are of the *ordinal level* of measurement, which does not allow the computation of means and standard deviation

Researcher typically avoid that problem assigning numerical values to the verbal response options; in the above case of three response options typically 3 for 'very happy', 2 for 'pretty happy' and 1 for 'not too happy' and they on this basis compute means and standard deviations. Thus they create a (semi) *interval* scale, which does allow mathematical computation.

In doing so, researchers implicitly assume that the distance between 'very' and 'pretty' happy is the same as between 'pretty' and 'not too' happy. This may not be the

case, possibly respondents see a greater distance between the latter than between the former. If so, unhappiness will be underestimated. Individual scores of 3, 2 and 1 will not fully reflect the real differences in happiness and as a result the variance shared with other variables will not be fully reflected in correlations. Aggregated scores will also be affected, mean scores will be higher and standard deviations lower than factual.

In this study we assess what people have in mind when they tick 'very' or 'pretty' happy and we will use that information to estimate more accurate numerical values for verbal response options to questions about happiness. For example, it might appear that the value of 'pretty happy' on scale 1 to 3 is actually 2.5 instead of 2 and 'not too happy' 0.8 instead of 1. We then use these estimates to re-analyze available data and check whether this refinement really makes a difference.

2.2 Standardize scales to range 0-10

In synthetic studies on happiness, researchers often combine findings obtained using slightly different questions. As they try to maximize the number of observations, they accept some diversity in the studies they include. This causes several problems, which can to a large extent be solved by the results of this study.

Overcoming difference in wording of response options

Researchers typically assign the same numbers for response options denoted with slightly different words. For instance, if the third response option in the above question is 'unhappy' instead of 'not too happy', they also code this response as 1. Obviously, this involves a loss of information.

This study will help us to do better, because it should generate more accurate estimates of the numerical values to be used for specific response options. For example, it might possibly show that the average respondent would equate 'not too happy' with 1.1 point on a 1 to 3 scale and 'unhappy' with 0.8 points. These values can then be used to compute weighted scores that more accurately reflect the actual differences in the happiness at stake.

Overcoming differences in number of response options

Another problem in research synthesis is that response scales differ in the number of response options. The above example of a survey question in happiness involves three response options, but there are also questions with four responses options and even questions that offer seven.

Researchers solve this problem by transforming observed scores to one common scale. One way is to downsize the longer scales, e.g. shortening a 4-step scale to range 1-3 by lumping the last two options together. This involves a loss of information and the danger of distortion. Alternatively one can stretch the scales linearly to a common range, e.g. when drawn out to range 0-10, score 2 on range 1-3 becomes 5. This method is discussed in more detail in Veenhoven & Kalmijn (2005: 447-9). In an earlier application on happiness it appeared to produce implausible results, in particular when applied on short scales.

In this study we solve the problem in another way. We ask proto-respondents to assign the values of response options on a common numerical scale. We present them with a scale ranging from 0 to 10 and ask them to partition that scale into intervals that

correspond to the degrees of happiness denoted by the words used for response options to questions about happiness. The intervals will be greater with shorter response scales, at least on an average, and the meaning of words may vary accordingly. For example, on a 3-step scale the response option 'very happy' may be seen to cover the range 10 to 7, with a mean of 8.3, whereas on a 5-step scale the option 'very happy' may be seen to denote range 10 to 9 with a midpoint of 9.5. Once obtained, these values will enable us to calculate comparable scores from available frequency distributions of responses to such different questions.

Overcoming language differences

Still another problem is that the same question is often asked in different languages. The meaning of the words used to denote a degree of happiness may rate differently in different tongues. This is a problem in nations where multiple languages are spoken, such as in South Africa, Switzerland and Belgium. It is also a problem for cross-national comparisons of happiness.

The most common solution for this difficulty is to reduce translation error, typically by using the technique of forth and back translation. Yet perfect translation is often not possible. A commonly mentioned example is that the English word 'happy' can only be translated in French to the word 'heureux', which however, may denote a higher degree of satisfaction.

With this study we also by-pass this translation problem. We ask native speakers to rate the numerical value of words used for response options in their own language. If it is true that the French are more choosy about how they use the word 'happy', they might place the option 'tres heureux' in the range 10 to 9, whereas the English raters would place 'very happy' on range of 10 to 8. This would result in different numerical mid-values, respectively 9.5 and 9.0. These differences are taken into account when we use the values to compute weighted averages from available frequency distributions. .

3 GOALS

The above described approach serves the following aims:

3.1 Better comparability of happiness over time

The average happiness of citizens was assessed for the first time in the USA in 1945. Since then, more than 200 assessments have followed, but it is still not clear whether Americans have become happier or not. One problem is that increments tend to be small close to the ceiling, but another problem is that the survey questions have differed slightly over time and that this is likely to obscure the overview of the small trend. That latter problem is typically solved by limitation to identical questions, but this means that about half of the available data must then be left out, yet we need large amounts of data to discern the trend from random variation.

This study alleviates this problem in two ways. Firstly it enhances the comparability of responses to questions that differ only slightly in the wording of responses options; e.g. 3-step items using 'pretty happy' for the second option instead of 'fairly happy'. As argued above, the subtle difference between such words will reflect in

the different values assigned by our raters, which are then taken into account in the weighted mean we calculate. Secondly, this study enables the comparison of responses to questions that differ in number of response options. As indicated above, the rating procedure is likely to neutralize the differences in length of response scales. Together this will broaden the database that can be used for analyzing change in happiness in nations.

3.2 Better comparability of happiness across nations

Currently, the section on ‘Happiness in Nations’ of the World Database of Happiness contains 112 nations where general population surveys had included questions on happiness. Yet again the questions used are not identical. The most commonly used question had only been applied in 78 nations and the translation of this item into the different languages is questionable in some cases. This impedes our ability to make comparative analyses of happiness in nations.

This study can lessen the above problem. As in the case of comparison over time it should broaden the database available, as differences in phrasing of response options and the difference in number of response options should cease to be a problem, and finally, translation error should be much reduced by our method.

3.3 Better comparability of correlational findings on happiness

In July 2006 the World Database of Happiness included 9.233 correlational findings on happiness yielded by 912 empirical studies. This collection has been gathered for the purpose of facilitating research synthesis and for meta-analysis in particular. However, the collection has hardly been used for this purpose as yet and one of the main reasons is that the measures of happiness used in the various studies differ too much.

This study should also serve to resolve this problem. All the benefits mentioned above apply also for this problem of heterogeneity of measurement of happiness. Meta-analysis will benefit from more accurate estimates of happiness from the available data and from better comparability of responses to questions that differ in wording, number of response options and language. However these benefits can only be reaped if the full distribution of responses is available, which is not always the case with correlational findings. Still, the obtained weights can be used in secondary analysis of available datasets.

3.4 Better measurement of happiness

This study should improve the measurement of happiness. Ideally, the meaning of response options to a question is the same for all respondents. Yet in practice there are always differences in interpretations of words and some words give rise to more differences than others. The use of such words must be avoided and therefore it is worth knowing which words cause confusion.

In this study we can identify such words using the standard deviation of the ratings. For example, if our English respondents differ more in their ratings of the term ‘rather happy’ than of ‘fairly happy’ while the midpoints are the same, future researchers should better avoid the former term.

4 METHOD

As noted above, this study is about survey questions on happiness using verbal response options. Since the main aim is to improve comparison of happiness across nations, we restrict to such questions as those that have been applied in studies of general populations in nations. These survey questions are specified on the 'item list' on [scheme 2](#).

4.1 Assessment

Native speakers will be asked to consider questions separately and estimate the degree of happiness denoted by the different response options provided in their language. This will be done on a bar scale, ranging from 10 to 0. The native speaker's task is to partition the scale into sections that correspond with the meaning denoted by the words used to qualify happiness. The native speakers will be asked to do this for each of the response options of several questions.

4.2 Instrument

The assessments will be made on a computer screen, which displays the survey question and a vertical bar scale. Next to this will be the verbal response options, such as 'very happy' or 'not happy'. On the vertical bar there will be horizontal lines that can be moved by the cursor to section the bar. The user can move these lines up and down and thereby divide the bar into sections of different size. The response options next to the line will also move. The native speakers must shift the boundaries until they feel that the segments on the bar correspond with the meaning of the words as used for the response options. This tool is named a 'Scale Interval Recorder' and is available on request (Veenhoven & Hermus 2005). An example of possible scores on this instrument is presented on [scheme 3](#).

4.3 The judges

University students will be invited to do this job. We opted for university students rather than 'average citizens', because the task requires a more than average verbal intelligence. We ignore that this may be at the cost of the validity, since the significance of words may depend on the subculture of their users.

Recruiting

Volunteers will be recruited through professors who participate in the study. Professors will explain the study to the students in class and then hand out a flyer with further details and a log-on code. Interested students can then do the assessments on a PC with Internet connection. This task is expected to take about 10 minutes per session, in which about 8 scales are judged consecutively.

Number

Estimates will be more reliable the more judges are involved. We aim at a precision of 0.1 and a 0 to 10 scale in a 95% confidence interval, which will require about 200 student judges to participate per language/country.

4.4 Analysis

The responses will be automatically recorded and transformed to an MS-Excel file. Using this file we will compute the average value allotted to each response option on a scale of 0 to 10. We will then use these scores to recalculate the means and standard deviations from distributional findings on happiness in nations already available in the section 'Happiness in nations' of the World Database of Happiness mentioned above.

Example

The response options with three response options were rated as follows in the Dutch language: 8.5 for 'very happy', 5.7 for 'fairly happy' and 2.5 for 'not too happy' (study dutch1). When used in a survey in The Netherlands in 1985 the frequency of responses to these options was respectively 71%, 27% and 2%. The values obtained were then used to compute a weighted average from these frequencies: $0.71 \times 8.5 + 0.27 \times 5.7 + 0.02 \times 2.5 = 7.6$. The ratings for the 'same' response options after translation into the English language were slightly different, 8.2 for 'very happy' and 2.6 for 'not too happy' (study english1). Given a same distribution of responses that would yield a slightly lower average: $0.71 \times 8.2 + 0.27 \times 5.7 + 0.02 \times 2.6 = 7.4$.

4.5 Validation

The question then is: will these weighted averages provide a more accurate estimate of happiness in nations than the currently used unweighed means, which assumes equidistance and absence of language differences? To answer that question we will compare the performance of the calibrated and uncalibrated means in a cross-national analysis, to find out if objective country characteristics such as income per head correlate stronger with the former than with the latter.

The data will also be used to calibrate standard deviations of happiness, which are used as a measure of inequality in nations (Kalmijn & Veenhoven 2005)

5 ILLUSTRATIVE RESULT

One of the standard questions in the World Values Survey reads: *Taking all things together, would you say you are.....? very happy, quite happy, not very happy or not at all happy?* A Dutch version of that question figured in the 1999 survey in the Netherlands and in that same year the English version was fielded in the UK. The observed frequency of responses to each of the answer options are presented in [scheme 4](#).

Application of ratings from 1 for 'not at all happy' to 4 for 'very happy' results in a mean score of 3.40 in the Netherlands and 3.21 in the UK. This difference of 0.19 on the 1-4 scale is 6.3% of the scale range.

Application of the method described in this paper yields a slightly different result. The numerical values assigned to the four response categories were not identical in the Dutch language (study dutch1) as in English language (study english1), the British judges give somewhat more weight to the lower categories. See [scheme 4](#). The resulting mean

score are respectively 7.29 in the Netherlands and 6.97 in the UK, and the difference is now only 3.2% of the scale range.

This preliminary result suggests that the difference in average happiness between the Netherlands and the UK is smaller than earlier findings on the basis of this question suggest. (Scheme 4)

6 CALL FOR COOPERATION

As noted above, this study aims to cover questions in 72 different languages and requires the participation of some 15,000 students. At this stage we have co-investigators for the following languages: Arabic, Chinese, English, Danish, Dutch, Portuguese, Spanish and Turkish. The latest update can be found on the web-site mentioned above. We welcome help from colleagues, especially in language areas not covered as yet.

7 WEBSITE

This study entirely web-based, Not only are the data gathered using a website, but also is all the documentation on the website, including the questionnaires and the findings. There is no paper archive in Rotterdam. So this research is entirely open to inspection. Anyone who may doubt the conclusions can check. The address of the study web site is:

http://worlddatabaseofhappiness.eur.nl/scalestudy/scale_fp.htm

Scheme 1

Some survey questions on happiness

- Taken all together, how would you say things are these days? Would you say that you are?
 - very happy
 - pretty happy
 - not too happy

 - On the whole how satisfied are you with the life you lead?
 - very satisfied
 - fairly satisfied
 - not very satisfied
 - not at all satisfied

 - In most ways my life is close to ideal
 - strongly agree
 - agree
 - slightly agree
 - neither agree nor disagree
 - slightly disagree
 - disagree
 - strongly disagree

 - How satisfied are you with your life as a whole these days?
 - 7 completely satisfied
 - 6
 - 5
 - 4 neutral
 - 3
 - 2
 - 1 completely dissatisfied
-

Scheme 2

Questions on happiness involving verbal response options, ever used in general population surveys

Item code*	Survey program in which question was used
2-step verbal happiness (table 111A in Happiness in Nations)	
O-HL/u/sq/v/2/a	The Netherlands, NIPO survey 1947
3-step verbal happiness (table 111A in Happiness in Nations)	
O-HL/c/sq/v/3/aa	Australian national Science Survey 1984, South Korean national survey 1980, Polish surveys 1991-2000, Spanish CIRES survey 1985, US General Social Survey since 1957,
O-HL/c/sq/v/3/ab	Euro-barometer (periodical multi-nation survey) bi-annual 1975-1986
O-HL/c/sq/v/3/ah	US General Social Survey since 1957,
O-HL/g/sq/v/3/a	US incidental Gallup polls 1947-1990
O-HL/g/sq/v/3/d	International Gallup/Kettering World survey 1975, US Gallup polls 1975-2004
O-HL/g/sq/v/3/f	International Gallup World Survey 3, 1965
O-HL/g/sq/v/3/g	US Gallup polls 1956, 1957
O-HL/g/sq/v/3/j	International Gallup World Survey 1946-48 (probably nr 1)
O-HL/g/sq/v/3/k	USA Gallup poll 1995
O-HL/g/sq/v/3/l	USA Gallup polls 1948, 1956
O-HL/g/sq/v/3/i	USA Gallup polls 1948, 1956, 1966, 1970
O-HP/u/sq/v/3/a	Netherlands, NIPO survey 1948
O-HP/u/sq/v/3/b	Germany, Allensbacher periodical polls 1954-1992
O-H?/?/sq/v/3/a	Japan periodical Life in Nations Surveys, International Gallup Poll 1970
4-step verbal happiness (table 111B in Happiness in Nations)	
O-HL/c/sq/v/4/a	Scandinavian one-time Welfare Survey 1972, German Wohlfahrt surveys 1980, 1984, 1988, 1990, 1993, 1998
O-HL/c/sq/v/4/f	International Gallup survey 1946, ISSP 1991, Polish GSS 1992, 1993, US GSS 1998, 2002
O-HL/c/sq/v/4/g	International Euromodul 1999, Former communist countries LLH survey 2001, Hong Kong and Taiwan Social Trends 2000
O-HL/c/sq/v/4/h	Russian panel study RUSSET 1993-1999
O-HL/c/sq/v/4/o	Poland periodical Social Diagnosis Survey 1993, 2003
O-HL/c/sq/v/4/p	World Value Survey (Japanese version) 1980, 1990, 1996, 2000
O-HL/g/sq/v/4/b	USA Gallup polls 1946-1956
O-HL/g/sq/v/4/f	International Latino Barometro 2002
O-HL/g/sq/v/4/g	Brazil 2001 (ISSP)
O-HL/g/sq/v/4/h	Ireland 2001 (ISSP)

O-HL/ly/sq/v/4/a	South Korea one-time Social Development Surveys 1981, 2001
O-HL/u/sq/v/4/a	International World Values Surveys 1980, 1990, 1996, 2000
O-HL/u/sq/v/4/c	World Youth Surveys 1988, 1993
O-HL/u/sq/v/4/e	Romanian Public Opinion Barometer 2005
O-HP/c/sq/v/4/a	Canadian GSS 1986
O-HP/g/sq/v/4/b	Canadian GSS 1985, 1989, 1990
5-step verbal happiness (table 111C in Happiness in Nations)	
O-HL/c/sq/v/5/e	International Leisure Development Survey 1979
O-HL/c/sq/v/5/g	South Africa periodical Quality of Life Surveys 1983-2003
O-HL/c/sq/v/5/m	International Asia barometer 2003
O-HL/g/sq/v/5/b	Sweden QOL survey 1985
O-HL/u/sq/v/5/b	Austria QOL survey 1984
O-HP/u/sq/v/5/a	Netherlands periodical QOL survey since 1977
O-HP/u/sq/v/5/d	Netherlands Dutch Household Survey, yearly panel since 1993
O-HP/u/sq/v/5/e	Netherlands Cultural Change Survey 2004
7-step verbal happiness (table 111D in Happiness in Nations)	
O-HL/g/sq/v/7/a	International periodical Social Survey Program (ISSP) 2003
2-step verbal life satisfaction (table 121A in Happiness in Nations)	
O-SLP/c/sq/v/2/a	USA Gallup polls 2001-2004
O-SLP/g/sq/v/2/b	USA Gallup polls 1973-2001
3-step verbal life satisfaction (table 121B in Happiness in Nations)	
O-SLL/c/sq/v/3/a	USA, various surveys 1968, 1972, 1976
O-SLL/u/sq/v/3/a	Canada various surveys 1968, 1974, 1977
O-SLL/u/sq/v/3/b	Eastern Europe New Democracies Barometer 1991
O-SLS/c/sq/v/3/a	International Tension Study 1948
O-SLW/c/sq/v/3/a	Eastern Europe New Democracies Barometer Survey 1991
4-step verbal life satisfaction (table 121C in Happiness in Nations)	
O-SLL/u/sq/v/4/a	Japan Life-in-Nation Survey s, yearly 1958-1963
O-SLL/u/sq/v/4/b	Euro-barometer survey in all member states, since 1973
O-SLL/u/sq/v/4/c	Japan Life in Nation Survey 1984
O-SLL/g/sq/v/4/d	Romania, Public Opinion Barometer, bi-annual since 1996
O-SLP/c/sq/v/4/a	USA Gallup polls 2001-2004
O-SLP/u/sq/v/4/a	USA Gallup poll 1991
O-SLP/?/sq/v/4/c	Hong Kong and Taiwan periodical Social Trend Studies 2000

O-SLu/c/sq/v/4/e	Japan Life-in-Nation Survey, yearly 1964-1983
O-SLu/c/sq/v/4/f	Japan Life-in-Nation Survey, yearly 1980-1991
O-SLu/g/sq/v/4/b	International Latino Barometro 1997, 2000
O-SLu/g/sq/v/4/c	International Latino Barometro 2001, 2002, 2004
O-SLu/u/sq/v/4/b	USA Gallup poll 2003
O-SLu/u/sq/v/4/c	USA Gallup poll 2004
O-SLW/c/sq/v/4/b	Canada survey 1989
O-SLW/c/sq/v/4/c	Former communist countries, LLH survey 2001
O-SLW/c/sq/v/4/e	Japanese Life-in-Nation Survey, periodical) since 1992
O-SLW/c/sq/v/4/f	China Gallup polls 1997, 1999, 2004, US Gallup poll 2004
O-SLW/g/sq/v/4/b	USA Gallup poll 1997
O-SLW/u/sq/v/4/b	International Candidate Countries Euro-barometer 2001, 2002, 2003
O-SLW/u/sq/v/4/d	Canadian GVP 1997, 2000
O-SLW/u/sq/v/4/e	USA Gallup poll 1991
O-SLW/?/sq/v/4/b	Finland, survey 1960
O-SQL/u/sq/v/4/a	Italian Barometro Sociale 1996
5-step verbal life satisfaction (table 121D in Happiness in Nations)	
O-SLL/c/sq/v/5/d	Netherlands periodical QOL survey since 1974
O-SLL/u/sq/v/5/a	European Readers Digest Survey 1990
O-SLu/c/sq/v/5/a	International Leisure Development Survey 1979
O-SLu/g/sq/v/5/c	International IRMC Survey 1991
O-SLu/u/sq/v/5/a	Polish survey 1960
O-SLW/c/sq/v/5/b	Austrian QOL survey 1984
O-SLW/c/sq/v/5/g	South Africa QOL survey 1983, Spanish survey 1993
O-SLW/c/sq/v/5/h	Russia panel RLMS panel, 1992-2002, Kyrgyzstan KMPS 1993
O-SLW/c/sq/v/5/k	Japanese Future Life Surveys 1995, 1997, 1999
O-SLW/c/sq/v/5/l	Japanese Future Life Survey 2001
O-SLW/c/sq/v/5/o	China Gallup survey 1994
O-SLW/c/sq/v/5/p	India, World value Survey 2002
O-SLW/u/sq/v/5/e	Romania Diagnosis of Quality of Life surveys, since 1993
O-SLW/u/sq/v/5/g	Japan Life Style Preference Survey, tri-annual since 1978
O-SLW/u/sq/v/5/m	Candidate Countries Euro-barometer 2004.1
O-SLW/u/sq/v/5/h	Japanese Life Style Preference Survey 2001
O-SL?/c/sq/v/5/a	Taiwan Living Conditions Survey 1998
7-step verbal life satisfaction (table 121F in Happiness in Nations)	
O-SLW/c/sq/v/7/a	Sweden QOL survey 1982

7-step verbal delighted-terrible life (table 141A in Happiness in Nations)	
O-DT/c/sq/v/7/aa	Michalos's International Student Survey 1985
O-DT/c/sq/v/7/ab	Diener's International Student Survey 1995
O-DT/u/sq/v/7/a	Australian QOL survey 1978, US QOL surveys 1972, 1978
O-DT/u/sq/v/7/d	USA Social Indicator Study 1972
8-step verbal delighted-terrible life (table 141B in Happiness in Nations)	
O-DT/u/sq/v/8/b	Australian National Social Science Survey 1995
Affect	
A-AOL/m/sq/v/5/b	Russia one-time Erasmus Survey 1991
A-AOL/md/sq/v/5/a	USA one-time Gallup poll 2001
Contentment	
C-RG/ly/sq/v/4/b	South Korea, Social Development Survey 1981 and Democracy barometer 2001
Mixed items	
M-FH/u/sq/v/3/a	Netherlands, NIPO 1988
M-FH/u/sq/v/3/d	Netherlands, NIPO 1965, 1995
M-FH/g/sq/v/4/a	Poland PGS-BS 1992, 1993
M-FH/cw/sq/v/6/a	Romania Diagnosis of Quality of Life surveys, annual 1993-1997
M-PL/c/sq/v/4/b	South Korea, Social Development Survey 1981 and Democracy barometer 2001
M-TH/cm/sq/v/6/a	Item in health inventory SF-36. Used in one-time surveys in Denmark 1994, Croatia 1999,2003 and Hong-Kong, Romania 1996
M-TH/g/sq/v/5/c	Romania Diagnosis of Quality of Life survey 1993

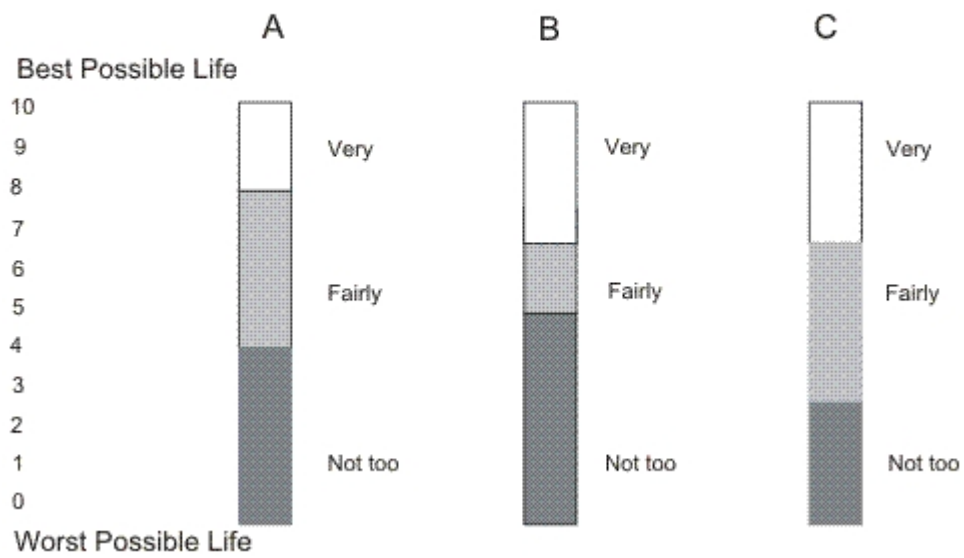
* The full text of the above mentioned questions on happiness can be found in the 'Item bank' of the World Database of Happiness (http://www1.eur.nl/fsw/happiness/hap_quer/hqi_fp.htm), using the item codes in the left column.

Scheme 3
Scale interval recorder

Taking all together, how happy would you say you are?
Would you say you are:

- very happy
- fairly happy
- not too happy

Examples of three possible ratings



Scheme 4
Illustrative results

Question	Taking all things together, would you say you are.....				Mean	
	<i>very happy</i>	<i>quite happy</i>	<i>not very happy</i>	<i>not at all happy</i>	<i>range 1-4</i>	<i>range 0-10</i>
Dutch						
Frequency of responses in %	45.9	49.1	4.2	0.6		
Standard rating	4	3	2	1	3.40	8,00 ¹
Assessed ratings	9,03	6,08	3,74	0,89	-	7.29
English						
Frequency of responses in %	33.1	57.1	7.2	2.4		
Standard ratings	4	3	2	1	3.21	7.37 ¹
Assessed ratings	8.6	6.67	3.6	1.04	-	6.97

¹) after linear scale transformation

REFERENCES

- Cummins, R.A. (1995)
On the trail of the gold standard for subjective well-being
Social Indicators Research, vol. 35, pp. 179-200
- Kalmijn, W.M. & Veenhoven, R. (2005)
Measuring inequality of happiness in nations: In search for proper statistics
Journal of Happiness Studies, Special issue on 'Inequality of Happiness in Nations', vol. 6, pp. 357-396
- Kalmijn, W.M. (2007)
International Scale Interval Study, Part 2: Methodological considerations
In preparation
- Jones, L.V. & Thurstone, L.L. (1955)
The psychophysics of semantics: An experimental investigation
Journal of Applied Psychology, vol. 33, pp. 31-36
- Smith, T.W., Mohler, P. Ph., Harkness, J. & Onodere, N. (2005)
Methods for assessing and calibrating response scales across countries and languages
Comparative Sociology, vol. 4, pp. 365-415
- Veenhoven, R. (2006)
World Database of Happiness: Continuous register of scientific research on subjective enjoyment of life
Erasmus University Rotterdam, The Netherlands. Available at:
<http://worlddatabaseofhappiness.eur.nl>
- Veenhoven, R. & Hermus, P. (2006)
Scale Interval Recorder: Tool for assessing relative weights of verbal response options on survey questions
Web survey program, Erasmus University Rotterdam, Dpt. of Social Sciences & Risbo
Contract research.
- World Database of Happiness, see Veenhoven (2006)

¹ This study limits as yet to only three languages. Unlike our study, it rates response options irrespective of the lead question.